Technische Universität München
TUM School of Engineering and Design

Advancing hydrological modeling in high-elevation Alpine catchments

Florentin Benedikt Hofmeister

Vollständiger Abdruck der von der TUM School of Engineering and Design der Technischen Universität München zur Erlangung eines

Doktors der Ingenieurwissenschaften (Dr.-Ing.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Nils Rüther

Prüfer*innen der Dissertation:

1. Prof. Dr. Gabriele Chiogna
2. Prof. Dr.-Ing. Markus Disse
3. Prof. Dr. Ralf Ludwig

Die Dissertation wurde am 30. 03. 2023 bei der Technischen Universität München eingereicht und durch die TUM School of Engineering and Design am 04. 07. 2023 angenommen.

To my wife Franzi

“What goes up must come down”

– Some old wise guy –

Contents

Abstract

Mountainous regions supply lower elevated regions around the world with fresh water and therefore mountains are often called “water towers”. More than half of the world’s population depends on the watershed function of mountains. Although rich in fresh water, the demand for water resources has a high potential for user conflicts (e.g., drinking, hydropower, irrigation, tourism). Additionally, climate change is having a strong impact on the high mountain regions with dramatic consequences for the cryosphere (i.e., glacier, snow and permafrost). The European Alps are classified as particularly vulnerable and are experiencing an accelerated warming since the 1980s. However, insufficient data availability, data quality, and process understanding limit the assessment of consequences for the hydrology in high-elevation Alpine areas. These limitations significantly reduce the applicability of hydrological models, which are required to provide a spatially and temporally highly resolved assessment of possible climate induced changes on Alpine hydrology as a whole or its individual components. This dissertation comprises of four published research articles which each deals with an important component of hydrological modeling in high-elevation Alpine catchments and aims to increase its reliability and efficiency within a holistic approach.

Since every scientific study of time series analysis or modelling depends on observational data, the first article examins the quality of hydrometeorological data of two high-elevation Alpine catchments and investigats to what extent the data quality influences the model results. We observed that inhomogeneties contained in the temperature and precipitation time series that drive a hydrological model, have a direct impact on main components of the cryosphere (i.e., snow and glacier accumulation and melt). They only indirectly influence the runoff generation, since the runoff generation includes superposition effects of a nonlinear hydrological system.

Another requisite for a comprehensive model calibration and validation is continuous and reliable runoff observation data with a low uncertainty in terms of measurement error and extrapolation uncertainty (i.e., consideration of rare streamflow events). The second article of this dissertation assesses an innovative system for event-based streamflow measurements based on automated tracer injections at three contrasting sites in the Alps. In addition to regular mean flow measurements, the system was able to detect and measure rare high flow events at each site and its measurements revealed a change in the cross-section at one measurement site after a high flow event. Another conclusion is that regular maintenance and the experimental setup are crucial for the reliability of the system and the reduction of measurement errors.

Due to high non-linearity of hydrological systems, especially in high-elevation catchments, it is necessary to optimize and validate hydrological models in a process-based approach. The

third article describes the use of highly resolved spatiotemporal snow cover maps, derived from optical remote sensing images (i.e., Sentinel 2), to optimize and validate the snow redistribution routine of a distributed hydrological model. This area information fills a data gap left by in-situ snow measurements, which cannot convey the highly heterogeneous snow cover distribution. The optimized snow model was able to reproduce the spatiotemporal dynamic of snow cover in a high-elevation Alpine catchment reliably.

Physically-based and fully distributed hydrological models demand considerable computational resources in simulation runs over long timeperiodes (> 30 years) with high spatial (25 m) and temporal (1h) resolution. This limits the application of these models, even for mesoscale catchments (i.e., up to 100 km²). Article four presents a solution for reducing the required computational resources by using a machine learning approach that emulates high-resolution model results from a coarser resolution model setup (i.e., 100 m and daily). The results show a high reliability of the prediction (NSE=0.89) and a significant reduction of computational time by 93 %.

A key aspect in improving the models reliability is the increase of observational data availability as well as a thorough evaluation of data quality. Basing the mashine learning approach on a physics based reference model ensures the results remain within the models physical limits while increasing the overall computational efficiency. Hence, this dissertation advances the reliability and efficiency of hydrological modeling in high-elevation Alpine catchments.

Zusammenfassung

Gebirgsregionen versorgen tiefer gelegene Regionen mit Süßwasser auf der ganzen Welt, weshalb Berge oft als "Wasserschloss" bezeichnet werden. Mehr als die Hälfte der Weltbevölkerung ist von der Einzugsgebietsfunktion der Berge abhängig. Obwohl sie reich an Süßwasser sind, birgt die Nachfrage nach Wasserressourcen ein hohes Potenzial für Nutzungskonflikte (z. B. Trinkwasserverorgung, Wasserkraft, Bewässerung und Tourismus). Darüber hinaus hat der Klimawandel starke Auswirkungen auf die Hochgebirgsregionen mit dramatischen Folgen für die Kryosphäre (d.h. Gletscher, Schnee und Permafrost). Die europäischen Alpen werden als besonders gefährdet eingestuft und erleben seit den 1980er Jahren eine beschleunigte Erwärmung. Unzureichende Datenverfügbarkeit, Datenqualität und Prozessverständnis schränken jedoch die Abschätzung der Folgen für die Hydrologie in hochalpinen Gebieten nach wie vor ein. Diese Limitierungen haben einen erheblichen Einfluss auf die Anwendbarkeit hydrologischer Modelle, um eine räumlich und zeitlich hochaufgelöste Untersuchung möglicher klimabedingter Veränderungen auf die einzelnen Komponenten und die alpine Hydrologie als Ganzes zu ermöglichen. Diese Dissertation besteht aus vier Forschungsartikeln, die sich jeweils mit einem wichtigen Aspekt der hydrologischen Modellierung in hochalpinen Einzugsgebieten befassen und darauf abzielen, deren Zuverlässigkeit und Effizienz durch einen holistischen Ansatze zu erhöhen.

Da jede wissenschaftliche Studie, ob Zeitreihenanalyse oder Modellierung, von Beobachtungsdaten abhängt, betrachtet der erste Artikel die Qualität hydrometeorologischer Daten von zwei hochalpinen Einzugsgebieten und untersucht inwieweit die Datenqualität die Modellergebnisse beeinflusst. Wir haben festgestellt, dass Inhomogenitäten in den Temperatur- und Niederschlagszeitreihen, die das hydrologische Modell antreiben, einen direkten Einfluss auf Teile der Kryosphäre haben (d.h. auf die Akkumulation und Schmelze von Schnee und Gletschern). Sie beeinflussen jedoch die Abflussbildung nur indirekt, da die Abflussbildung Überlagerungseffekte eines nichtlinearen hydrologischen Systems beinhaltet.

Eine weitere Voraussetzung für eine umfassende Modellkalibrierung und -validierung sind kontinuierliche und zuverlässige Abflusszeitreihen mit einer geringen Unsicherheit in Bezug auf Messfehler und Extrapolationsunsicherheit (d.h. Berücksichtigung seltener Abflussereignisse). Im zweiten Artikel dieser Dissertation wird ein innovatives System für eventbasierte Abflussmessungen, das auf automatesierter Tracer-Injektion basiert, an drei unterschiedlichen Standorten in den Alpen getestet. Zusätzlich zu den regelmäßigen Messungen bei mittleren Abflussbedingungen war das System in der Lage seltene Hochwasserereignisse an jedem Standort zu erkennen und zu messen. Zudem haben die Messungen eine Veränderung des Gerinnequerschnitts an einem Messstandort nach einem Hochwasserereignis gezeigt. Eine weitere Schlussfolgerung ist, dass die regelmäßige

Wartung und der experimentelle Aufbau entscheidend für die Zuverlässigkeit des Systems und die Verringerung von Messfehlern sind.

Aufgrund der hohen Nichtlinearität hydrologischer Systeme, insbesondere in hochalpinen Einzugsgebieten, ist es notwendig, hydrologische Modelle mit einem prozessbasierten Ansatz zu kalibieren und zu validieren. Der dritte Artikel beschreibt die Nutzung räumlich und zeitlich hochaufgelöster Schneekarten, die aus optischen Fernerkundungsbildern (z. B. Sentinel 2) abgeleitet wurden, um die Schneeumverfrachtungsroutine eines flächendifferenzierten hydrologischen Modells zu optimieren und zu validieren. Mit diesen flächenhaften Schneeinformationen wird eine Datenlücke geschlossen, die auf der Limitierung von in-situ Schneemessungen die sehr heterogene Schneebedeckung wiederzugeben beruht. Das optimierte Schneemodell war in der Lage, die räumlich-zeitliche Dynamik der Schneedecke in einem hochalpinen Einzugsgebiet zuverlässig zu reproduzieren.

Physikalisch-basierte und flächendifferenzierte hydrologische Modelle benötigen bei Simulationsläufen über lange Zeiträume (> 30 Jahre) mit hoher räumlicher (25 m) und zeitlicher (1h) Auflösung erhebliche Rechenressourcen. Dies schränkt die Anwendung dieser Modelle selbst für mesoskalige Einzugsgebiete (d. h. bis zu 100 km²) ein. In Artikel vier wird eine Lösung zur Verringerung der erforderlichen Rechenressourcen vorgestellt, der auf einem maschinellen Lernenansatz basiert und hochaufgelöste Modellergebnisse aus einem gröber aufgelösten Modellsetup (d. h. 100 m und täglich) emuliert. Die Ergebnisse zeigen eine hohe Zuverlässigkeit der Prediktionen (NSE=0,89) und eine erhebliche Reduzierung der Rechenzeit um 93 %.

Ein Schlüsselaspekt für eine verbesserte Modellzuverlässigkeit ist eine erhöhte Verfügbarkeit von Beobachtungsdaten sowie eine gründliche Prüfung der Datenqualität. Die Verwendung eines physikalisch-basierten Referenzmodells als Grundlage für einen maschinellen Lernenansatz stellt sicher, dass die Ergebnisse innerhalb der physikalischen Grenzen des Refernenzmodells bleiben und gleichzeitig die Gesamtrechenzeit reduzierent wird. Diese Dissertation leistet somit einen wichtigen Beitrag, um die Zuverlässigkeit und Effizienz der hydrologischen Modellierung in hochgalpinen Einzugsgebieten zu verbessern.

Declaration of authorship

I hereby declare that the work presented in this Doctoral thesis is authentic and original unless clearly indicated otherwise, and in such instances full reference to the source is provided. I further declare that no unethical research practices were used. This dissertation was not submitted in the same or in a substantially similar version to another examination board.

Munich, 25 March 2023

Florentin Hofmeister

Acknowledgements

I would like to take this opportunity to thank the people without whom this work would not have been possible and who have accompanied me over the last four years. First and foremost, I would like to thank Gabriele Chiogna, who has contributed significantly to the success of this work. With his enthusiasm for science, he encouraged me to go beyond my own limits. At the same time, he has given me great confidence and freedom to develop myself further. Even after four years, I can still learn a lot from him. Moreover, he was always available for feedback or discussions no matter how difficult the circumstances were. I would like to thank Markus Disse for including me in his very pleasant team. The positive atmosphere at the chair contributed greatly to the successes of this dissertation and, besides the work, you could also have a lot of fun. Markus Disse also had a critical look at my work from a different, more engineering perspective. Additionally, I would like to thank Ralf Ludwig for taking on the role of external examiner of this dissertation. Our paths cross from time to time, and I hope they will also cross in the future.

Of the many colleagues at the chair, I would first like to thank Christiane Zach-Cretaine, who kept track of everything in the administrative jungle and reimbursed a million business trips and materials for the SEHAG project. Speaking about business trips, which would not have been possible without Michael Tarantik. With a lot of zest for action and good humor, we set off together on the numerous "SEHAG hikes" in the Alps and always returned safely. I would also like to thank the Augustenstrasse PhD-crew. I will try to list them: Soham Adla, Teresa Pérez Ciria, Leonardo Arias-Rodriguez, Francesca Perosa, Francesca Ziliotto, Moha Al-Qadi, Ala Al Dwairi, Mónica Basilio Hazas, Pablo Merchan Rivera, Sisay Simachew Mekonen, Beatrice Richieri, Daniel Bittner, Sonja Teschemacher, Lu Tian, Lucas Alcamo, Timo Schaffhauser, Fabian Merk, and Michael Neumayer. It was a very special time and a pleasure to have shared the office with you.

Gabe taught us the magic of turning salt into streamflow automatically. Although working from across the big pond, he has been in the field every year and has not been put off by the eerie mountains and outlandish people of Munchenland. In the very varied research activities, some students have also assisted and contributed significantly to the production of the publications, of which Alice Spadian, Brenda Rubens-Venegas, Floriana Graziano, and Thulasi Vishwanath-Harish deserve special mentions. In addition, I would like to thank all co-authors for the productive collaboration. Furthermore, I would like to thank the entire SEHAG research group for their cooperation over the past four years. The interdisciplinary exchange was very exciting and opened up completely new perspectives. I hope we will stay in contact in the future. The research group would not have been possible without funding, so I would like to thank the German Research Foundation (DFG) for funding the first and second SEHAG project phases.

Thanks also go to Uli Müller from the Zufallhütte, who never left us standing out in the rain and always provided us with the best food in the Alps.

I would like to thank my friends and family for their support over the years aloof from work and the dissertation. There will be more time again for the pleasant things in life. I would also like to thank John Peitz for checking the language of this thesis and the many discussions over the last years. Last but not least, I would like to thank my wife Franzi. The support and affection is hard to put into words. I am very happy to have you by my side and without you I would not have had the strength and persistence to complete this work. Driven by your positive nature, I have persevered even through the difficult times.

Research articles and author contributions

Presented research articles

The presented cumulative dissertation is based on four research articles. The first three articles are peer-reviewed and published in relevant international journals. The fourth article is a peer-reviewed proceeding article from the 39th IAHR World Congress with the title “from Snow to Sea” held in Granada (Spain) in 2022. Along with the third article, all generated data and scripts for the processing was published on the Mendeley Data platform and is open access (Hofmeister, Notarnicola, et al., 2022). The given author contributions are based on the CRediT (Contributor Roles Taxonomy) criteria (L. Allen, O’Connell, & Kiermer, 2019).

Article 1

Hofmeister, F., Graziano, F., Marcolini, G., Willems, W., Disse, M., & Chiogna, G. (2023). Quality assessment of hydrometeorological observational data and their influence on hydrological model results in Alpine catchments. *Hydrological Sciences Journal*, 02626667.2023.2172335. doi: 10.1080/02626667.2023.2172335.

- 1 of 3 required first author publications
- Chapter 2
- Author contributions are based on CRediT statements:
    - **Florentin Hofmeister:** Conceptualization, Methodology, Software, Formal analysis, Data Curation, Visualization, Writing - Original Draft.
    - **Floriana Graziano:** Software, Formal analysis, Data Curation.
    - **Giorgia Marcolini:** Conceptualization, Methodology, Writing - Review & Editing.
    - **Winfried Willems:** Conceptualization, Methodology, Writing - Review & Editing.
    - **Markus Disse:** Writing - Review & Editing, Funding acquisition.
    - **Gabriele Chiogna:** Conceptualization, Methodology, Writing - Review & Editing, Funding acquisition.

Article 2

Hofmeister, F., Rubens Venegas, B., Sentlinger, G., Tarantik, M., Blume, T., Disse, M., & Chiogna, G. (2023). Automated streamflow measurements in high-elevation Alpine catchments. *River Research and Applications*, rra.4203. doi: 10.1002/rra.4203.

- 2 of 3 required first author publications
- Chapter 3

- Author contributions are based on CRediT statements:
    - **Florentin Hofmeister:** Conceptualization, Methodology, Software, Investigation, Formal analysis, Data Curation, Visualization, Writing - Original Draft.
    - **Brenda Rubens Venegas:** Software, Investigation, Formal analysis, Data Curation, Visualization.
    - **Gabe Sentilinger**: Software, Investigation, Formal analysis, Data Curation.
    - **Michael Tarantik:** Investigation.
    - **Theresa Blume:** Writing - Review & Editing, Funding acquisition.
    - **Markus Disse:** Writing - Review & Editing, Funding acquisition.
    - **Gabriele Chiogna:** Conceptualization, Methodology, Writing - Review & Editing, Funding acquisition.

Article 3

Hofmeister, F., Arias-Rodriguez, L. F., Premier, V., Marin, C., Notarnicola, C., Disse, M., & Chiogna, G. (2022). Intercomparison of Sentinel-2 and modelled snow cover maps in a high-elevation Alpine catchment. *Journal of Hydrology X*, *15*, 100123. doi: 10.1016/j.hydroa.2022.100123.

- 3 of 3 required first author publications
- Chapter 4
- Author contributions are based on CRediT:
    - **Florentin Hofmeister:** Conceptualization, Methodology, Software, Formal analysis, Data Curation, Visualization, Writing - Original Draft.
    - **Leonardo F. Arias-Rodriguez:** Methodology, Software, Data Curation, Formal analysis.
    - **Valentina Premier:** Software, Data Curation, Formal analysis.
    - **Carlo Marin:** Conceptualization, Writing - Review & Editing.
    - **Claudia Notarnicola:** Conceptualization, Writing - Review & Editing.
    - **Markus Disse:** Conceptualization, Writing - Review & Editing, Funding acquisition.
    - **Gabriele Chiogna:** Conceptualization, Writing - Review & Editing, Supervision, Funding acquisition.

Article 4

Hofmeister, F., Spadina, A., & Chiogna, G. (2022). Coupling Support Vector Machine and physically-based Hydrological Modeling for Reducing the Computational Time in Climate Change Studies. *Proceedings of the 39th IAHR World Congress*. Presented at the 39th IAHR WORLD CONGRESS, Granada, Spain. Granada, Spain. doi: //10.3850/IAHR-39WC2521716X2022902.

- Chapter 5
- Author contributions are based on CRediT:
    - **Florentin Hofmeister:** Conceptualization, Software, Methodology, Formal analysis, Visualization, Writing - Original Draft.
    - **Alice Spadina:** Software, Formal analysis, Validation, Visualization.
    - **Gabriele Chiogna:** Conceptualization, Writing - Review & Editing, Supervision, Funding acquisition.

Further scientific contributions

Research articles

Altmann, M., Piermattei, L., Haas, F., Heckmann, T., Fleischer, F., Rom, J., Betz-Nutz, S., Knoflach, B., Müller, S., Ramskogler, K., Pfeiffer, M., **Hofmeister, F.,** Ressl, C. & Becht, M. (2020). Long-Term Changes of Morphodynamics on Little Ice Age Lateral Moraines and the Resulting Sediment Transfer into Mountain Streams in the Upper Kauner Valley, Austria. *Water*, *12*(12), 3375. doi: 10.3390/w12123375.

Knoflach, B., Ramskogler, K., Talluto, M., **Hofmeister, F.,** Haas, F., Heckmann, T., … Stötter, J. (2021). Modelling of Vegetation Dynamics from Satellite Time Series to Determine Proglacial Primary Succession in the Course of Global Warming—A Case Study in the Upper Martell Valley (Eastern Italian Alps). *Remote Sensing*, *13*(21), 4450. doi: 10.3390/rs13214450.

Ramskogler, K., Knoflach, B., Elsner, B., Erschbamer, B., Haas, F., Heckmann, T., **Hofmeister, F,** … Tasser, E. (2023). Primary succession and its driving variables – a sphere-spanning approach applied in proglacial areas in the upper Martell Valley (Eastern Italian Alps). Biogeosciences, 20(14), 2919–2939. doi: 10.5194/bg-20-2919-2023.

Rom, J., Haas, F., **Hofmeister, F.,** Fleischer, F., Altmann, M., Pfeiffer, M., … Becht, M. (2023). Analysing the Large-Scale Debris Flow Event in July 2022 in Horlachtal, Austria Using Remote Sensing and Measurement Data. *Geosciences*, *13*(4), 100. doi: 10.3390/geosciences13040100.

Oral presentations

Hofmeister, F., Disse, M. & Chiogna, G. (2019). SEHAG Teilprojekt 2. Klimawandelbedingte Veränderung hydrologischer Prozesse in hochalpinen Einzugsgebieten. *3. Workshop zur Alpinen Hydrologie*. Obergurgl (Austria).

Hofmeister, F., Arias-Rodriguez, L.F., Premier, V., Marin, C., Notarnicola, C., Disse, M. & Chiogna, G. (2021). Generation of a high-resolution snow cover dataset from Sentinel-2 images for snow model calibration. *EGU General Assembly 2021*. Vienna (Austria).

Hofmeister, F., Arias-Rodriguez, L.F., Premier, V., Marin, C., Notarnicola, C., Disse, M. & Chiogna, G. (2022). Validation of snow redistribution dynamics in hydrological modeling using optical remote sensing products. *SnowHydro2022*. Grenoble (France).

Hofmeister, F., Arias-Rodriguez, L.F., Premier, V., Marin, C., Notarnicola, C., Disse, M. & Chiogna, G. (2022). Vergleich verschiedener Sentinel-2 und modellierter Schneebedeckungskarten in einem hochalpinen Einzugsgebiet. *4. Workshop zur Alpinen Hydrologie*. Obergurgl (Austria).

Hofmeister, F., Rubens Venegas, B., Sentlinger, G., Disse, M. & Chiogna, G. (2022). Automated discharge measurements with salt dilution in Alpine creeks and uncertainty quantification. *EGU General Assembly 2022*. Vienna (Austria).

Hofmeister, F., Spandina, A. & Chiogna, G. (2022). Coupling Support Vector Machine and Physically-based Hydrological Modeling for Reducing the Computational Time in Climate Change Studies. *39th IAHR World Congress*. Granada (Spain).

Posters

Chiogna, G., **Hofmeister, F.,** Pfeiffer, M., Labuhn, I., Marzeion, B. & Disse, M. (2022). Limitierungen in der Modellierung historischer Flutereignisse aufgrund von Gletscherseeausbrüchen in einem alpinen Einzugsgebiet. *TDH 2022.* Munich (Germany).

Schaffhauser, T., Chiogna, G., Disse, M. & **Hofmeister, F**. (2022). SWAT-G, a New Glacier Routine for the Hydrological Model SWAT. *AGU Fall Meeting 2022*. Chicago (USA).

List of figures

List of tables

1. Introduction

1.1. Motivation

Mountainous regions supply a substantial part of fresh water to environmental and human water demands all over the globe. More than half of the world's population depends on the watershed function of mountains (Immerzeel et al., 2020). Due to their orography, mountains usually have a higher specific discharge (ratio of catchment area and discharge) than lower elevated areas and are therefore called the 'water towers' of the world. In addition, runoff generation in mountainous regions shows a pronounced seasonality due to summer melting of snow and glaciers. The meltwater from snow and glaciers provides a relatively constant supply of freshwater to the downstream areas over the warm summer months, which are otherwise very dry in some regions of the world (e.g., Central Asia) (Barandun et al., 2020, 2021). In Europe, the Alps are the most relevant water supplying mountain range (Immerzeel et al., 2020). Despite the high availability of freshwater, socio-economic conflicts arise over water use, e.g., between agriculture and hydropower or between hydropower and tourism (Beniston, 2012). The water use conflict is further intensified by the production of machine-made snow in Alpine areas with intensive ski tourism (Grünewald & Wolfsperger, 2019). The water demand for snow-making can surpass 50 % of the communal water consumption during the winter season (Vanham, Fleischhacker, & Rauch, 2008).

Global warming of the climate system is unequivocal and has led to an exceptionally high temperature increase of about +2 °C in the Alps between the 19th and early 21st century, which is more than twice the rate of the average warming of the Northern hemisphere. Therefore, climate change makes the Alps especially vulnerable and intensifies the already existing water use conflicts (European Environment Agency, 2009). However, Tudoroiu et al. (2016) detected a negative elevation-dependent warming in the Eastern Italian Alps for the period 1975 to 2010. Further studies confirm similar patterns of elevation-dependent warming trends in the Alpine region (Kuhn & Olefs, 2020; Rottler, Kormann, Francke, & Bronstert, 2019). In addition, the warming of the Alpine region was not linear in the 20th century. Battaglia & Protopapas (2012) observed a linear warming trend of 0.01°C per year from the beginning of the 20th centuray, but around 1980 the warming trend increased dramatically, exceeding the rate of 0.05°C per year. There is still no evidence of a decline in the warming trend (Battaglia & Protopapas, 2012). The impact of climate change on precipitation changes is not as clear as it is on the temperature increase in the Alps. Brunetti et al. (2006) detected a slight trend towards an increase in the north Alpine region and a significant decrease in the south over the last two centuries by analyzing 192 homogenized station series. Additionally, they noticed a regular and smooth signal toward an increase in winter and spring precipitation. The investigation of

climate change impact on the precipitation is further complicated by feedback from the North Atlantic Oscillation (NAO), which does not affect all Alpine regions equally (Brunetti et al., 2006). Brugnara & Maugeri (2019) analyzed 18 homogenized daily precipitation series from the southern European Alps and detect a negative trend in precipitation frequency over the period 1890-2017. This trend is related to a step-like reduction of cyclonic weather types over central Europe that occurred around 1940. Strong trends on a decadal scale are related not only to the NAO but also to the Atlantic Multidecadal Oscillation (AMO), although the influence of the NAO is present only in recent years. Contrarily, trends in heavy precipitation indices do not show a coherent pattern in the Alps (Brugnara & Maugeri, 2019). Scherrer et al. (2016) found an increase in precipitation intensity (i.e., exceedance of the 99th percentile) in Swiss observations over the years 1901-2014/15. Besides the intensity, the seasonal occurrence of moderate and extreme precipitation events in the Alps shifted to cooler months (i.e., early summer and autumn) (Brönnimann et al., 2018).

The impact of climate change on snow is much more explicit in the Alps than for precipitation. Snow is a major driver of Alpine hydrology as it stores precipitation in solid form over the winter period and releases the water in spring/early summer (Matiu et al., 2021). The amount and duration of snow in the Alps has a high economic significance in terms of tourism and hydropower (Marty, 2008). Multiple studies depict a clear negative trend in snow days due to a sudden climate change induced temperature increase at the end of the 1980's (Marty, 2008; Notarnicola, 2022; Olefs, Koch, Schöner, & Marke, 2020). Matiu et al. (2021) detected negative trends in monthly snow depth records between 1971 and 2019 in an Alpine-wide analysis form most stations from November to May. However, it is not possible to generalize the negative trends because the trends vary strongly by region (i.e., climatic forcing zone) and elevation (Matiu et al., 2021).

The Intergovernmental Panel on Climate Change Ipcc (2022) has very high confidence that the mass losses of the glaciers and increased permafrost temperature over the last decades are caused by global warming. The latest report lists, with high confidence, that the reduction in snow cover (i.e., duration, extent and depth), is also a consequence of global warming (Ipcc, 2022). Depending on the emission scenarios, it is projected that the glacier mass of regions with mostly small glaciers (e.g. Central Europe) could decline by more than 80 % of their current ice mass by 2100 under the RCP8.5 scenario (medium confidence), and many glaciers are projected to disappear regardless of future emissions (very high confidence). Regardless of the emissions scenario, snow depth at lower elevations in high mountain regions is likely to decrease by about 10-40 % (high confidence) by 2031-2050, compared to 1986-2005. The projected decrease is likely about 50-90 % for RCP8.5 scenario for 2081-2100. The runoff generation in snow-dominated and glacier-fed high mountain catchments is projected to

change regardless of the emission scenario (very high confidence), resulting in an increase of average winter runoff (high confidence) and earlier spring peaks (very high confidence). In regions with little glacier cover (e.g., European Alps) most glaciers have already passed the peak runoff (i.e., 'peak water'), after which it declines (high confidence) (Ipcc, 2022). Even if the temperature increase is limited to +1.5°C, about half of the world's glaciers will disappear by 2100 (Rounce et al., 2023).

Based on these facts, several conclusions can be drawn: i) meteorological observations with long temporal coverage show a clear warming trend in the Alpine region, which directly propagates into the cryosphere (snow, glaciers, and permafrost). ii) climate change intensifies already existing conflicts over the use of water as resources in the Alps. iii) the projected climate-induced changes in the cryosphere (e.g., recession of glacier and decrease in snow) have a high confidence, especially projections into the first half of the 21st century.

However, preavious research clearly indicates that a generalization of the climate-induced changes in the Alpine region is not possible. It is worth to mention that the detected changes dependent on: i) the analyzed variable, i.e., changes in precipitation frequency and intensity are less confident as changes in temperature. ii) spatial scale, changes can be regional or local. iii) elevation, changes can manifest differently depending on the altitude. iiii) temporal scale, changes can be detected in climate periods, decadal periods, annual, seasonal, daily, hourly or sub-hourly intervals.

A further aspect that complicates the study of climate-induced changes in the Alpine region is the complex feedback between the individual components of the geosystem (e.g., hydrosphere and cryosphere). Moreover, anthropogenic changes in the geosystem (e.g., deforestation, reduction of aerosol emission, ski tourism) can lead to a superposition of supposed climate change signals (Kuhn & Olefs, 2020; Philipona, 2013; Rottler et al., 2019; Tudoroiu et al., 2016). Therefore, all individual components of the geosystem, such as the Alpine hydrosphere, have to be considered in a holistic approach to identify and quantify climate change-induced changes on each individual component.

1.2. Current state of research in the field of Alpine hydrology

Before introducing the main research hypothesis and research questions, the current state of research regarding the four main topics of this dissertation (i.e., time series analysis, automated streamflow measurements, monitoring and modeling of snow cover, and coupling hydrological models with machine learning) will be summarized briefly. The current state of research is essential for identifying research gaps on which the respective research questions are based.

1.2.1. Time series analysis

Hydrometeorological observation data are essential for the understanding of hydrological systems and the analysis of potential changes whether natural or anthropogenic. In some areas of the Alpine region, hydrometeorological observations already started in the second half of the 18th century. Although these records are a valuable treasure, acquiring these datasets is very time-consuming due to frequent changes in the geopolitical situation in the Alpine region over the last centuries, which have caused the records to be located in various different archives. More and more of these records have now been digitalized and published (Auer et al., 2007). The HISTALP project collected and performed quality controll on a unique dataset of 557 meteorological time series with a monthly resolution. The quality control ensures that changes in the measurement record are not falsely detected as climate change signal in long climate time series, as changes mainly occur due to changes in relocation, surroundings of sites, instruments, time of observations and observers. The first step of homogenization is the detection of potential break points in a time series. Metadata from the stations can give hints for the existence and reasons of possible inhomogeneities (i.e., break points) in time series (Auer et al., 2007). The homogeneity of a time series can be evaluated with two procedures, either a relative homogeneity test, in which a time series is compared to neighboring series (Auer et al., 2005), or an absolute homogeneity test that detects possible inhomogeneities within a time series. The result of relative homogeneity tests is highly dependent on the quality of the reference series. Consequently, the relative homogeneity tests should always include an absolute test of the reference series (Yozgatligil & Yazici, 2016). Additionally, the selection of a minimum coefficient of variation of the ratio between two series controls the reliability of test results (Auer et al., 2005).

While the focus of break detection and homogenization has mainly focused on temperature and precipitation series, recent studies have been focusing on snow depth (Buchmann et al., 2022; Marcolini et al., 2019a; Resch et al., 2022; Schöner, Koch, Matulla, Marty, & Tilg, 2019). Li, Yan, Zhu, Freychet, & Tett (2020) generated the first homogenized daily data set of relative humidity for China over the period 1960-2017 by means of the MASH method, which stands for Multiple Analysis of Series for Homogenization (Szentimrey, 1999). Most (i.e., 92 %) of the 746 relative humidity series were inhomogeneous, containingat least one break point. Break points mainly occurred due to the implementation of automated observation across China in the early 2000's (Li et al., 2020). Homogenized daily solar radiation records are available for Spain (Sanchez-Lorenzo, Calbó, & Wild, 2013), Italy (Manara et al., 2016) and China (Yang, Wang, & Wild, 2018). The homogenized radiation data enables the detection of dimming and brightening in the atmosphere due to anthropogenic aerosol emissions (Manara et al., 2016). Lucio-Eceiza, González-Rouco, Navarro, Beltrami, & Conte (2018) checked the quality of 526 wind speeds records from across northeastern North America. About 16 % of those wind speed

records had to be corrected, mainly due to unrealistic low wind speeds (Lucio-Eceiza et al., 2018). Chen, Saunders, & Whan (2021) developed a workflow for quality control and bias adjustment of crowdsourced wind speed observations. Volchak, Sidak, & Parfomuk (2021) applied parametric (i.e., Buishand Range Test and Standard Normal Homogeneity Test) and non-parametric tests (i.e., Pettitt's Test) for the detection of inhomogeneities in annual discharge time series across Belarus. However, the test results generated with parametric tests need to be interpreted with caution, since parametric tests actually assume normally distributed data without autocorrelation (i.e., serial dependency) between adjacent members of the series (Volchak et al., 2021) and this is not the case for runoff time series. Furthermore, serial dependency, also called persistence, can be induced by the sampling of data. High-resolution time series often have stronger persistence than low resolution time series (Mudelsee, 2010). Zanchettin, Traverso, & Tomasino (2008) checked an extended discharge time series of the Po River for inhomogeneities through basic trend and change point analyses and comparison with approximated catchment-averages of precipitation and evapotranspiration for the period 1831–2003. They found concurrent changes in the persistence (i.e., autocorrelation) of precipitation and discharge data. The analysis of historic discharge time series must also be conducted with caution because stage-discharge conversion may lead to considerable uncertainty in the discharge estimation even when historic stage measurements are accurate and homogeneous (Zanchettin et al., 2008). The non-stationarity of the stage-discharge relation considerably affects the quality of the derived discharge time series and creates uncertainty for interpretation of climate-induced changes on hydrological regime of a river (Farsi, Mahjouri, & Ghasemi, 2020).

1.2.2. Discharge measurements and rating curves for Alpine creeks

The hydrological response of small headwater catchments in the Alpine region can be very heterogonous due to differences in the degree of glaciation, topography (e.g., orographic precipitation), geology and soil (Meißl et al., 2021; Penna et al., 2014; Penna, van Meerveld, Zuecco, Dalla Fontana, & Borga, 2016; Penna et al., 2017). However, stream gauges are usually located in the main valleys or at the inflow of reservoirs. This is often the case because the operation of gauging stations in mountainous regions is associated with a high expenditure of time and money. Especially due to the high sediment and coarse bedload transport of glacio-nival rivers (Comiti et al., 2019), constant maintenance of the gauges is necessary and damages caused by extreme discharge events have to be expected. Besides the high turbulences impeding hydrometric measurements, the cross-sections are difficult to define and are constantly changing (Leibundgut, Maloszewski, & Külls, 2011; Morgenschweis, 2018). In addition, creeks in high alpine areas are often located in protected areas, which makes it difficult to set up gauging stations, or they are located in inaccessible areas. Nevertheless, the operation of additional stream gauges enables the investigation of the hydrological response

of small tributaries and is an additional validation source for quality checks (i.e., relative homogeneity tests).

Tracer-based discharge measurements have proven successful for decades in alpine rivers and creeks (Allen & Taylor, 1923; Moore, 2004; Østrem, 1964) when the application of "conventional" velocity-area methods (e.g., current meter or Acoustic Doppler Current Profiler (ADCP)) has not been possible due to difficult meansuring conditions such as strong turbulence, unknown cross-section geometry, and high fluctuation of water level (Morgenschweis, 2018; Richardson, Sentlinger, Moore, & Zimmermann, 2017). Noncontact methods based on remote sensing of water surface velocity using Doppler radars have been developed over the past two decades (Son, Kim, Kim, & Roh, 2023; Welber et al., 2016). Surface velocity radars can either be mounted permanently or used in a portable configuration. Welber et al. (2016) extensively tested a portable surface velocity radar in a field study on rivers and creeks with different characteristics in terms of river width, roughness and turbulences. They concluded that surface velocity radars can produce results with higher uncertainty than conventional intrusive methods because of the limited knowledge of bed changes and the value of the velocity coefficient, which represents the ratio between depth-average and surface velocity. Moreover, the velocity coefficient showed large fluctuations if computed on individual verticals, especially in high roughness channels (Welber et al., 2016). Hauet, Morlot, & Daubagnan (2018) analyzed a very large sample set (i.e., > 3600) of gaugings and determined velocity coefficients for natural and artificial channels. Nevertheless, the use of the surface velocity method requires knowledge about the geometry of the cross-section. This is not the case for tracer-based flow measurements. Therefore, water-soluble hydrological tracers are the predominant flow measurement method in alpine creeks and rivers and can be subdivided into three groups: i) salt and other chemical substances, ii) dyes, especially fluorescent dyes and iii) radioactive tracers. Selection of a tracer type depends mainly on the required tracer properties. Although radioactive tracers have nearly perfect tracer proparties, they are rarly usered in hydrologic studies because strict radiation safty regulation make their use impossible. In surface waters, salt is usually used as a tracer because salt concentration can be easily determined via the electrical conductivity (EC) as both have a linear relation. Sodium Chloride (NaCl) is the most commonly used salt tracer since it is relatively cheap, available almost everywhere, and highly soluble in water (Morgenschweis, 2018). Moreover, salt dilution is very accurate with a measurement error about 5 % if the key assumptions (e.g., no salt loss during the measurement) and requirements (e.g., the salt needs to be fully mixed across the channel at the point where EC measurements are taken) are met (Richardson, Moore, & Zimmermann, 2017; Richardson, Sentlinger, et al., 2017). Hauet (2020) developed a first complete framework for uncertainty quantification of salt discharge measurement following the GUM (Guide to the expression of uncertainty in measurement) method that take

into account all error sources. In general, salt dilution gauging is applicable to a wide range of stream sizes, from small first-order streams (Q from 1 to 10 l/s) to large rivers (Q from 100 to 1000 m³/s) (González-Pinzón, Haggerty, & Dentz, 2013; Richardson, Moore, et al., 2017).

It can be difficult to obtain a full range of flow measurements for a hydrometric rating curve at remote locations of rivers in the Alps. Field visits depend on weather and snow conditions, thus, sites are only visited at times when these allow for safe access to the location, which in turn often coincides with an environmental situation that displays similar flow conditions at each visit. Flow measurements during peak flow events can be especially dangerous (Sentlinger, Fraser, & Baddock, 2019). These rare peak events are particularly important for the construction of robust and consistent rating curves with a reduced extrapolation uncertainty. The rating curve is based on the stage-discharge relation and is an indirect measurement method to generate a continuous discharge time series. In the case of natural measurement cross-sections in unconsolidated sediments, measurement frequency must be increased to 10 to 12 measurements per year, in order to validate the rating curve (Morgenschweis, 2018). Sentlinger et al. (2017) developed the first automated salt dilution system (AutoSalt) that can achieve acceptable results (< 7 % uncertainty) using as little as 100g per m³/s of flow by increasing the Signal to Noise Ratio (SNR). An optimized injection dose minimizes potential for detrimental effects on the ecology (Wood & Dykes, 2002) and enables more measurements per tank filling. The AutoSalt system has already been tested on various smaller streams around the world, but discharges were generally in the range of 1 to 10 m³/s. Using the AutoSalt, a rating curve can be established in a relativly short timeperiod (1 year) and changes in the cross-section can be quickly detected. Due to the extensive ascertainment of the measurement error of the system, the error propagation can be taken into account when constructing rating curves (Di Baldassarre & Montanari, 2009a; Kiang et al., 2018). Nevertheless, discharge data retains a measure of uncertainty and therefore it is useful to monitor other hydrological variables that significantly influence runoff generation in high-elevation Alpine catchments, such as the snow component.

1.2.3. Modeling and monitoring spatial snow distribution

Since the discharge of an Alpine catchment is an integrated information of different runoff sources (e.g., precipitation, snow and glacier melt) and storages (e.g., glacier, soil and geology), a growing number of hydrological modeling studies also consider spatial snow cover maps for constraining model parameters and validation of the model results (Hanzer, Helfricht, Marke, & Strasser, 2016; Thornton, Brauchli, Mariethoz, & Brunner, 2021; Vionnet et al., 2021; Warscher et al., 2013). In contrast to the snow depth measurement, which gives only point information, snow cover maps derived from high-resolution optical satellites (e.g., Sentinel 2, Landsat) can capture the dynamical spatio-temporal snow distribution in such complex terrain

(Gascoin, Grizonnet, Bouchet, Salgues, & Hagolle, 2019). With a much coarser spatial resolution (250 m), the MODIS sensor has been providing daily snow cover maps since 2000 (Matiu, Jacob, & Notarnicola, 2019). This valuable data set allows the investigation of global trends in snow cover duration (Notarnicola, 2020, 2022). However, MODIS is not able to capture the spatial heterogeneity of snow cover induced by solar radiation in mountainous regions (Bouamri et al., 2021). Another valuable information that can be derived from snow cover maps are the snow line altitudes (SLAs), which are mainly controlled by temperature, precipitation, and solar radiation. Although SLA and snow-covered area are complementary metrics, the SLA has a strong potential for understanding local-scale snow cover dynamics and their controlling mechanisms (Girona- Mata, Miles, Ragettli, & Pellicciotti, 2019). Additionally, the SLA is an indicator of the equilibrium line altitude (ELA), which is a key parameter for calculating and assessing glacier mass balance (Li, Wang, & Wu, 2022). Hence, the end-of-season SLA is needed for hydrological applications and for understanding the response of glaciers to climate variability (Racoviteanu, Rittger, & Armstrong, 2019).

Physically-based hydrological models are typically used to predict snow accumulation and melt in mountainous regions and to project the effects of climate change on the runoff generation from snow (Krogh et al., 2022; Strasser et al., 2019) and glacier (Beniston et al., 2018; Huss et al., 2017; Huss, Jouvet, Farinotti, & Bauder, 2010; Rounce et al., 2023). The spatial variably of the snowpack controls the timing and magnitude of the snowmelt runoff. A homogenous snowpack generally results in a higher snowmelt peak than a highly heterogeneously distributed snowpack since snowmelt rates have large spatial differences (Freudiger, Kohn, Seibert, Stahl, & Weiler, 2017). Depending on the complexity of the hydrological model, snow redistribution due to wind and gravitational slides (i.e., avalanches) are considered in the snow routine. With higher model complexity (e.g., Alpine3D model developed by Lehning et al. (2006)), more input data (e.g., local wind fields) are needed, but these are not available for every area or period. Therefore, snow redistribution routines are often based on empirical assumptions in the hydrological models. Generally, the consideration of snow redistribution has improved the simulation of snow patterns and snow water equivalent (SWE) and consequently the prediction of discharge in mountainous catchments (Freudiger et al., 2017). The almost continuous snow observation products derived from satellite images allow the snow distribution routines to be validated not only for individual images of the year (Warscher et al., 2013), but also for all seasons (Thornton et al., 2021) and even for multiple years (Hofmeister, Arias-Rodriguez, et al., 2022). Painter, Gentile, & Ferraris (2023) calibrated a fine-scale Cellular Automaton model that describes snow coverage across a high-elevation catchment via a randomized parameter search, fitting simulation data against snow cover masks estimated from Sentiel-2 satellite images.

However, a major limitation of the snow cover maps is the fact that they provide only binary information (i.e., snow and no snow). With complex procedures, areal snow depths can be generated by means of stereoscopic or active optical (e.g., LIDAR) sensors for individual days and specific areas or regions. Based on active microwave sensors (e.g., ALOS 2, Sentinel 1, RADARSAT-2, TerraSAR-X), spatial snow depth and snow wetness information can be derived with a relative high spatial resolution (up to 1 m). However, the oblique geometry viewing of SAR systems amplifies geometric distortions, making interpretation in mountainous regions particularly difficult. Areal SWE by means of passive microwave are limited by the coarse spatial resolution (around 25 km) and interference of the microwave signal from liquid water content in the snow pack at the beginning of the melting period (Largeron et al., 2020).

Remotely-sensed data is the most natural source of information, as opposed to extrapolated insitu snow observations, to constrain snow cover simulations, especially in data scarce regions (Largeron et al., 2020). Thus, epistemic uncertainty in the model results can be reduced by incorporating snow information from remote sensing products into a multi-objective calibration and validation approach. Tiel, Stahl, Freudiger, & Seibert (2020) concluded in a comprehensive literature review that in about 50 % of glacio-hydrological models have been calibrated and validated using a multi objective approach. In recent years research has increased its focus on data assimilation of remote sensing products into physically-based models (Alonso-González et al., 2022; Deschamps- Berger et al., 2022; ESA, 2023a). The combination of snow observations and physically-based snow modeling enables not only accurate probabilistic SWE prediction, for instance for runoff and flow predictions, but also a better knowledge of physical snowpack properties (Largeron et al., 2020). A disadvantage of the high complexity (i.e., consideration of more processes) of physically-based models with high spatiotemporal resolution are the increasing computational demands, which is a limitation especially for long simulation periods in climate change studies.

1.2.4. Coupling of physically-based hydrological model and machine learning

Physically-based models mainly construct a simplified watershed system and express the interior behavior through solving mathematical equations that represent the hydrological process to the best current knowledge. In contrast, data-driven (e.g., machine learning) models establish a direct mapping between hydrological variables and extract their relationship from historical measured data by the algorithms developed in the fields of statistics, computational intelligence, machine learning, and data mining (Young, Liu, & Wu, 2017). Therefore, data-driven models are highly flexible in adapting data and are amenable to finding unexpected patterns (Reichstein et al., 2019). In data scarce regions, physically-based modeling confronts a much higher challenge to reach the required accuracy (Young et al., 2017; Yuan & Forshay,

2022) However, the predictions from data-driven models may be physically inconsistent or implausible, owing to extrapolation or observational biases, for example (Reichstein et al., 2019).

Historically, physical modelling and machine learning have often been treated as two different fields with very different scientific paradigms (i.e., theory-driven versus data-driven) (Reichstein et al., 2019). In recent years, the coupling of physically-based hydrological models and machine learning methods has become more and more popular in hydrological studies investigating and predicting nonlinear and non-stationary hydrological behavior of complex systems (Yuan & Forshay, 2022). The reasons for the coupling are very diverse. The integration of machine learning methods can reduce the risk of overfitting of the model parameters and enhance the generalization of the models, especially in the case of Support Vector Machines (SVM) (Smola & Schölkopf, 2004; Young et al., 2017). Moreover, the important hydrological processes involved in a physically-based model can compensate the black-box nature of data driven models (Young et al., 2017). This enables direct interpretation of the results as well as the potential of extrapolation beyond observed conditions (Reichstein et al., 2019). The powerful data-driven methodology can alleviate the difficulty in accurate physical modeling. In general, the two methods,based upon different philosophies, complement each other with respect to their inherent strengths and limitations (Young et al., 2017). Reichstein et al. (2019) defined five points of potential synergies between physically-based and machine learning models:

1. **Improving parametrizations.** Machine learning can lean parametrizations to optimally describe the ground truth that can be observed or generated from detailed and high-resolution models through first principles.
2. **Replacing a "physical" sub-model with a machine learning model.** Semi-empirical sub-models can be replaced by a machine learning model if a sufficient number of observations are available. This leads to a hybrid model, which combines the strengths of physically modeling (theoretical foundations, interpretable compartments) and machine leaning (data-adaptiveness).
3. **Analysis of model-observation mismatch.** Machine learning can help to identify, visualize and understand the patterns of model error, which allows us also to correct model outputs accordingly. This helps to improve the physical model and theory.
4. **Constraining sub-models.** The replacement of another (potentially biased) sub-model in an offline simulation with the output from a machine learning algorithm can help to disentangle model error, which originates from the sub-model of interest from errors of coupled sub-models. This simplifies and reduces biases and uncertainties in model parameter calibration or the assimilation of observed system state variables.

5. **Surrogate modeling or emulation.** Emulation of the full or specific parts of a physical model can be useful for computational efficiency and tractability. Trained machine learning emulators can achieve simulation results orders of magnitude faster than the original physical model without scarifying much accuracy. This allows for fast sensitivity analysis, model parameter calibration, and derivation of confidence intervals for the estimates.

Another clear advantage of the coupling is the high flexibility of the application. Chiogna, Marcolini, Liu, Pérez Ciria, & Tuo (2018) coupled the Soil and Water Assessment Tool (SWAT) model and Support Vector Regression (SVR) method to predict hydropeaking for the Upper Adige watershed in northeast Italy and applied a wavelet method to analyze the price of energy. Yuan & Forshay (2022) developed a new hybrid model that integrates the SWAT model with a SVR calibration method coupled with discrete wavelet transforms (DWT) to better support modeling watersheds with limited data availability. Young et al. (2017) developed a hybrid model, which integrates the physically-based HEC-HMS model into the SVR for accurate rainfall-runoff modeling with an application in the Chishan Creek basin in southern Taiwan. Huang et al. (2022) proposed a hybrid framework by taking the advantages of interpretability of physical hydrodynamic modeling and the adaptability of machine learning to improve river modeling with complex boundary conditions (i.e., lake-connected river system).

1.3. Hypothesis and research questions

Based on the current state of research and the identified research gaps in the field of Alpine hydrology, the main research hypothesis of this dissertation was developed:

A holistic approach of high-resolution spatio-temporal data, comprehensive quality assessment of observation data and machine learning constitute an advantage for reliable and efficient process-based hydrological modeling in high-elevation Alpine catchments.

In general, the research hypothesis is based on the fact that, on one hand, additional observation data is needed to reduce data scarcity to better describe and understand spatio-temporal variability of snow hydrological processes in complex terrains, and, on the other, to consider the quality of already existing observation data before they are used in a hydrological model application. The collecting of additional high-resolution data includes the use of a inovative automated tracer injection system to collect event-based streamflow data and the generation of continuous and high-resolution snow cover maps from Sentinel-2 images. Consequently, a reliable and efficient hydrological model for high-elevation catchments can be achieved by four aspects: i) forcing with quality controlled observation data, ii) calibration with reliable streamflow data iii) process-based multi-objective calibration including snow cover

information, and iiii) the coupling with machine learning to reduce the computation time. In other words, the uncertainty in the hydrological model results can be reduced because the model is forced by homogeneous observation data and is able to reproduce the spatio-temporal variability of snow hydrological processes. In addition, since the computation time of high-resolution (i.e., in space and time) physically-based models can be very high, the efficiency was increased by coupling it with a machine learning algorithm.

The main research hypothesis is subdivided into four research questions correspondent to the four research articles. Since this dissertation addresses many different keywords of the field of Alpine hydrology (Table 1-1), they are roughly subdivided into three main subgroups (namely *research area*, *observational data*, and *hydrological models*) to better highlight the links between the individual articles. The group of *research area* includes *Alpine catchments,* where all four studies are located, and *spatio-temporal complexity* of snow hydrological processes, which is addressed in all articles. In general, all articles discuss the difficulties of observing or modeling snow hydrological processes in the complex topography of high-elevation areas from the point to the catchment scale.

The category of *observational data* is divided into the sub-categories of *data scarcity, data quality, input uncertainty, and uncertainty in the calibration objective*. *Data scarcity* is the main limitation and challenge for studies in high-elevation areas, which links all four articles. Followed by *data quality,* as low data quality can further exacerbate data scarcity. *Uncertainty in the calibration objective* is mainly related to discharge and snow observations, which also includes the generated snow cover maps from article three. *Input uncertainty* deals with the uncertainty that comes with the meteorological forcing of the hydrological model. The uncertainty originates from inhomogeneities in the observed time series or from systematic (i.e., epistemic) underestimations in the case of solid precipitation in winter as mentioned in the third article. The first article on time series analysis addresses all four sub-categories.

The last category of *hydrological modeling* contains three aspects, namely *constrain model parameters*, *model efficiency*, and *model reliability*. The subtopic of *constraining model parameters* is only addressed in the third article of the observed and modeled snow cover maps. *Model efficiency* plays an important role in reducing the computational time of the physically-based model WaSiM that is discussed in the fourth article. *Model reliability* is an important aspect in all articles except of the article about the automated discharge measurement, where no modeling studies were performed.

Table 1-1: List of the research questions and addressed keywords.

Adressed keywords										
	Research area		Observational data					Hydrological modeling		
Research question	**Alpine catchments**	**Spatio-temporal complexity**	**Data scarcity**	**Data quality**	**Input uncertainty**	**Uncertainty in the calibration objective**	**Constrain model parameters**	**Model efficiency**	**Model reliability**
What is the impact of inhomogeneous time series on hydrological model results?									
How beneficial and reliable are automated stream discharge measurements in high-elevation areas to reduce data scarcity?									
Can we reduce data scarcity of spatial snow information in high-elevation Alpine catchments by remote sensing data?									
How fast and reliable are coupled machine learning and physically-based hydrological models?									

The following section highlights the motivation for the individual research questions and how it is addressed in the corresponding publication.

1.3.1. Research question 1: What is the impact of inhomogeneous time series on hydrological model results?

Since long time series (> years) are very likely to be affected by inhomogeneities, it is inevitable to check the quality of these before using them in a model application (Beven et al., 2022). Unfortunately, this is often not the case in most hydrological model applications. Unchecked station data is used, on one hand, as forcing and, on the other, to calibrate and validate the models. However, it is not clear to what extent the inhomogeneities can propagate to the model results and how reliable a model optimized against possibly inhomogeneous observational is. In high-elevation Alpine areas, where station data coverage is often poor and data collection is far more difficult than in lower-elevated areas, the quality of fewer time series is even more significant for model forcing and model calibration/validation. In the first article (Hofmeister, Graziano, et al., 2023), we applied a holistic approach, illustrated in Figure 1-1, to test the quality of 59 time series of two high-elevation Alpine catchments (i.e., Martelltal and Horlachtal). Before performing the actual quality tests by means of the HyStat software (Willems, Dick, Stricker, & Kasper, 2013), we eliminated the seasonality from each time series in order to reduce the number of significant autocorrelation coefficients. As conclusion, we found that inhomogenities in the model forcing (i.e., temperature and precipitation) influenced the cryosphere (i.e., snow and glaciers) directly and the runoff generation indirectly. Due to the non-linear behavior of hydrological processes, the asessment of consequences of including or excluding suspicious time series in hydrological setups is particularly difficult.

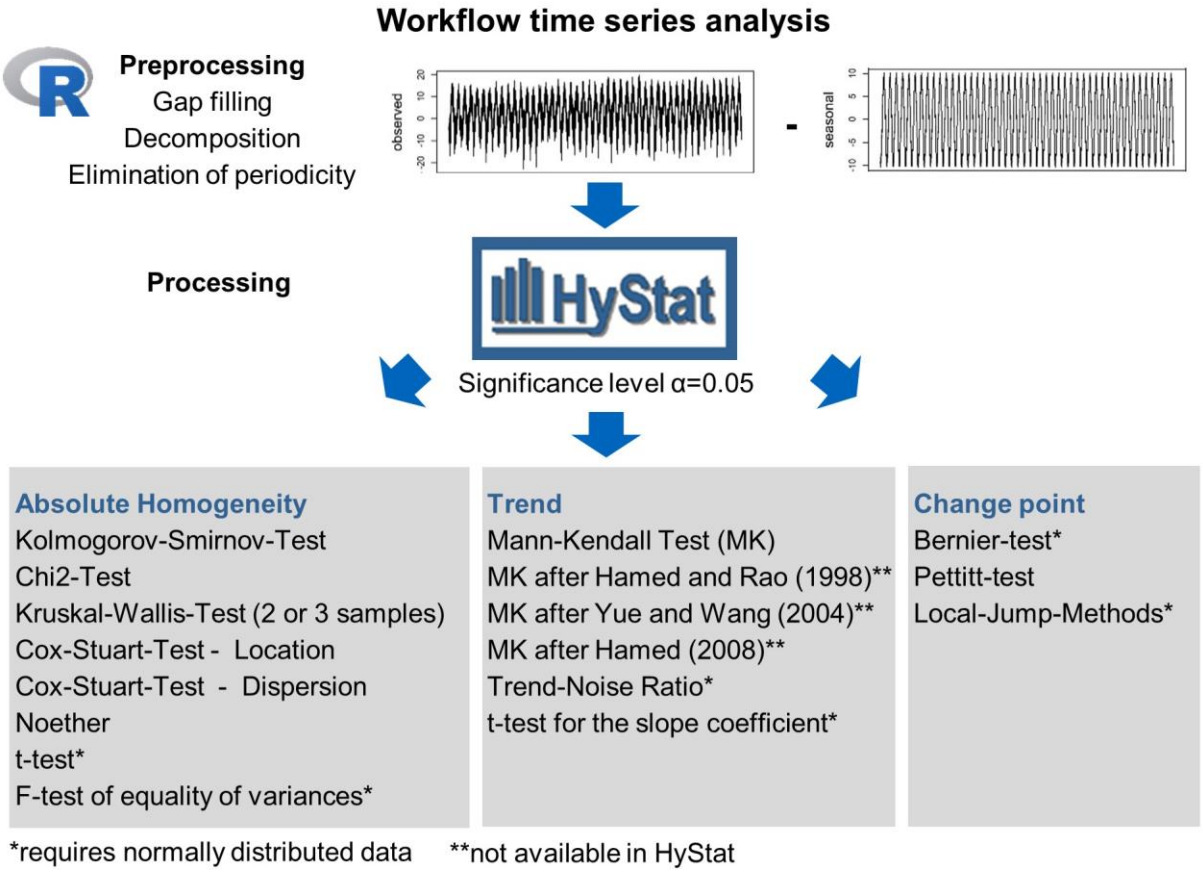


Figure 1-1: Workflow of time series quality check from preprocessing to processing (i.e., homogeneity, trend and change point detection) by means of the statistical HyStat software (Willems et al., 2013). Parametric tests that require normally distributed data are labeld with *. Adjusted Mann-Kenndall Tests (labeld with **) are not available in HyStat.

1.3.2. Research question 2: How beneficial and reliable are automated streamflow measurements in high-elevation areas to reduce data scarcity?

In order to reduce the data scarcity and to better investigate runoff generation in high-elevation Alpine catchments, event-based streamflow measurements are a key technology to capture rare events. In addition, rare events are important for rating curve construction and reducing uncertainty in extrapolation range. The main challenges consist in the proper installation and the regular maintainence of the automated system, to keep the experimental uncertainty as small as possible. In the second article (Hofmeister, Rubens Venegas, et al., 2023), we tested and evaluated the benefit and reliability of an automated salt injection system that was originally developed in Canada. For this purpose, we installed and assessed three automated salt injection systems over two years in three diverse areas of the Alps (i.e., electrical background conductivity, drained and glaciated area, and channel geometry). Figure 1-2 shows an example of one deployed AutoSalt system, the change of temperature compensated electrical conductivity (EC_T) in stream water due the salt injections, and the stage-discharge relationship based on two years of measurements. The blue box in Figure 1-2 illustrates a

hydrological extreme event that occurred in a side valley of the Horlachtal (i.e., Grastal) and partly destroyed one AutoSalt system in July of 2022. This particular event was analyzed by Rom et al. (2023). In the second article, we were able to demonstrate the reliable functioning of the system at contrasting sites with a majority (81 %) of small measurement errors (< 7 %) and the recording of rare events with peak streamflow of up to 20 m³/s. Moreover, the use of the AutoSalt system enables us to detect non-stationarity in the cross-section.

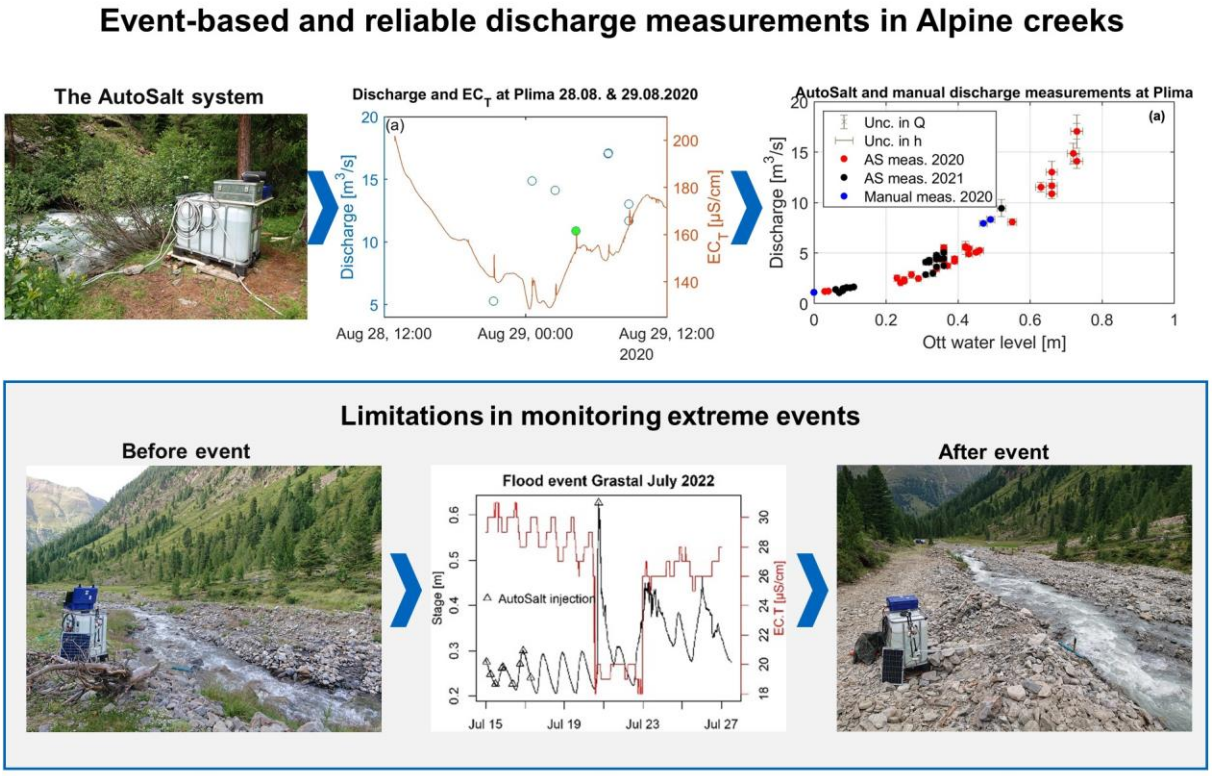


Figure 1-2: Automated salt injection system (AutoSalt) for unsupervised and event-based streamflow measurements in Alpine catchments. The extreme event occurred in a side valley of the Horlachtal (i.e., Grastal) in July 2022.

1.3.3. Research question 3: Can we reduce data scarcity of spatial snow information in high-elevation Alpine catchments by remote sensing data?

Data scarcity in high-elevation Alpine catchments is not only characterized by a lack of observational station data, but also by the spatial coverage in the complex terrains because station data mainly represent a lumped information (e.g., meteorological and snow observations). However, snow amount and cover is particularly heterogeneously distributed in high-elevation areas, which in turn affects, for example, the runoff generation in spring. Therefore, spatial snow information is essential for estimating spring runoff or for validating hydrological models. The third article (Hofmeister, Arias-Rodriguez, et al., 2022) shows the potential of high-resolution snow cover maps derived from optical remote sensing images for: i) constraining model parameters that control the gravitational snow redistribution and ii) for

validating the spatio-temporal model results (Figure 1-3). The optimized snow model is able to reproduce spatio-temporal variability in snow cover duration due to surface energy balance dynamics, wind and gravitational redistribution. It can be assumed that the validated hydrological model gives more realistic runoff predictions than the model setup without the optimization and additional validation by means of observed snow cover maps. Through these promising results, we have shown the potential of observed high-resolution snow cover maps for consideration in a multi-objective optimization approach, which reduces the uncertainty in the model results.

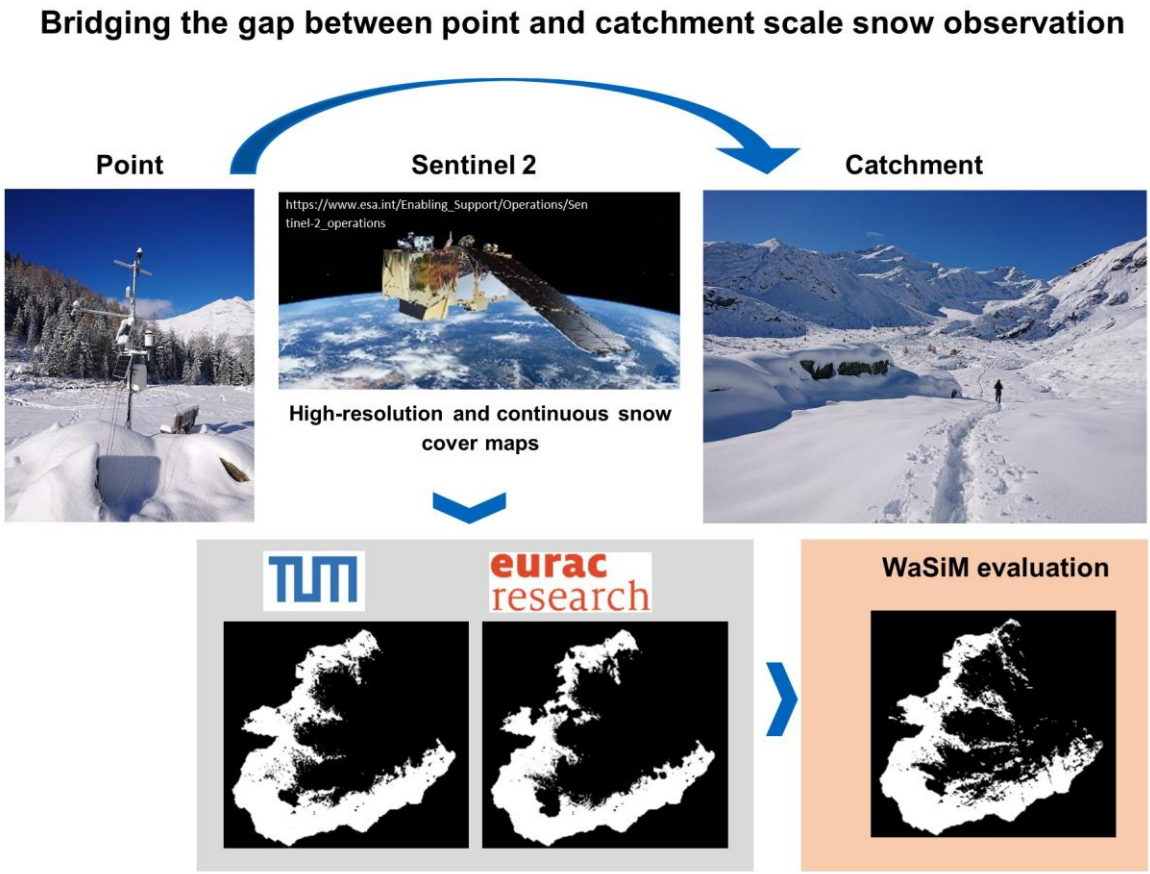


Figure 1-3: Illustration of the research objective to reduce data scarcity of spatial snow information in high-elevation Alpine catchments by remote sensing data (grey box). The derived snow cover maps are beneficial for the validation of the model results (red box) and for constraining the model parameters of a physically-based and distributed snow model (i.e., WaSiM).

1.3.4. Research question 4: How fast and reliable are coupled machine learning and physically-based hydrological models?

Although physically-based hydrological models are powerful to reliably and consistently simulate hydrological processes, they may be limited in their applicability because of data availability, process representation, or computational time. Coupling physically-based hydrological models with machine learning has great potential as it can be used very flexibly. The coupling results in synergies that can reduce the respective limitations of the other

approach. In the fourth article (Hofmeister, Spadina, & Chiogna, 2022), we coupled WaSiM with a Suppport Vector Machine (SVM) to speed up the computational time for long simulation runs. Until then, it was still unknown i) how a coupling can work (e.g., surrogate model), ii) how large the reduction in computing time is, and iii) how reliable the predictions generated with the emulator are (Figure 1-4) for a high-elevation Alpine catchment. Although article four is only a case study, we were able to show that coupling with a Support Vector Regression (SVR) algorithm provides reliable results and that a significant reduction in computation time is possible. Thus, the study is a good foundation for further investigations and tests in other areas and for longer time periods.

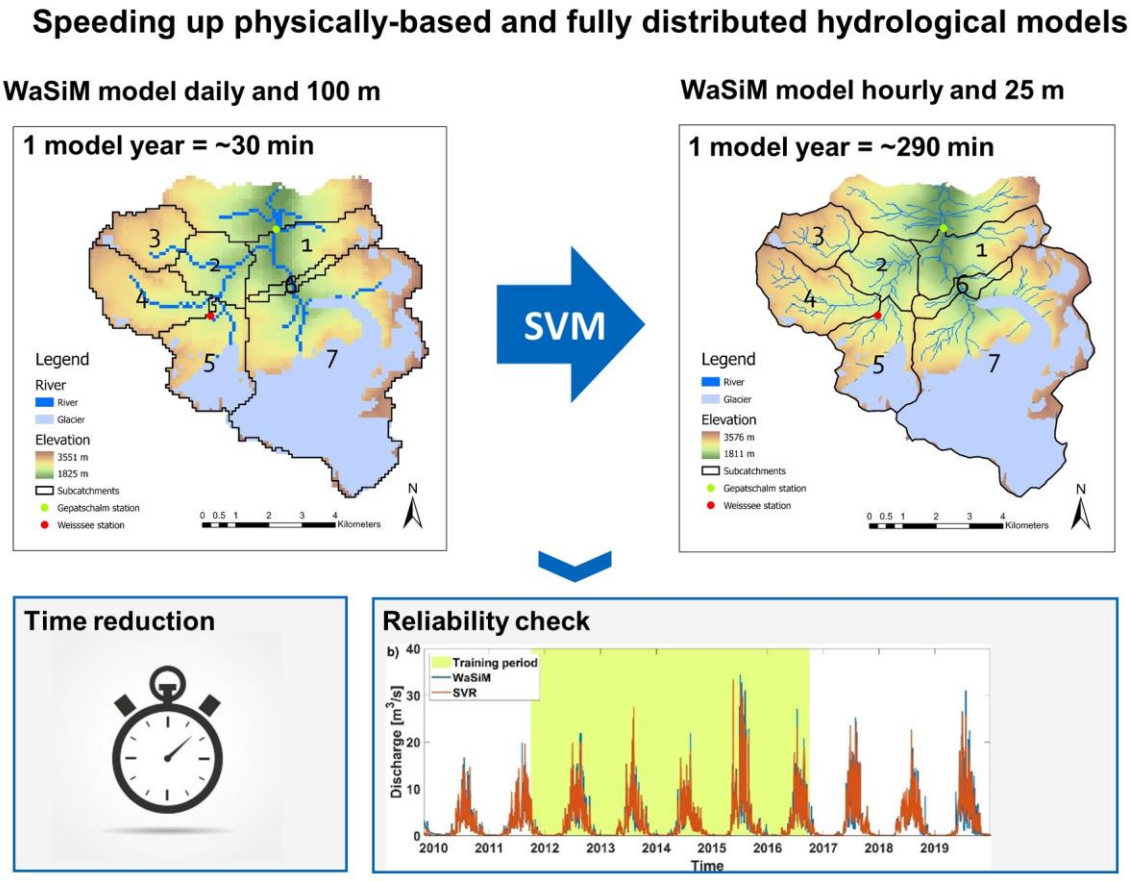


Figure 1-4: Emulation of a high-resolution hydrological model setup (i.e., 25 m spatial and hourly time step) by a Support Vector Machine (SVM) and a courser model setup (i.e., 100 m spatial and daily time step). The evaluation of the emulator focuses on time reduction and reliability of the model predictions.

1.4. Research areas

The research areas of this dissertation are mainly focused on two high-elevation headwater catchments in the Central Alps (Kaunertal and Horlachtal) and one in the Southern Alps (Martelltal), as illustrated in Figure 1-5. Accordingly, the catchments of the Central Alps are located north of the main Alpine divide and the Martelltal south of it. All three catchments correspond to the research areas of the SEHAG project funded by the DFG (German Research

Foundation) and FWF (Austrian Science Fund). The selection of these three research areas was motivated by multiple reasons. They have a similar geology (i.e., metamorphic Austroalpine nappes (Schmid, Fügenschuh, Kissling, & Schuster, 2004)), topography and size (i.e., ~60 km²). As the three study areas provide similar conditions regarding geology (i.e., siliceous metamorphic rocks), topography (i.e., rather steep slopes with a high morphodynamic activity) and vegetation, the soil formation is also similar. Soil types range from shallow Leptosols to deeper developed Cambisols and Podzols. Differences mainly consist in the climate and the glacier coverage. A detailed description of each research area is included in the following section.

Figure 1-5: Outlines of the three research areas Horlachtal, Kaunertal and Martelltal (from north to south).

1.4.1. Horlachtal

Horlachtal, located in the Ötztal part of Tyrol, Austria, is a 55 km² Alpine catchment drained at Gasthof Stuibenfall gauge station, with glaciers covering 1.8 % of the catchment area in 2015 (Buckel & Otto, 2018). The catchment elevation varies from 1484 to 3277 m.a.s.l. and the mean elevation is 2443 m.a.s.l. The rock mainly consist of gneiss, mica-schists and scattered amphibolites (Schmid et al., 2004). The climate of Horlachtal corresponds to the typical dry valley character of the upper Inn basin and the vegetation has a characteristic altitudinal alpine

gradation (Rom, Haas, Heckmann, et al., 2023). Since the Horlachtal has a typical geomorphic dynamic, some studies have already been conducted to investigate the geomorphic processes (Becht, 1995; Geitner & Becht, 1999; Rom, Haas, Heckmann, et al., 2023). Mean annual precipitation computed using all available stations (Horlachalm, Leiter, Küthai and Umhausen) over the period 1991-2019 is 916 mm with a standard deviation (SD) of ±106 mm. The mean annual temperature computed using the same stations over the same period is -0.1 °C with a SD of ±0.64 °C. The available stations from Horlachtal and the surrounding area are shown in Figure 1-6.

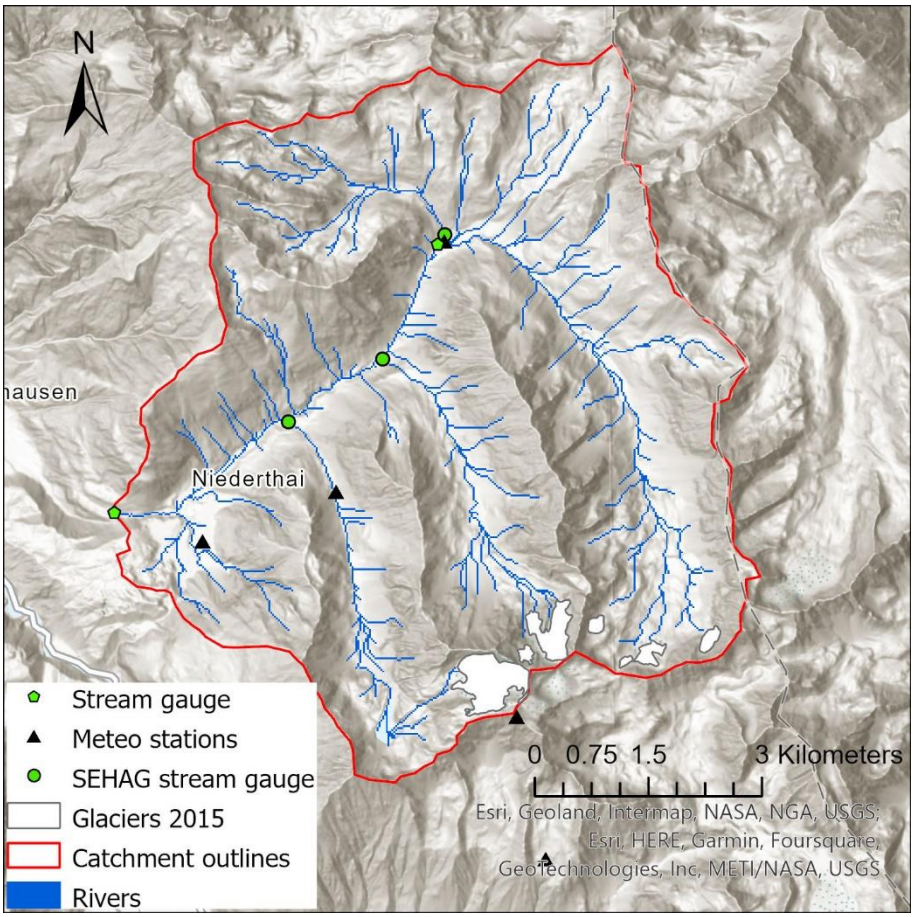


Figure 1-6: Overview map of the Horlachtal with the glacier extents from 2015 (Buckel & Otto, 2018) and locations of hydrometeorological stations. Green circles indicate the stream gauges installed in the SEHAG project.

The Tiroler Wasserkarft AG (TIWAG) and the Avalanche Warning Service (LWD) are the main operators of meteorological stations in Horlachtal and the neighboring valleys. In addition, the Central Institution for Meteorology and Geodynamics (ZAMG) runs a meteorological station in the main Ötztal at 1035 m.a.s.l. next to Umhausen, which has been recording daily meteorological variables since 1946. The highest station is Lampsenspitze Windstation at 2870 m.a.s.l.. The TIWAG station Leiter has the longest temporal coverage within Horlachtal beginning in 1973. From 1990 on, the number of available stations has increased. However, stations located at higher elevations (above 2000 m.a.s.l.) are available only from 2010

onwards. High-resolution (15 minutes) hydrological observations are available at gauge Horlach Fassung and Niederthai Gasthof Stuibenfall from 1985 onwards. At the Horlach Fassung, most of the stream water is diverted to the Finstertal reservoir near Kühtai. In the framework of the SEHAG project, additional stream gauges were installed at three tributary valleys (i.e., Finstertal, Larstigtal and Grastal) in 2019 to better understand the runoff generation in these small headwater catchments.

1.4.2. Kaunertal

The Kaunertal is part of the Oetztaler Alps and located in the central Alps in Tyrol/Austria. The valley has a south-north orientation with an altitudinal range from 1895 m.a.s.l. to 3576 m.a.s.l. and is drained by the Fagge river, a tributary of the Upper Inn River system. The research area of this study covers the upper Kaunertal south of the Gepatsch reservoir and draines at the gauge Gepatschalm with an approximated size of 54 km². The geology is mainly characterized by paragneis, orthogneis and interacted amphibolites (Schmid et al., 2004). The mean annual temperature computed using the Weisssee station (2540 m.a.sl.) over the 2006-2021 period is -0.26 °C with a SD of ±1.15 °C. Mean annual precipitation sum at Weisssee is 813 mm with a standard deviation (SD) of ±335 mm over the same period. The runoff regime of the Fagge has a glacio-nival characteristic with a long low flow period in winter (November to May) and a shorter period of larger discharge fluctuations in summer (June to October). Mean discharge of the Fagge at gauge Gepatschalm was 3 m³/s over the period 2009/11 to 2019/12 and the maximum hourly discharge was 49.18 m³/s on 22.08.2011. The Gepatsch glacier, the second largest glacier of Austria, together with the Weißseeferner, covers approximately 37 % of the catchment (Förster et al., 2016). In adition to multiple hydrological (Förster et al., 2016; Rogger et al., 2017) and cryospheric investigations (Fey, Schattan, Helfricht, & Schöber, 2019; Fleischer et al., 2021; Schattan, Schwaizer, Schöber, & Achleitner, 2020; Schöber et al., 2014; Wagner, Brodacz, Krainer, & Winkler, 2020; Wagner, Pauritsch, & Winkler, 2016), a number of other processes have been studied in the Kaunertal, such as changes in the sediment transport (Baewert & Morche, 2014; Heckmann et al., 2012; Schöber & Hofer, 2018) and in the morphodynamics of this high Alpine area (Altmann et al., 2020; Haas, Heckmann, Hilger, & Becht, 2012). Within the framework of the SEHAG project, additional stream gauges were installed at three sites (i.e., upper Fagge, Riffler creek and Krummgampental) in 2019. The reason of the installation was to subdivide the upper Kaunertal in smaller sub-catchments with different shares of glacier cover. The hydrometeorological stations as well as the glacier cover of the upper Kaunertal from 2015 is shown in Figure 1-7.

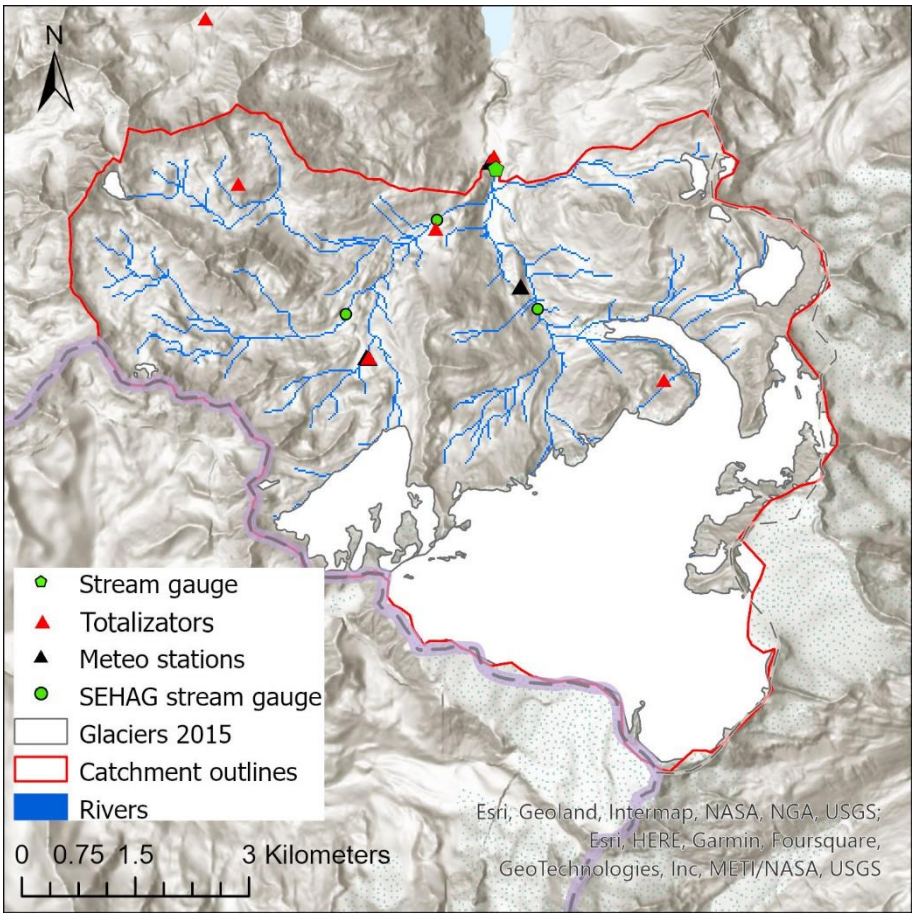


Figure 1-7: Overview map of the Kaunertal with the glacier extents from 2015 (Buckel & Otto, 2018) and locations of hydrometeorological stations. Green circles indicate the stream gauges installed in the SEHAG project. Precipitation totalizators of the TIWAG are indicated with red triangles.

1.4.3. Martelltal

Martelltal, situated in South Tyrol, Italy, is a 62 km² Alpine catchment discharging in the artificial Zufritt (Giovaretto) reservoir and had a glacier coverage of 15 % in 2019 (Knoflach et al., 2021) (see Figure 1-8). The catchment elevation varies from 1838 to 3747 m.a.s.l. and the mean elevation is 2814 m.a.s.l.. In Martelltal, para-gneis, mica-schists, phyllites and orthogneis are the prevailing metamorphic units (Schmid et al., 2004). The climate in Martelltal is characterized by a transition zone of Mediterranean climate and the drier climate of the Central Alpine region (Knoflach et al., 2021). Based on the data of the Stausee Zufritt station, we could compute a mean annual precipitation of 910 mm with a SD of ±217 mm and a mean annual temperature of -1.7 °C with a SD of ±1.21 °C over the period 1973-2020.

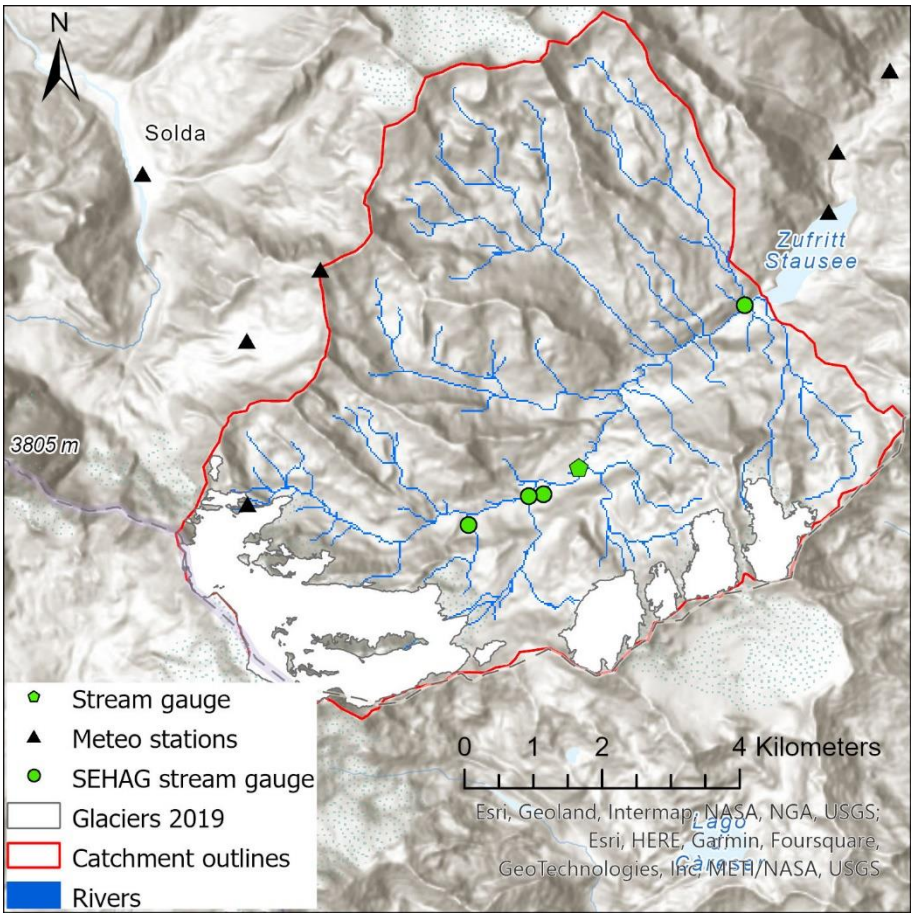


Figure 1-8: Overview map of the Martelltal with the glacier extents from 2019 (Knoflach et al., 2021) and locations of hydrometeorological stations. Green circles indicate the stream gauges installed in the SEHAG project.

The meteorological station at Langenferner is operated by the Institute of Atmospheric and Cryospheric Sciences of the University of Innsbruck, the Autonomous Province of Bozen/Bolzano - South Tyrol runs all other stations. The meteorological station network covers a large altitude gradient from 1720 m.a.s.l. (Hintermartell) to 3328 m.a.s.l. (Schöntaufspitze). However, temporal coverage of metrological observations is very sparse prior 2005. The Zufritt station displays the largest temporal coverage, having measured daily temperature and precipitation since 1973. Previous hydrological studies investigating the mass balance of the Langenferner glacier (Galos et al., 2017; Galos & Klug, 2015) and the impact of glacier shrinking on hydropower production (Puspitarini, François, Zaramella, Brown, & Borga, 2020) also focused on this study area. In addition, the catastrophic flood event from 1987 was the subject of several vulnerability assessment studies (Totschnig & Fuchs, 2013; Papathoma-Köhle, Zischg, Fuchs, Glade, & Keiler, 2015).

1.5. Hydrological model WaSiM

The selection of the deterministic, grid-based **Wa**ter Flow and Balance **Si**mulation **M**odel WaSiM (Schulla, 2021) is mainly based on the model requirements within the SEHAG project (**Se**nsitivity of **h**igh **A**lpine **g**eosystems to climate change since 1850), in which climate-induced changes on the hydrosphere of high-elevation catchments since the last Little Ice Age (LIA) are investigated Hence, the hydrological model has to be able to simulate elevation dependent (i.e., temperature dependent) hydrological processes such as snow and glacier melt. WaSiM is a physically-based model because it simulates water fluxes in the unsaturated zone by solving the 1D-Richards equation, evapotranspiration can be computed using the Penman-Monteith equation as well as snow ablation by solving the energy balance of the snow surface. The application of physically-based algorithms is strongly limited by the availability of forcing data and spatial information, especially soil texture, horizons and thickness. Applying the Penman-Monteith and the energy balance algorithms requires meteorological records of relative humidity, wind speed and short wave incoming radiation, in the research areas the latter is available, at the earliest, since the beginning of the 1990's. If these forcing data are not available, WaSiM can simulate the evapotranspiration and the snowmelt with less complex algorithms such as Hamon or Haude for evapotranspiration and temperature index (T-index) based on degree-days for snowmelt. Accordingly, temperature and precipitation observations are the minimum input for forcing WaSiM with the lowest model complexity. The use of physics-based algorithms ensures that model results remain within plausible physical limits even over long simulation periods (> 30 years). However, the higher degree of model complexity complicates the setup, calibration and validation of a hydrological model as it requires a deep understanding about the hydrological cycle and how it is implemented in the model. WaSiM has a logical model structure similar to the natural water cycle beginning with the atmosphere at the top and the water fluxes in the unsaturated and saturated zone as illustrated in Figure 1-9. A great advantage of the modular structure is that each user can configure the model complexity according to the research question, area, and data availability.

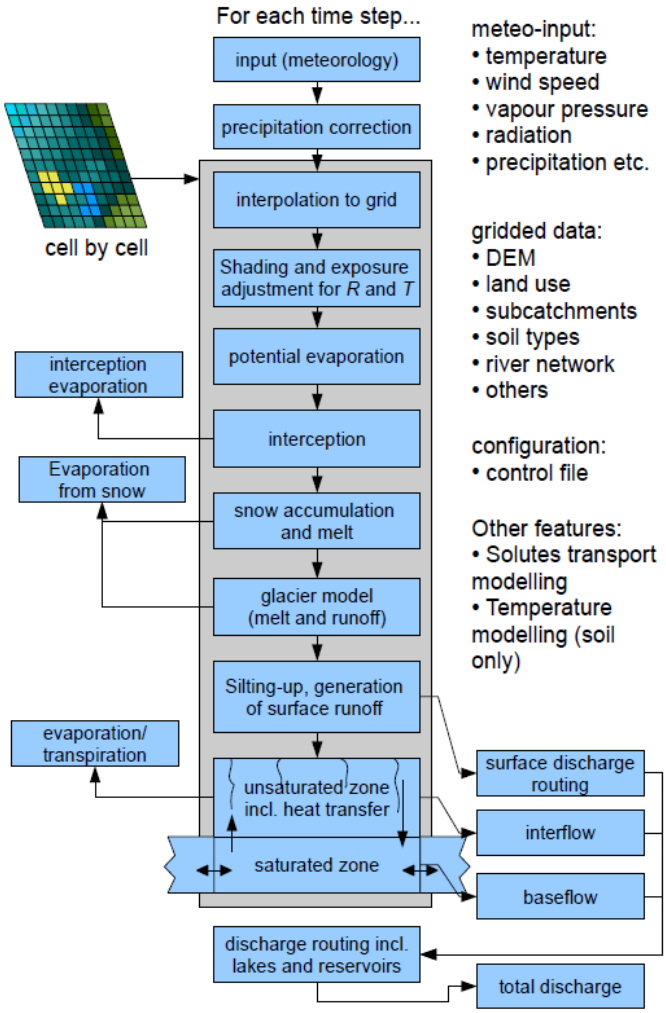


Figure 1-9: Model structure of the grid- and physically-based hydrological model WaSiM (Schulla, 2021).

With the implementation of snow redistribution routines by Warscher et al. (2013), WaSiM is able to simulate wind-induced snow redistribution and gravitational snow slides that avoid the so-called 'snow-tower' at the top of mountain ridges. Both redistribution algorithms are not event-based, they are designed to simulate the snow distribution at the end of the accumulation period in complex topographies. Event-based snow redistribution processes are available, for example, in the Alpine3D model (Lehning et al., 2006). Another important argument for the selection of WaSiM was the simulation of glacier mass balances as well as glacier dynamics (i.e., recession and advance) due to climatic changes. The implemented algorithm is the volume-area power law scaling, a widely used method for estimating the future response of ice caps and glaciers to environmental change (Bahr, Meier, & Peckham, 1997; Bahr, Pfeffer, & Kaser, 2015). Ice and firn melt can be simulated either with an T-index or by including radiation after Hock (1999). A new coupling of WaSiM with the explicit ice flow model **O**pen **G**lobal **G**lacier **M**odel (OGGM), developed by Maussion et al. (2019), is currently in development for a better representation of glacier geometry and ice dynamics (Pesci & Förster,

2022). Although not considered in the model configuration for the model runs within the SEHAG framework, WaSiM has the option to compute the heat transfer in the snowpack and subsurface to simulate permafrost dynamics. The heat transfer model was mainly developed and tested in Arctic tundra (Bui, Lu, & Nie, 2020; Debolskiy et al., 2021; Liljedahl et al., 2016).

Because of the detailed representation of land cover, WaSiM has been used not only for climate change studies (Kunstmann, Schneider, Forkel, & Knoche, 2004; Meyer, Blaschek, Duttmann, & Ludwig, 2016; Poschlod, Willkofer, & Ludwig, 2020), but also for land use change studies (Bormann & Elfert, 2010; Hounkpè, Diekkrüger, Afouda, & Sintondji, 2019) or for combined scenarios (Alaoui et al., 2014; Springer, Ludwig, & Kienzle, 2015; Strasser et al., 2019). Moreover, WaSiM has the capability to consider anthropogenic streamflow modifications such as irrigation, reservoirs or water bypasses. The latter was particularly relevant for the model setup of Horlachtal as about 80 % of the water from the upper catchment are diverted to a neighboring valley.

A relatively high spatial model discretization of 25 m x 25 m was chosen as a compromise between a reasonable computational time for the three research areas and the ability to resolve small-scale processes such as snow redistribution and correct deviation of flow paths of Alpine creeks. Additionally, it is possible to compare and provide the simulated outputs (e.g., snow cover duration) to other subprojects of the SEHAG research unit (e.g., botany, sediment transport) as they are conducting their studies on a meter resolution (Knoflach et al., 2021; Ramskogler et al., 2023). Moreover, as described in chapter 4, WaSiM is able to reproduce small-scale snow redistribution processes with the 25 m x 25 m resolution.

The model time step of WaSiM was chosen to be 1h in order to simulate diurnal runoff generation from snow and glacier melt. Another reason for the 1h time step was the modeling and investigation of hydrological extreme events that usually occur in short time windows due to quick runoff concentration time (i.e., only a few hours) in the steep headwater catchments.

2. Quality assessment of hydrometeorological observational data and their influence on hydrological model results in Alpine catchments

Hofmeister, Graziano, et al., 2023[1]

Abstract

Hydrometeorological observational data can be affected by trends, change points and inhomogeneities. It is often difficult to identify the cause of these statistical properties of the time series and therefore they represent a source of uncertainty that can propagate to the hydrological model results. We examine two case studies of high-elevation Alpine catchments for which 59 hydrometeorological time series are analyzed to show how often trends, change points and inhomogeneities in the time series occur. We consider daily and monthly aggregated data, as well as a set of nine homogeneity test, six trend tests and three change point tests. The results show that inhomogeneities in meteorological time series have a stronger and more direct influence on the cryosphere (snow and glacier processes) than on runoff generation. The non-linear behavior of hydrological processes greatly complicates the judgement about the consequences of including or excluding suspicious time series in hydrological model set-ups.

[1]Material from:
Hofmeister, F., Graziano, F., Marcolini, G., Willems, W., Disse, M., & Chiogna, G. (2023). Quality assessment of hydrometeorological observational data and their influence on hydrological model results in Alpine catchments. *Hydrological Sciences Journal*, 02626667.2023.2172335. doi: 10.1080/02626667.2023.2172335.

2.1. Introduction

The editorial by Cudennec, Lins, Uhlenbrook, & Arheimer (2020) highlighted the importance of advancing hydrological sciences of FAIR (i.e., findable, accessible, interoperable, reusable data; Wilkinson et al., 2016) and SQUARE data (i.e., those that support quality action and research). In fact, errors in hydrological and meteorological time series are often present, and it is important to identify them since experimental hydrology is a very challenging discipline. Pitfalls in experimental hydrology exist not only in field measurements (Dembélé, Oriani, Tumbulto, Mariéthoz, & Schaefli, 2019; Di Baldassarre & Montanari, 2009b; Rasmussen et al., 2012) but also in laboratory analysis (Orlowski et al., 2018; Penna et al., 2010) and processing of remote sensing datasets (Anagnostou et al., 2018; Borga, Degli Esposti, & Norbiato, 2006; Di Marco et al., 2020a; Hofmeister, Arias-Rodriguez, et al., 2022). Measurement errors in hydrological and meteorological time series not only represent input errors for hydrological models but can also challenge hydrological model validation and calibration (Bittner, Richieri, & Chiogna, 2021; Brath, Montanari, & Toth, 2004; Mcmillan, Jackson, Clark, Kavetski, & Woods, 2010). It has been shown that the uncertainty in meteorological time series in both accumulation and ablation phases is a crucial factor for the model performance of snow models. The input uncertainties may even be more relevant than the parameter choice and model structure of physically-based snow models (Günther, Marke, Essery, & Strasser, 2019). The fact that measurement errors can be compensated by parameter calibration (Bárdossy & Singh, 2008; Zhang, Li, Huang, Wang, & Cheng, 2016) further increases the importance of properly proofing the quality of the available datasets.

Testing time series is often considered a tedious exercise and it often does not lead to publications with high scientific impact, but it is an important prerequisite for robust hydrological studies that should be better acknowledged by the scientific community. Moreover, some hydrometeorological variables such as temperature, precipitation and discharge (Hänsel, Medeiros, Matschullat, Petta, & de Mendonça Silva, 2016; Kocsis, Kovács-Székely, & Anda, 2020; Shen et al., 2018; Yozgatligil & Yazici, 2016; Whitfield & Pomeroy, 2017; Gudmundsson, Do, Leonard, & Westra, 2018; Pandžić et al., 2020) are more often tested than others like snow depth, humidity, wind and radiation (Boris Faybishenko et al., 2021; Fiebrich, Morgan, McCombs, Hall, & McPherson, 2010; Marcolini et al., 2019b; Schöner et al., 2019). However, all hydrometeorological drivers are important to describe the hydrological cycle properly and should be carefully validated before being used for hydrological purposes. Despite being mathematically sound, available statistical tests for time series analysis still require interpretation for practical applications in hydrology. Relevant examples are outliers tests and their typical inability to distinguish between (physical) extreme events and (unphysical) measurement errors (Crochemore et al., 2020). Moreover, hydrological time series are often

autocorrelated with consequences on the significance of the statistical tests (El-Shaarawi & Damsleth, 1988; Hirsch, Slack, & Smith, 1982).

In this work, we illustrate the problem of using suspicious time series (i.e., time series that display inhomogeneity, trends or change points) as input for a hydrological model. We consider two Alpine catchments, i.e., Horlachtal in Tyrol/Austria and Martelltal in South Tyrol/Italy, in which precipitation and temperature time series, respectively, are of suspicious quality. We focus on Alpine catchments for four reasons. First, they are sensitive to small changes in temperature (Crespi, Matiu, Bertoldi, Petitta, & Zebisch, 2021; Jennings & Molotch, 2019; Zampieri, Scoccimarro, Gualdi, & Navarra, 2015). Second, they are typically data-scarce regions and therefore excluding a single time series from the hydrological analysis of a catchment can be a difficult choice that, in extreme cases, can make performing any analysis on the catchment impossible (Hingray, Schaefli, Mezghani, & Hamdi, 2010). In addition, the measurement of precipitation and snow depth at higher altitudes (> 2000 m.a.s.l.) is particularly prone to error due to strong winds that lead to undercatch of solid precipitation (Kochendorfer et al., 2022, 2016; Rasmussen et al., 2012; Sevruk, 1991). Finally, hydrometeorological time series in Alpine catchments are spatially heterogeneous and therefore spatial interpolation is more complex than in catchments displaying low elevation gradients (Foehn, García Hernández, Schaefli, & De Cesare, 2018). However, since well-configured high-resolution atmospheric models have become available, a better representation of the spatial and temporal precipitation distribution is possible (Lundquist, Hughes, Gutmann, & Kapnick, 2019). This is beneficial for runoff and snow simulation results, but these data are often only available for a relatively short period.

The paper is structured as follows. In the Material and methods section, we introduce the study areas and the hydrometeorological dataset available for this work. Afterwards, we review the statistical tests performed for hydrological time series analysis and we present the grid-based **Wa**ter Flow and Balance **Si**mulation **M**odel (WaSiM) (Schulla, 2021) applied in this study. The Results section presents the quality check results for the dataset and the effect of excluding or maintaining in the model exercise time series of suspicious quality. In the subsequent Discussion, we critically present how to interpret the results of the statistical tests and the implications of different statistical anomalies in the time series for hydrological modeling. Finally, the conclusion highlights the importance of performing accurate time series analysis to obtain accurate and reliable model predictions in Alpine catchments.

2.2. Material and methods

2.2.1. Study area

2.2.1.1. Horlachtal

Horlachtal, located in the Ötztal part of Tyrol, Austria, is a 55 km² Alpine catchment drained at Gasthof Stuibenfall gauge station, with glaciers covering 1.8 % of the catchment area in 2015 (Buckel & Otto, 2018). The catchment elevation varies from 1484 to 3277 m.a.s.l. and the mean elevation is 2443 m.a.s.l. The climate of Horlachtal corresponds to the typical dry valley character of the upper Inn basin. Mean annual precipitation computed using all available stations (Horlachalm, Leiter, Küthai and Umhausen) over the period 1991-2019 is 916 mm with a standard deviation (SD) of ±106 mm. The mean annual temperature computed using the same stations over the same period is -0.1 °C with a SD of ±0.64 °C. The available stations from Horlachtal and the surrounding area are shown in Figure 2-1a.

The Tiroler Wasserkarft AG (TIWAG) and the Avalanche Warning Service (LWD) are the main operators of meteorological stations in Horlachtal and the neighboring valleys. In addition, the Central Institution for Meteorology and Geodynamics (ZAMG) runs a meteorological station in the main Ötztal at 1035 m.a.s.l. next to Umhausen, which recodes daily meteorological variables since 1946. The highest station is Lampsenspitze Windstation at 2870 m.a.s.l.. The TIWAG station Leiter has the longest temporal coverage within Horlachtal beginning in 1973. From 1990 on, the number of available stations increased. However, stations located at higher elevations (above 2000 m.a.s.l.) are available only from 2010 onwards. High-resolution (15-minute) hydrological observations are available at the gauge Horlach Fassung and Niederthai Gasthof Stuibenfall from 1985 onwards.

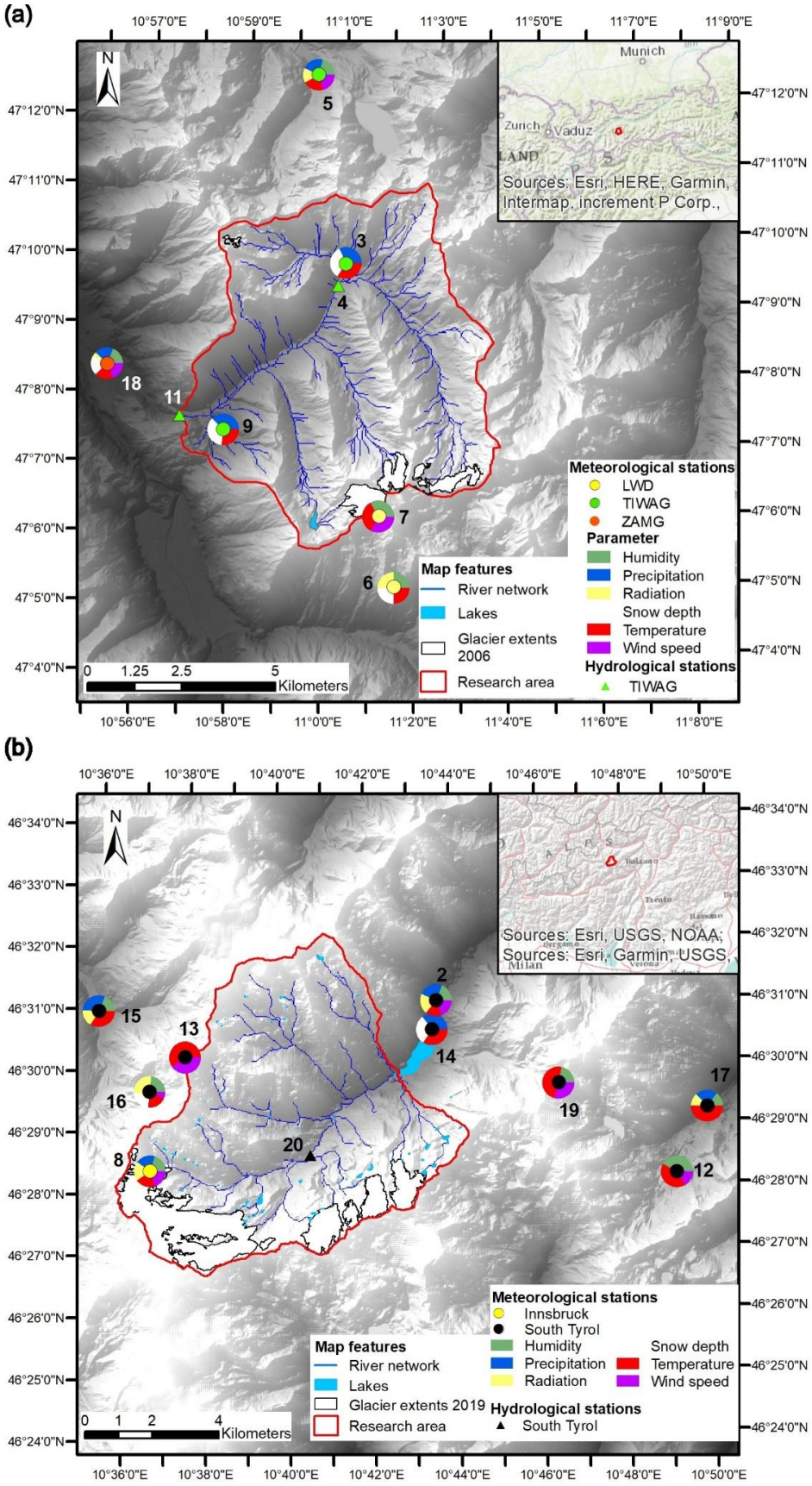


Figure 2-1: Overview map of (a) Horlachtal and (b) Martelltal with locations of meteorological and hydrological stations and their corresponding observation variables. Station numbers correspond to station numbering in the hydrometeorological dataset section, as listed in Table 2-1.

2.2.1.2. Martelltal

Martelltal, situated in South Tyrol, Italy, is a 62 km² Alpine catchment discharging into the artificial Zufritt (Giovaretto) reservoir and had a glacier coverage of 15.6 % in 2015 (see Figure 2-1b). The catchment elevation varies from 1838 to 3747 m.a.s.l. and the mean elevation is 2814 m.a.s.l. The climate in Martelltal is characterized by a transition zone of Mediterranean climate and the drier climate of the Central Alpine region (Knoflach et al., 2021). Based on data from the Stausee Zufritt station, we computed a mean annual precipitation of 910 mm with an SD of ±217 mm and a mean annual temperature of -1.7 °C with an SD of ±1.21 °C over the period 1973-2020.

Except for the meteorological station at Langenferner, which is operated by the Institute of Atmospheric and Cryospheric Sciences of the University of Innsbruck, the Autonomous Province of Bozen/Bolzano - South Tyrol runs all other stations. The meteorological station network covers a large altitude gradient, from 1720 m.a.s.l. (Hintermartell) to 3328 m.a.s.l. (Schöntaufspitze). However, temporal coverage of metrological observations is very sparse before 2005. The Zufritt station displays the largest temporal coverage; it has measured daily temperature and precipitation since 1973. Since only daily minimum and maximum temperature are recorded at Zufritt station, we computed the daily average by taking the mean of both observations, which is a traditional approach in meteorological applications (Bernhardt, Carleton, & LaMagna, 2018).

2.2.2. Hydrometeorological dataset

In total, 60 time series from 20 stations are available for the two study areas. We reduced the total number of time series to 59 from 19 stations because we only analyzed time series that cover at least five years. Hydrometeorological variables measured at one station can generate multiple time series, if data gaps larger than 14 days are present. We filled smaller data gaps (< 14 days) in the time series by linear interpolation with neighboring stations if the correlation coefficient was larger than 0.7. Table 2-1 lists information about the analyzed time series. The station numbering includes a jump in Table 1 because only short time series (< 5 years) were available from the station Langenferner-Felsköpfl, and therefore were excluded from the analysis. An additional numbering in the variable column differentiates between multiple time series of the same hydrometeorlogical variable belonging to the same station (e.g., Hintermartell T 1 and T 2) due to the occurrence of gaps in the recording. The following variables are considered: 10 relative humidity time series (H), 10 precipitation time series (P), four radiation time series (R), 20 temperature time series (T), eight wind speed time series (WS), four snow depth time series (HS) and three discharge time series (Q). For the Horlachtal we analyzed 27 time series from eight stations, while for the Martelltal we analyzed 32 time

series from 11 stations. A table containing all metadata of the analyzed and not analyzed hydrometeorological dataset is provided in the supplemental material A.1.1.

Table 2-1: Characteristics of the tested hydrometeorological time series. Variables are abbreviated as Q = discharge, P = precipitation, WS = wind speed, HS = snow depth, T = temperature, H = humidity, R = shortwave incoming radiation.

Station Name	Station code	Variable	Number of years	Latitude	Longitude	Elevation m.a.s.l.	Observation period	Provider
Durraplatt	1	P	16	46.5274	10.7373	1616	1989-2005	Bozen
Hintermartell	2	H	11	46.5169	10.7269	1720	2009-2020	Bozen
Hintermartell	2	P	11	46.5169	10.7269	1720	2009-2020	Bozen
Hintermartell	2	R	5	46.5169	10.7269	1720	2009-2014	Bozen
Hintermartell	2	T 1	5	46.5169	10.7269	1720	2009-2014	Bozen
Hintermartell	2	T 2	6	46.5169	10.7269	1720	2014-2020	Bozen
Hintermartell	2	WS	11	46.5169	10.7269	1720	2009-2020	Bozen
Horlachalm	3	HS 1	6	47.1577	11.0125	1910	1999-2005	TIWAG
Horlachalm	3	HS 2	10	47.1577	11.0125	1910	2009-2018	TIWAG
Horlachalm	3	P	29	47.1577	11.0125	1910	1989-2018	TIWAG
Horlachalm	3	T	31	47.1577	11.0125	1910	1987-2018	TIWAG
Horlach Fassung	4	Q	33	47.1577	11.0125	1912	1985-2018	TIWAG
Kühtai	5	H	29	47.2071	11.0059	1918	1990-2019	TIWAG
Kühtai	5	P	26	47.2071	11.0059	1918	1990-2016	TIWAG
Kühtai	5	R	29	47.2071	11.0059	1918	1990-2019	TIWAG
Kühtai	5	T	29	47.2071	11.0059	1918	1990-2019	TIWAG
Kühtai	5	WS	25	47.2071	11.0059	1918	1990-2015	TIWAG
Lampenspitze Schneestation	6	H	11	47.153	11.121	2111	2008-2019	LWD
Lampenspitze Schneestation	6	R	11	47.153	11.121	2111	2008-2019	LWD
Lampenspitze Schneestation	6	T	8	47.153	11.121	2111	2011-2019	LWD

Lampenspitze Windstation	7	H	10	47.156	11.096	2870	2009-2019	LWD
Lampenspitze Windstation	7	T	8	47.156	11.096	2870	2011-2019	LWD
Lampenspitze Windstation	7	WS	10	47.156	11.096	2870	2009-2019	LWD
Leiter	9	HS	41	47.1228	10.9691	1564	1976-2018	TIWAG
Leiter	9	P	43	47.1228	10.9691	1564	1976-2019	TIWAG
Leiter	9	T 1	8	47.1228	10.9691	1564	1991-1999	TIWAG
Leiter	9	T 2	20	47.1228	10.9691	1564	1999-2019	TIWAG
Malga Mare	9	P	54	46.4142	10.68	1950	1930-1984	Trentino
Niederthai	11	Q	33	47.1577	11.0125	1478	1985-2018	TIWAG
Rossbänke	12	H	18	46.4693	10.8194	2255	2002-2020	Bozen
Rossbänke	12	T	18	46.4693	10.8194	2255	2002-2020	Bozen
Rossbänke	12	WS	6	46.4693	10.8194	2255	2013-2019	Bozen
Schöntaufspitze	13	T	13	46.5029	10.6286	3328	2003-2016	Bozen
Schöntaufspitze	13	WS	5	46.5029	10.6286	3328	2003-2008	Bozen
Stausee Zufritt	14	P	47	46.509	10.7251	1851	1972-2019	Bozen
Stausee Zufritt	14	T	46	46.509	10.7251	1851	1973-2019	Bozen
Sulden	15	H	5	46.5159	10.5953	1907	2015-2020	Bozen
Sulden	15	P	7	46.5159	10.5953	1907	2003-2010	Bozen
Sulden	15	T 1	7	46.5159	10.5953	1907	2003-2010	Bozen
Sulden	15	T 2	5	46.5159	10.5953	1907	2015-2020	Bozen
Sulden Madritsch	16	H	17	46.4938	10.6144	2825	2003-2020	Bozen
Sulden Madritsch	16	HS	11	46.4938	10.6144	2825	2009-2020	Bozen
Sulden Madritsch	16	T	17	46.4938	10.6144	2825	2003-2020	Bozen

Sulden Madritsch	16	WS	6	46.4938	10.6144	2825	2008-2014	Bozen
Ulten Weißbrunn	17	H	10	46.4868	10.8318	1900	2010-2020	Bozen
Ulten Weißbrunn	17	P	11	46.4868	10.8318	1900	2004-2015	Bozen
Ulten Weißbrunn	17	T 1	12	46.4868	10.8318	1900	1990-2002	Bozen
Ulten Weißbrunn	17	T 2	8	46.4868	10.8318	1900	2002-2010	Bozen
Ulten Weißbrunn	17	T 3	10	46.4868	10.8318	1900	2010-2020	Bozen
Umhausen	18	H	73	47.1392	10.9289	1025	1946-2019	ZAMG
Umhausen	18	P	73	47.1392	10.9289	1025	1946-2019	ZAMG
Umhausen	18	R	8	47.1392	10.9289	1025	2011-2019	ZAMG
Umhausen	18	T 1	57	47.1392	10.9289	1025	1946-2003	ZAMG
Umhausen	18	T 2	16	47.1392	10.9289	1025	2003-2019	ZAMG
Umhausen	18	WS	73	47.1392	10.9289	1025	1946-2019	ZAMG
Weißbrunnspitze	19	H	6	46.494	10.774	3253	2012-2018	Bozen
Weißbrunnspitze	19	T	8	46.494	10.774	3253	2012-2020	Bozen
Weißbrunnspitze	19	WS	5	46.494	10.774	3253	2013-2018	Bozen
Zufallhuette	20	Q	6	46.4763	10.6768	2311	2014-2020	Bozen

2.2.3. Statistical tests

The preprocessing of the time series includes the aforementioned gap-filling procedure and the aggregation to daily and monthly mean (i.e., discharge, temperature, snow depth, humidity and wind speed) or sum values (i.e., precipitation and radiation). To reduce the number of significant autocorrelation coefficients, we eliminated the deterministic periodic component from the time series through the decompose function (version 3.6.2) in R (R Core Team, 2021) that is based on moving averages. Before the statistical analysis, we subtracted the computed periodicity from each time series. Afterwards we analyzed the time series with periodicity eliminated by means of HyStat (Willems et al., 2013), a software program for statistical analysis of hydrological time series. In a statistical test, there is usually a null hypothesis, which will eventually be rejected based on the evaluation of a test statistic and its comparison with a threshold value. The rejection of the null hypothesis is controlled by the significance level α, i.e., the probability of rejecting the null hypothesis when it is true, which was set to 0.05 for all performed tests. This implies that there might be some false positives in this study, in which several tests are performed on observational data. Moreover, when applying statistical tests,

one should keep in mind the hypothesis on which they rely. For instance, most tests require that the values of a time series are not autocorrelated. Since this hypothesis is often not fulfilled for hydrological time series, it can be assumed that the rejection of null hypotheses tends to be too frequent (El-Shaarawi & Damsleth, 1988). Although the elimination of periodicity reduces the probability of a false rejection of the null hypotheses, trends and stochastic short- and long-term (Hurst, 1951) correlated components are present in the time series that can influence the test result. While there have been some efforts to adjust a few statistical tests for autocorrelation (e.g., Hamed, 2008; Hamed & Rao, 1998; Yue & Wang, 2004), this inconvenience is often not explicitly addressed (Yozgatligil & Yazici, 2016).

2.2.3.1. Homogeneity tests

The term inhomogeneity is used very differently in the hydrological and water management literature. We follow the definition of Dyck (1980), who defines inhomogeneity as a sudden or gradual change in hydrologic regime, making the assumption of a common population for the sample unrealistic. We can distinguish between relative homogeneity and absolute homogeneity tests. Relative homogeneity tests are used to identify whether the changes observed in a time series are likely caused by anthropogenic factors (Mestre et al., 2013). In this work, we focus only on absolute homogeneity tests, which analyze the internal variability of the time series since relative homogeneity tests are not available for all considered hydrometeorological variables. In addition, data scarcity in high-elevation catchments makes it difficult to find suitable time series pairs, which limits the applicability of relative homogeneity tests. Absolute homogeneity tests can be subdivided into parametric tests, where the test decision is based on the assumption that the sample corresponds to a defined parametric class of distributions, and non-parametric ones, where no such assumption about the distribution of the sample is made (Willems et al., 2013).

In HyStat, six non-parametric tests are implemented (Table 2-2). This includes the Kolmogoroff-Smirnov (Berger & Zhou, 2014), the Chi-squared (Pearson, 1900) and the Kruskal-Wallis (Wallace, 1959) tests, which verify whether the time series divided into two or three segments has different statistical properties. The Kolmogoroff-Smirnov and Chi-squared tests are based in the comparison of pairs of elements, while the Kruskal-Wallis test evaluates the central tendency of subsegments (in our cases two and three). The Cox-Stuart test (Cox & Stuart, 1955) assesses the hypothesis that no trend is present in the location or in the dispersion of the time series. Cyclical trends within a time series are checked using the Noether test (Noether, 1956). A cycle is understood as successive values grouped into blocks of similar tendency (local trends), i.e., phases of steady increase are replaced by similar long phases of steady decrease. The t-test and the equality of variance F-test are, in contrast, parametric tests (McCuen, 2016; Willems et al., 2013). The null hypothesis of the t-test is that the mean values

of the two samples are identical, while the equality of variance F-test tests for identical variances. Both tests assume that the samples are normally distributed. In addition, the t-test also assumes that the two tested samples have nearly identical variances. Because the Kruskal-Wallis test was calculated with two different configurations (for two and three subsegments), there are nine homogeneity tests in total.

Table 2-2: Key information about the performed non-parametric and parametric homogeneity tests that were performed.

Test	**Null - Hypothesis H0**	**Important aspects**
Kolmogorov-Smirnov test	Identical distributions for partial samples	Non-parametric
Chi2 test	Identical distributions for partial samples	Non-parametric
Kruskal-Wallis test (2 or 3 samples)	Identical central tendency for partial samples	Non-parametric
Cox-Stuart test - Location	No upward/downward trend in central tendency	Non-parametric
Cox-Stuart test - Dispersion	No upward/downward trend in dispersion	Non-parametric
Noether	No cyclical trend in the time series	Non-parametric
t-test	Mean value is constant	Assumption: subsamples follow a normal distribution and have the same variance
F-test	Variance is constant	Assumption: subsamples follow a normal distribution

While our recommendation is that the most adequate test should be chosen according to the properties of the time series and the hypothesis of the test, one may also want to compare and validate the results with other tests requiring stronger or weaker assumptions. In this paper, to summarize the results of the nine tests that we performed, we considered a time series reliable if no more than half of the null hypotheses (H0) of the homogeneity tests are rejected, and considered it inhomogeneous (suspicions) if more than half but fewer than eight H0 are rejected. A time series was classified as corrupted if eight or even nine H0 are rejected.

2.2.3.2. Trend analysis

The trend analysis aims to detect long-term tendencies in hydrometeorological time series. For the detection of the trends, the Mann-Kendall test, the Trend-Noise Ratio (Schönwiese, 2013) and the t-test for the slope coefficient were performed (Table 2-3). The Mann-Kendall test (Kendall, 1975; Mann, 1945) is a non-parametric test used to detect consistently increasing or

decreasing trends over time within the data and is based on an analysis of the sum of the signs of differences between all possible pairs within the sample. Because of possible short- and long-term autocorrelation we applied three additional trend tests that consider stochastic autocorrelation, namely tests for short-term autocorrelated data after Hamed and Rao (1998) and Yue and Wang (2004), as well as one for long-term autocorrelated time series (Hamed, 2008). These adjusted Mann-Kendall tests are available as the R packages modifiedmk (Patakamuri & O'Brien, 2021) and HKprocess (Tyralis, 2016). The criterion of the trend-noise-ratio method is based on the ratio of the trend value to the standard deviation, where the trend value is calculated as the difference between the first and last values of a regression equation. We computed the trend-noise-ratio with two different algorithms the FQS (minimization of the error sum of squares) and the FAS (minimization of the absolute error sum) (Willems et al., 2013). Consequently, we obtain two different results for this test. Finally, the t-test for the slope coefficient is used to check the significant deviation of the slope parameter from zero (Willems et al., 2013). While the Mann-Kendall trend test is not affected by the actual distribution of the data and is less sensitive to outliers, the parametric trend tests require data to be normally distributed and are more sensitive to outliers (Hamed, 2008).

Table 2-3: Key information about the trend tests performed. AE stands for absolute error.

Test	**Null - Hypothesis H0**	**Important aspects**
Mann-Kendall Test (MK)	There is no significant trend (α =0.05)	Non-parametric test, requires uncorrelated data, not affected by the distribution of data
MK after Hamed and Rao (1998)	There is no significant trend (α =0.05)	Variance correction approach to address the issue of serial short-term correlation in trend analysis. ESS with an empirical formula
MK after Yue and Wang (2004)	There is no significant trend (α =0.05)	Variance correction approach to address the issue of serial short-term correlation in trend analysis. ESS with Monte Carlo simulations
MK after Hamed (2008)	There is a significant trend based on the estimated scaling coefficient *H*	Considers the effect of long range dependence in trend analysis (scaling hypothesis for the data)
Trend-noise ratio	There is no significant trend [AE(T/N) > 1 corresponds to α=0.2 and AE(T/N) > 2 corresponds to α=0.05]	For normally distributed data. Trend line approximation with two different methods minimization of the error sum of squares and minimization of the absolute error sum
t-test for the slope coefficient	There is no significant trend (the slope coefficient is zero)	For normally distributed data. This test is used to check the significant deviation of the slope parameter b from zero

2.2.3.3. Change points

The analysis of change points allows us to detect whether the time series present abrupt changes in measures like the central tendency. The change point test detects a particular point of inhomogeneity in a time series. We applied the Bernier (Bernier, 1994), Pettitt test (Pettitt, 1979) and local jump method (Willems et al., 2013), as listed in Table 2-4. The Bernier test is a parametric approach that identifies the most likely change point as the point of the time series corresponding to the maximum cumulative deviation between its logarithmic value and the average of all its logarithmic values. The non-parametric Pettitt method computes a sum of signs of differences between the values before and after a given location n. The most probable change point location will be the one maximizing the sum. Another parametric method is the local jump method, which assumes that a time series X can be represented as a superimposition of a continuous function g() with a jump function, leading to points of discontinuity (Willems et al., 2013). While the Bernier and Pettitt tests can only detect one change point per time series, the local jump method can detect multiple change points.

Table 2-4: Key information about the parametric and non-parametric change point tests that were performed.

Test	Null - Hypothesis H0	Important aspects
Bernier test	There is no significant change point	Parametric test
Pettitt test	There is no significant change point	Non-parametric test
Local jump methods	There is no significant change point	Parametric test

2.2.4. Hydrological model

We used the physically-based and fully distributed hydrological model WaSiM (version 10.05.05) as a test case to evaluate to what extent the quality of the meteorological time series influences model results. The spatial model resolution of the Horlachtal case study was 25 m with daily time steps for the period 1990-10-01 to 2019-12-31. However, the beginning of the evaluation period was shifted to 1992-10-01 to consider an initialization phase of the model. The spatial and temporal model resolution of the Martelltal case study was the same, but the modeling period differed. The simulation period was from 1973-10-01 to 2020-09-30 and the beginning of the evaluation was shifted to 1974-10-01 to consider an initialization phase. For both case studies, important WaSiM modules are the interpolation of meteorological input, the snow module, the evapotranspiration and the soil module. The evapotranspiration was computed with Penman-Monteith for Horlachtal and with Hamon (1961) for Martelltal, due to their differing data availability. For both case studies, the water dynamics in the unsaturated zone were calculated with the Richards equation. Similar to the evapotranspiration, the interpolation and snowmelt methods differ due to the available meteorological input data. The

larger data availability in Horlachtal in comparison to Martelltal enables the computation of snowmelt with the enhanced energy balance approach, where the energy balance is solved at the snow pack surface. In addition, gravitational and wind-driven snow redistributions are simulated according to Warscher et al. (2013). Since there are no shortwave radiation observation data available in Martelltal before 2006, snowmelt was simulated with the classical temperature index method based on time-variable snowmelt coefficients (degree-day). For the time variable degree-day factors, we used the standard parameter provided in the WaSiM user manual (Schulla, 2021). We applied different meteorological interpolation methods dependent on the altitudinal distribution of the stations. First, altitude-dependent regression was used to determine the mean temperature lapse rate based on the period with the largest data availability, i.e., Horlachtal -0.58 K/100 m for 2010-2019 and Martelltal -0.48 K/100 m for 2012-2020. Secondly, the spatially interpolated temperature based on the nearest neighbor is adjusted by respective constant lapse rates for the simulation runs (Horlachtal 1991-2019 and Martelltal 1973-2020). The same approach was used to define an appropriate precipitation lapse rate for Martelltal with 0.18 mm/100 m for 2012-2020, and to correct the interpolated daily precipitation by the constant lapse rate for the simulation run. For the interpolation of precipitation in Horlachtal, we used a simple inverse distance weighting (IDW). A 20 % offset is added to solid and liquid precipitation in order to account for systematic undercatch, especially during snowfall events (Rasmussen et al., 2012). The topography-dependent adjustment of radiation and air temperature follows the scheme devised by Oke (2002). Glacier melt was simulated with a classical temperature index method in Martelltal, while we could apply the extended temperature index method after Hock (1999) for the Horlachtal case study, which includes information on the global radiation during each time step on each grid cell in order to modify the melt. Moreover, WaSiM is also able to simulate glacier dynamics as a growing or shrinking surface after Stahl et al. (2008). Table 2-5 summarizes the main model parameters used for the two case studies.

Table 2-5: Main WaSiM parameters for meteorological interpolation, snow accumulation and ablation.

Process	WaSiM parameter	Description	Values Horlachtal	Values Martelltal
Interpolation of meteorological variables	r_{temp}	Temperature lapse rate (K/m)	-0.0058	-0.0048
	r_{prec}	Precipitation lapse rate (mm/m)	-	0.001848
	hv,l	Lower inversion for altitude dependent regression (m.a.s.l.)	2000 (Temp)	2000 (Temp)
			-	2200 (Prec)
	hv,u	Upper inversion for altitude dependent regression (m.a.s.l.)	2100 (Temp)	2100 (Temp)
			-	2600 (Prec)
Correction of Precipitation	*Snowa*	Correction parameters for liquid precipitation (-)	1.2	1.2
	Snowb	Correction parameters for solid precipitation (-)	1.2	1.2
Snow accumulation	*T0R*	Temperature limit for rain (°C)	0	0
	Ttrans	½ of temperature-transient zone for rain-snow (°C)	0.5	0.5
Snow ablation	*C0*	Time-variable degree-day factor (mm/d/°C) from January to December	-	1.4, 1.4, 1.4, 1.6, 1.8, 1.9, 1.9, 1.8, 1.8, 1.8, 1.4, 1.4
	T0	Temperature for beginning with snowmelt (°C)	-	0.5
	LWINcorr	Correction factor for incoming long-wave radiation (-)	1	-
	LWOUTcorr	Correction factor for outgoing long-wave radiation (-)	1	-

The WaSiM Horlachtal setup achieved a Nash-Sutcliffe model efficiency coefficient (NSE) of 0.72, Kling-Gupta efficiency (KGE) of 0.75, Root Mean Square Error (RMSE) of 0.56 m^3/s and percent bias (PBIAS) of -21.3 % at the stream gauge Horlach Fassung for daily time steps over a period from 2006 to 2018. The model performance indices show a systematic underestimation of simulated streamflow, which is predominantly due to the precipitation underestimation at meteorological station Horlachalm, as shown in the analysis in the following results section. The WaSiM Martelltal setup achieved a performance with NSE of 0.75, KGE of 0.69, RMSE of 1.27 m^3/s and PBIAS of -23 % at the stream gauge Zufallhuette for the period 2014-2020. The underestimation of simulated streamflow results from the general limitation of having only one precipitation station near the catchment outlet to adequately represent the spatio-temporal precipitation distribution in this very complex terrain. Since the focus of this study was to examine the hydrological response of various meteorological forcings with different properties based on our statistical analyses, we did not perform extensive model calibration and validation. The model parameters were adopted either from the WaSiM user manual (Schulla, 2021) or from previous studies (Förster, Garvelmann, Meißl, & Strasser, 2018; Hofmeister, Arias-Rodriguez, et al., 2022). Before we analyzed the simulated time series, we removed the periodicity from the daily time series with the same method (i.e., decompose function) as for the observed time series.

2.3. Results

2.3.1. Statistical results – entire dataset

Figure 2-2 summarizes the main statistical results obtained for the entire dataset. The quality of the data increases significantly from daily time series to monthly time series, as illustrated in Figure 2-2a, since the null hypothesis for the homogeneity tests is not rejected much more often for monthly time series than for daily. Figure 2-2b shows that trends are detected for 25 time series. However, trends are never detected univocally by all tests. Figure 2-2c shows that 26 out of the 59 time series have at least one statistically significant change point. In Table 2-6, we provide the classification of the time series into reliable, suspicious and corrupted for each variable and each study area. The most reliable time series in our dataset are those of precipitation, which are also generally affected by less short- and long-term autocorrelation in comparison to the other variables. All snow depth time series are classified as corrupted. Generally, the data collected in Horlachtal seem to be more reliable than the dataset available for Martelltal.

Table 2-6: Total analyzed time series per meteorological variables classified in the three categories reliable, suspicious and corrupted, where H is relative humidity, P is precipitation, R is radiation, T is temperature, WS is wind speed, HS is snow depth and Q is discharge. The number of time series for each valley is given in parentheses (HT for Horlachtal and M for Martelltal).

Variable	H	P	R	T	WS	HS	Q	Total	%
Reliable time series daily	1 (HT1/M0)	6 (HT2/M4)	1 (HT1/M0)	3 (HT2/M1)	1 (HT1/M0)	0 (HT0/M0)	0 (HT0/M0)	12 (HT7/M5)	20
Suspicious time series	7 (HT3/M4)	4 (HT2/M2)	2 (HT1/M1)	12 (HT5/M7)	4 (HT0/M4)	0 (HT0/M0)	2 (HT1/M1)	31 (HT20/M27)	53
Corrupted time series	2 (HT0/M2)	0 (HT0/M0)	1 (HT1/M0)	5 (HT1/M4)	3 (HT2/M1)	4 (HT3/M1)	1 (HT1/M0)	16 (HT8/M8)	27
Total tested time series	10 (HT4/M6)	10 (HT4/M6)	4 (HT3/M1)	20 (HT8/M12)	8 (HT3/M5)	4 (HT3/M1)	3 (HT2/M1)	59 (HT27/M32)	100

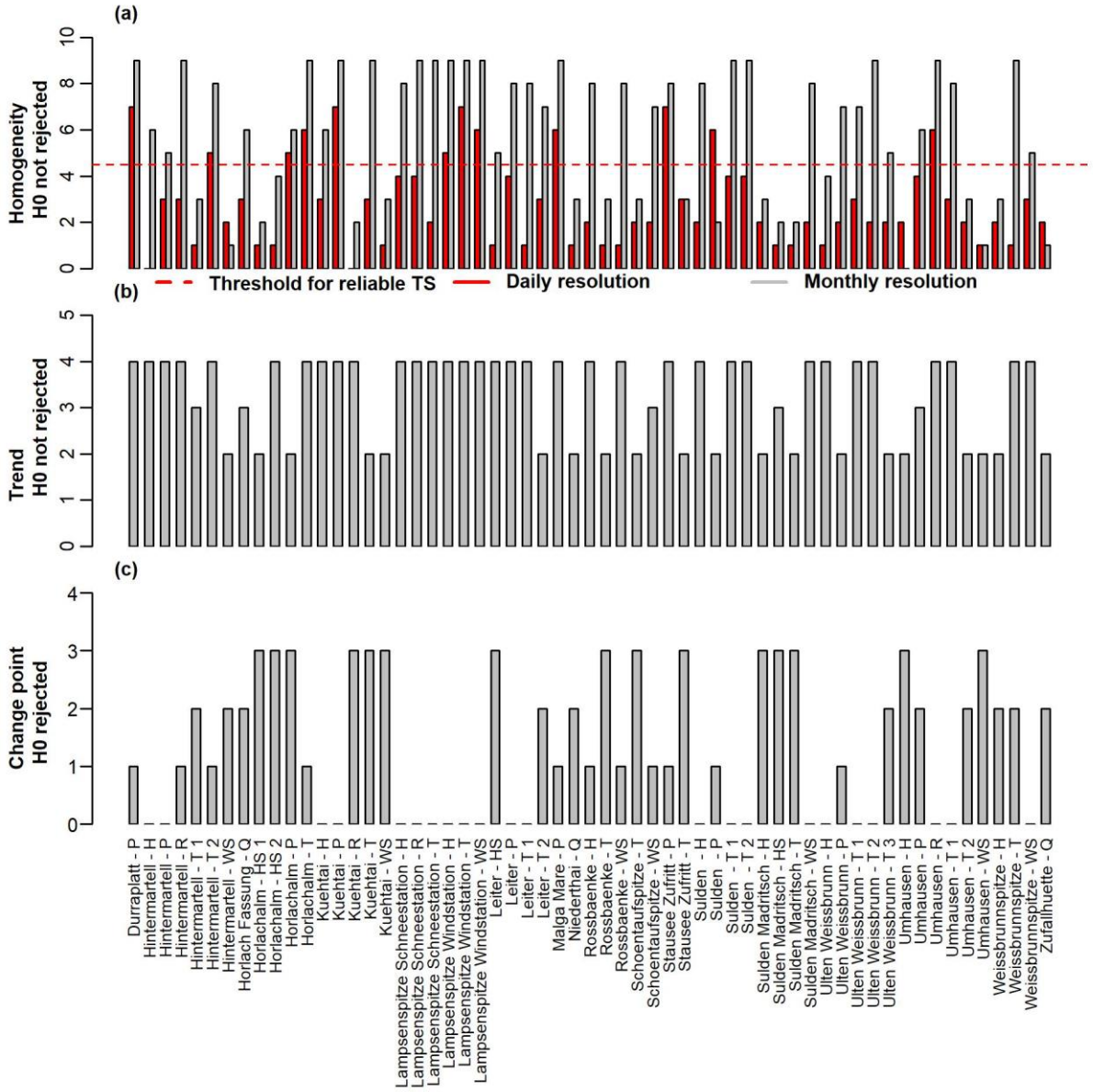


Figure 2-2: (a) Non-rejected null hypothesis (H0) of daily (red) and monthly (grey) homogeneity tests for each analyzed time series. The red dashed line shows the minimum number of non-rejected H0 that must be achieved for a daily time series to be considered reliable. (b) Non-rejected H0 of significant trend tests per monthly time series. (c) Rejected H0 of significant change points per monthly time series.

Figure 2-3 presents the frequencies of the statistical results aggregated for the hydrometeorological variables. The comparison of Figure 2-3a and b supports the results shown in Figure 2-2a. In fact, the H0 hypothesis of the homogeneity tests is not rejected for 53 % of the monthly time series, while at the daily scale H0 is generally only not rejected for precipitation time series. Moreover, discharge and snow depth can be identified as the hydrometeorological variable most prone to the occurrence of inhomogeneities. In general, we found that the elimination of periodicity leads to a higher rejection rate in the homogeneity tests compared to the test results with the original time series. Figure 2-3c shows that the occurrence of trends is particularly present for discharge, snow depth and wind speed. Discharge and

snow depth are also the hydrometeorological variables most affected by change points (Figure 2-3d).

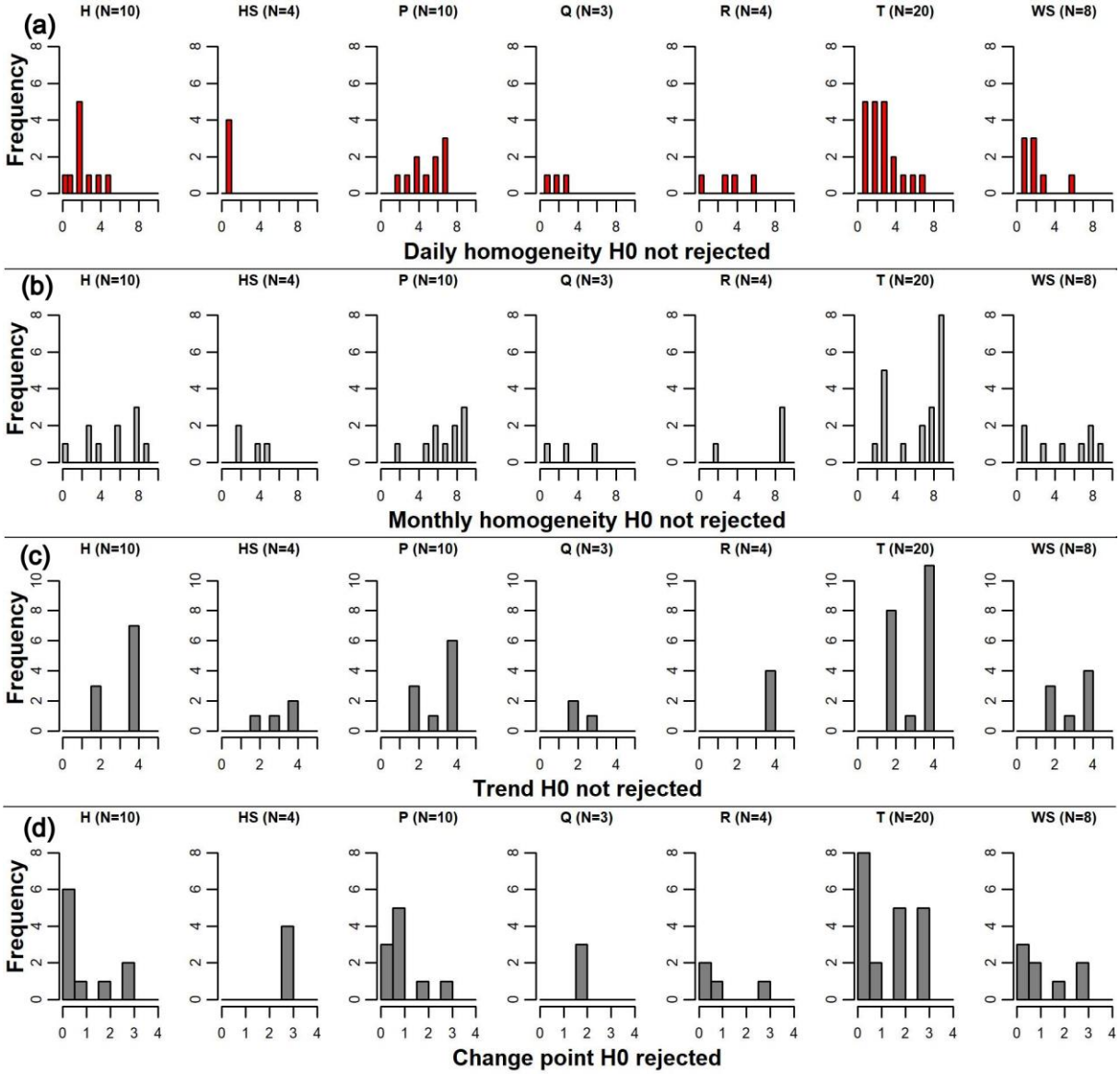


Figure 2-3: (a) Frequencies of non-rejected H0 of homogeneity tests on daily resolution per variable and (b) on monthly resolution. Frequencies of (c) non-rejected H0 of significant trend tests per variable and (d) rejected H0 of significant change points per variable.

While specific change point analysis and homogeneity tests have not been developed to explicitly consider autocorrelation, some trend tests attempt to do this. Hence, we have tested the effect of autocorrelation on the trend test considering only time series longer than 30 years (i.e.: snow depth time series of Leiter, two discharge time series from Horlach Fassung and Niederthai, temperature time series of Zufritt Stausee and precipitation timer series of Horlachalm). The test results are listed in Table 2-7. With the exception of the Leiter snow depth time series, for which all test algorithms detected no trend, a significant trend was detected for all other time series using the standard Mann-Kendall test. The Mann-Kendall test adjusted for long-term autocorrelation after Hamed (2008) only detected a significant trend in

the temperature time series of Zufritt Stausee, which was confirmed by the other tests. Furthermore, we found significant trends in the discharge time series of Niederthai and precipitation of Horlachalm with the adjusted trend test for short-term autocorrelation after Hamed and Rao (1998). The second Mann-Kendall adjusted for short-term autocorrelation after Yue and Wang (2004) detected significant trends in all time series except snow depth of Leiter. However, it should be noted that the discharge at stream gauge Niederthai is additionally affected by an artificial upstream water diversion at Horlach Fassung, which means that the trend is more likely due to an anthropogenic influence.

Table 2-7: Significant trends detected by standard Mann-Kendall (MK) test without considering autocorrelation and with compensation for autocorrelation after different approaches are applied to monthly time series.

Time series	Significant trend after MK	Significant trend after Hamed (2008)	Significant trend after Hamed and Rao (1998)	Significant trend after Yue and Wang (2004)
Q Niederthai	Yes	No	Yes	Yes
HS Leiter	No	No	No	No
Q Horlach Fassung	Yes	No	No	Yes
T Zufritt	Yes	Yes	Yes	Yes
P Horlachalm	Yes	No	Yes	Yes

2.3.2. Statistical results – Horlachalm precipitation and Zufritt temperature time series

In the following section, we focus on two specific time series of the database: the Horlachalm precipitation (Figure 2-4a) and the Zufritt temperature time series (Figure 2-4c). In the first case, the precipitation time series of Horlachalm station showed a significant change point in monthly precipitation in November 2000 detected by Bernier and Pettitt tests and a significant local jump in October 2007 (see Figure 2-4b). Moreover, half of the trend tests indicate a significant positive trend for the monthly precipitation sums. Four out of nine H0 of the homogeneity tests were rejected for daily precipitation and the time series was therefore not initially recognized as a suspicious case.

In the second case, the daily temperature time series from the Stausee Zufritt station (Martelltal) had only two non-rejected H0 out of nine homogeneity tests and three with monthly frequency. Half of the trend tests indicate a significant positive trend in the mean monthly temperature as well. In addition, the mean monthly temperature time series has a significant change point in August 1987 detected by Bernier and Pettitt as well as a local jump in May

2006 (see Figure 2-4d). Comparing the mean monthly temperature before and after the change point detected in 1987, we observed a difference in the mean temperature of 1.67°C.

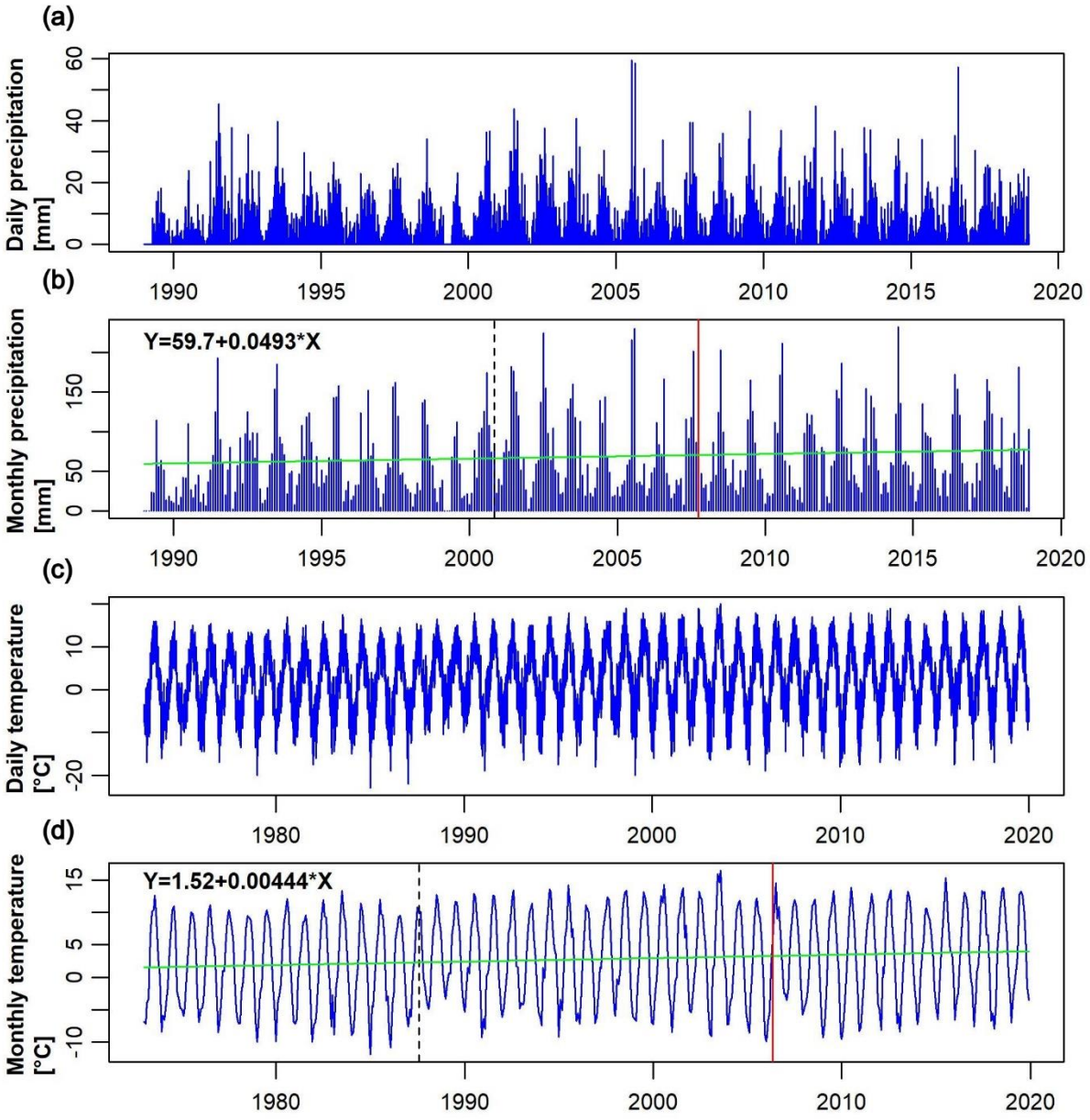


Figure 2-4: (a) Daily and (b) monthly precipitation sum at Horlachalm station as well as (c) daily and (d) monthly mean temperature at Zufritt station. The vertical dashed line shows a significant change point detected by the Bernier and by Pettitt tests, while the solid line indicates a significant local jump. The horizontal line indicates a significant positive trend detected by standard Mann-Kendall and t-test estimated with FQS. In addition, estimated parameters of the trend equation are given in the respective sub-plot.

2.3.3. Horlachtal case study

To assess the impact of suspicions time series on the results of the WaSiM hydrological model, we forced the model with and without consideration of the Horlachalm precipitation time series. Figure A.1.2a shows the impact of including or excluding the time series for the interpolation and computation of the mean catchment precipitation. In contrast to the observed monthly

precipitation at the Horlachalm station, there is no significant trend or change point in the interpolated precipitation considering or excluding Horlachalm precipitation. The difference between the daily mean catchment precipitation computed including the Horlachalm station and excluding it ranges from +20 mm to almost -20 mm and the interpolated positive precipitation anomalies (green bars) are mostly stronger in the summer period. The impact of station selection on the snow water equivalent (SWE) is evident in Figure A.1.2b, where the mean catchment precipitation computed without considering the Horlachalm station leads to higher SWE peaks than in the case of its consideration. This is particularly relevant for the periods 1995-2003 and 2013-2015, where the exclusion of Horlachalm precipitation led to a larger snow accumulation and consequently to a longer snow cover duration. In both simulated SWE time series, one significant change point was detected by the Bernier and Pettitt tests in 2004. The difference in mean SWE decreased slightly between the different precipitation forcings before (222 mm with Horlachalm and 245 mm without Horlachalm) and after the change point (157 mm with Horlachalm and 167 mm without Horlachalm). The t-test detected a significant negative trend in both simulated monthly SWE time series. The decrease in monthly SWE is larger in the case without Horlachalm (SWE=304-0.441*m, where m indicates the time in months) than with Horlachalm (SWE=262-0.338*m), which would correspond to a yearly SWE decrease of 5.3 mm without Horlachalm to 4.1 mm with Horlachalm over the period 1993 to 2019. Moreover, we used the empirical approach of Jonas et al. (2009) to estimate the SWE from snow depth observations at Horlachalm station and compared the empirical cumulative distribution function (ECDF) with the daily simulated SWE of both model configurations (Figure 2-5a). This comparison is particularly valuable to evaluate the snow model performance of the configuration without Horlachalm precipitation. Without Horlachalm, the model accumulates too much snow and is therefore far from the distribution function of the estimated SWE. On the other hand, we can observe that the configuration with Horlachalm matches the estimated SWE better, but includes a systematic underestimation of peak SWE, although we considered a 20 % undercatch of solid precipitation in the snow model.

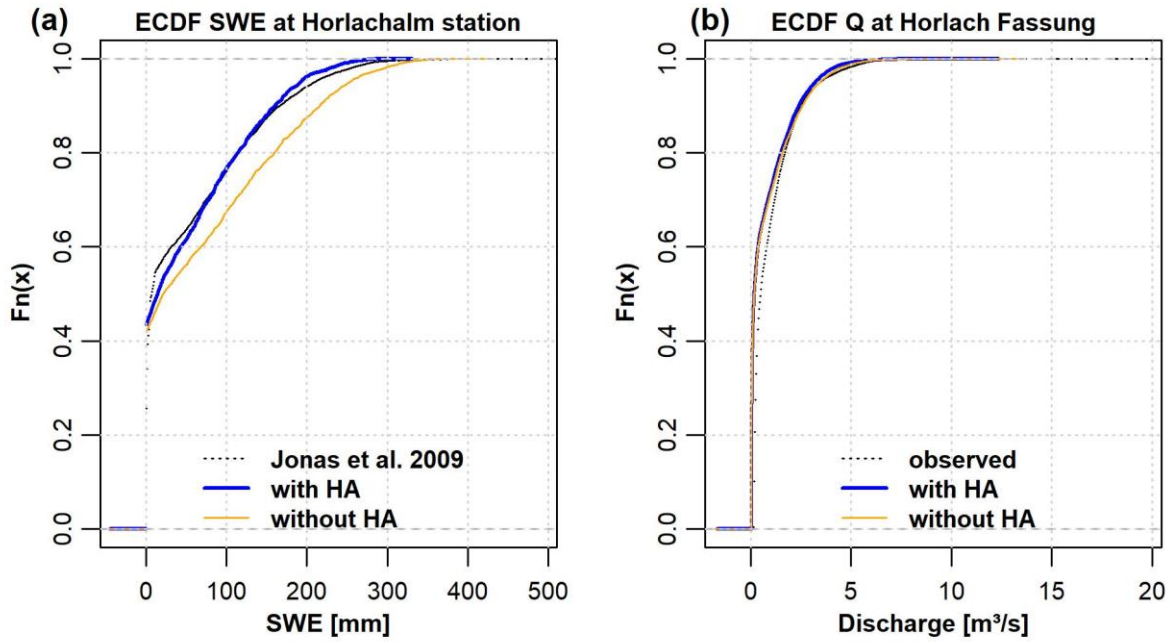


Figure 2-5: (a) Empirical cumulative distribution function (ECDF) of estimated SWE (dotted line) after Jonas et al. (2009) and simulated SWE with (thick solid line) and without Horlachalm (thin solid line) at Horlachalm station; and (b) ECDF of observed (dotted line) and simulated discharge with (thick solid line) and without Horlachalm (thin solid line) at Horlach Fassung.

An additional two-sample Kolmogorov-Smirnov test indicated that the simulated discharges do have the same pdf (p value of 0.7951), which can explain the small difference in discharges (Figure A.1.3 and Figure 2-5b). However, the differences in simulated discharge are present in both the mean discharges (1.5 m³/s with Horlachalm and 1.6 m³/s without Horlachalm) and the peak flows (24.6 m³/s with Horlachalm and 26.8 m³/s without Horlachalm). The 0.99 quantile of the discharge pdf, which corresponds to the 100-year return interval, is slightly higher for the case without Horlachalm (9.3 m³/s) than for the case with Horlachalm (8.9 m³/s). If we take a closer look at the discharge time series of the hydrological year 2002 (Figure A.1.3b), we can observe differences during the melting season (June to July) caused by the different amounts of SWE and differences in the flood event in August. Such differences also occur during other hydrological years. Both simulated discharges contain a significant change point (i.e., June 2003) and a slight negative trend.

According to the cumulative fluxes of the water balance components of Figure 2-6a, there is a constant underestimation of total precipitation when the Horlachalm precipitation time series is considered in the model forcing, which directly affects the simulated runoff. The simulated evapotranspiration sums are only slightly affected by the lower precipitation amounts, indicating a system that is mainly energy-limited. The aforementioned underestimation of solid precipitation also visible in Figure 2-6b consequently propagates to the runoff generation in June and the following months (Figure 2-6c). The timing and magnitude of the snowmelt-induced runoff are very similar for both configurations from April to June, suggesting a similar

temporal and spatial SWE distribution. However, the runoff lines diverge from the runoff peak in June onwards due to the lower snow accumulation of the model configuration with Horlachalm precipitation. The low flow period from November to April is not affected by the differences in snowfall amount or SWE.

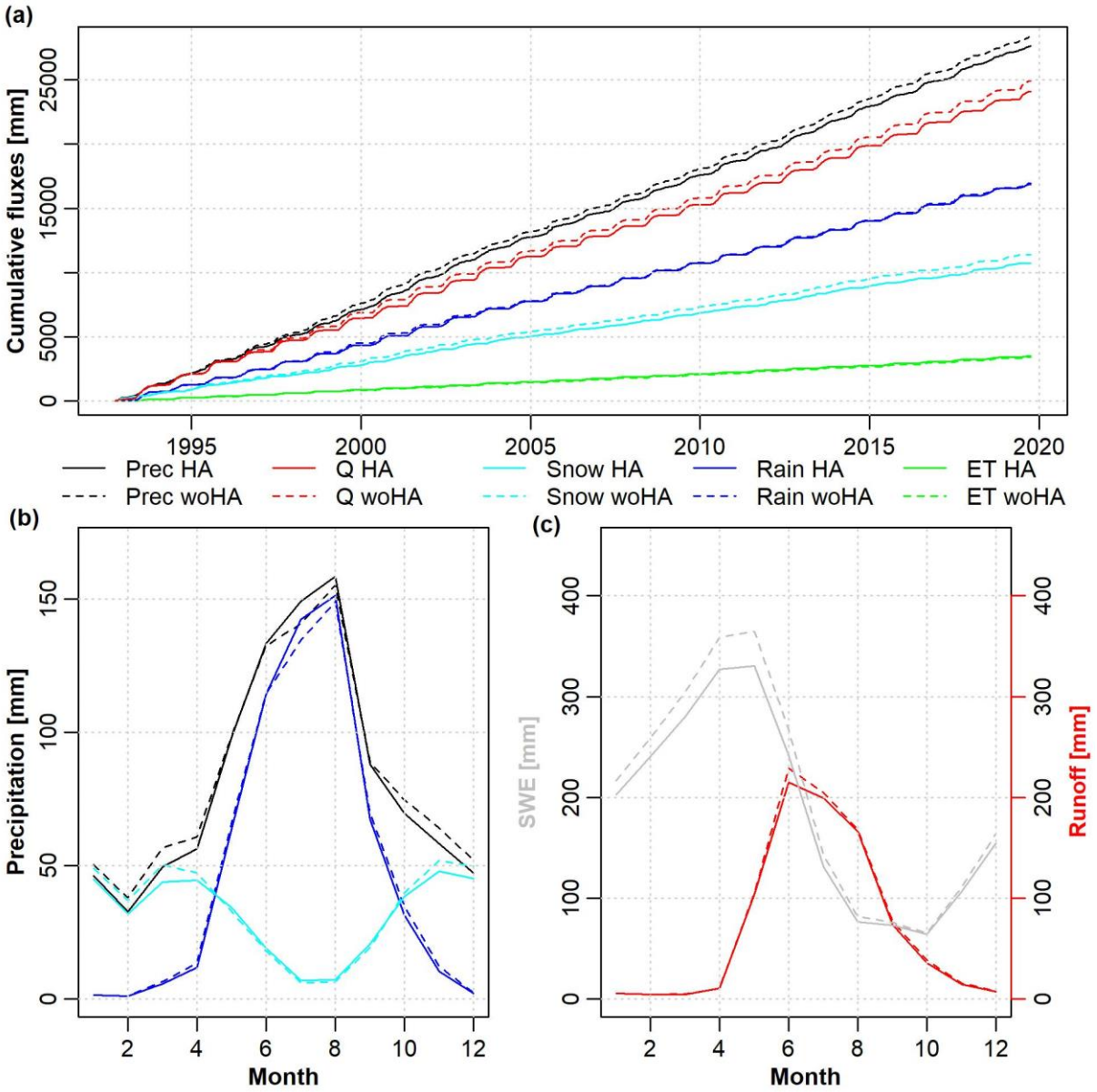


Figure 2-6: (a) simulated cumulative water balance components (Prec=precipitation (black), Q=runoff (red), Snow=snowfall (cyan), Rain=liquid precipitation (blue) and ET=evapotranspiration (green)) with (solid lines) and without (dashed lines) consideration of Horlachalm precipitation time series. (b) mean monthly simulated liquid and solid precipitation sum and (c) mean monthly SWE (grey) and runoff (red).

2.3.4. Martelltal case study

In the Martelltal case, we investigated the influence of a suspicious temperature time series on the simulated snow-hydrological response by WaSiM. The homogeneity tests had already detected a large inhomogeneity with seven rejected H0 in the daily Zufritt Stausee temperature time series; hence, including it in the model significantly affects its output. In fact, a change

point around the same year can be identified both in the interpolated catchment temperature (1987) and in the simulated SWE time series (1988). The sudden jump in temperature resulted in a comparable large difference in simulated mean SWE before (486 mm) and after (311 mm) the 1988 change point. In Figure 2-7a, we observe a sudden increase in the temperature in the catchment after 1987, which results in a decrease in the average SWE after the change point. Moreover, the standard Mann-Kendall and t-test detected a significant positive trend in interpolated monthly temperature (T=-3.17+ 0.00473*m), which propagated from the observed temperature time series of the Zufritt station. Particularly in the 1970s and the beginning of the 1980s, snow accumulation exceeds snowmelt. Nevertheless, there is a significant negative trend in simulated monthly SWE (SWE=512-0.369*m) detected by standard Mann-Kendall and t-test over the period 1973-2020. Since snowmelt is simulated with the degree-day approach in this model configuration, the temperature is the critical factor for SWE. The simulated discharge shows a significant change point in July 1997, and the standard Mann-Kendall and t-test detected a significant positive trend (Q=2.01+1.44E-03*m) in monthly discharge.

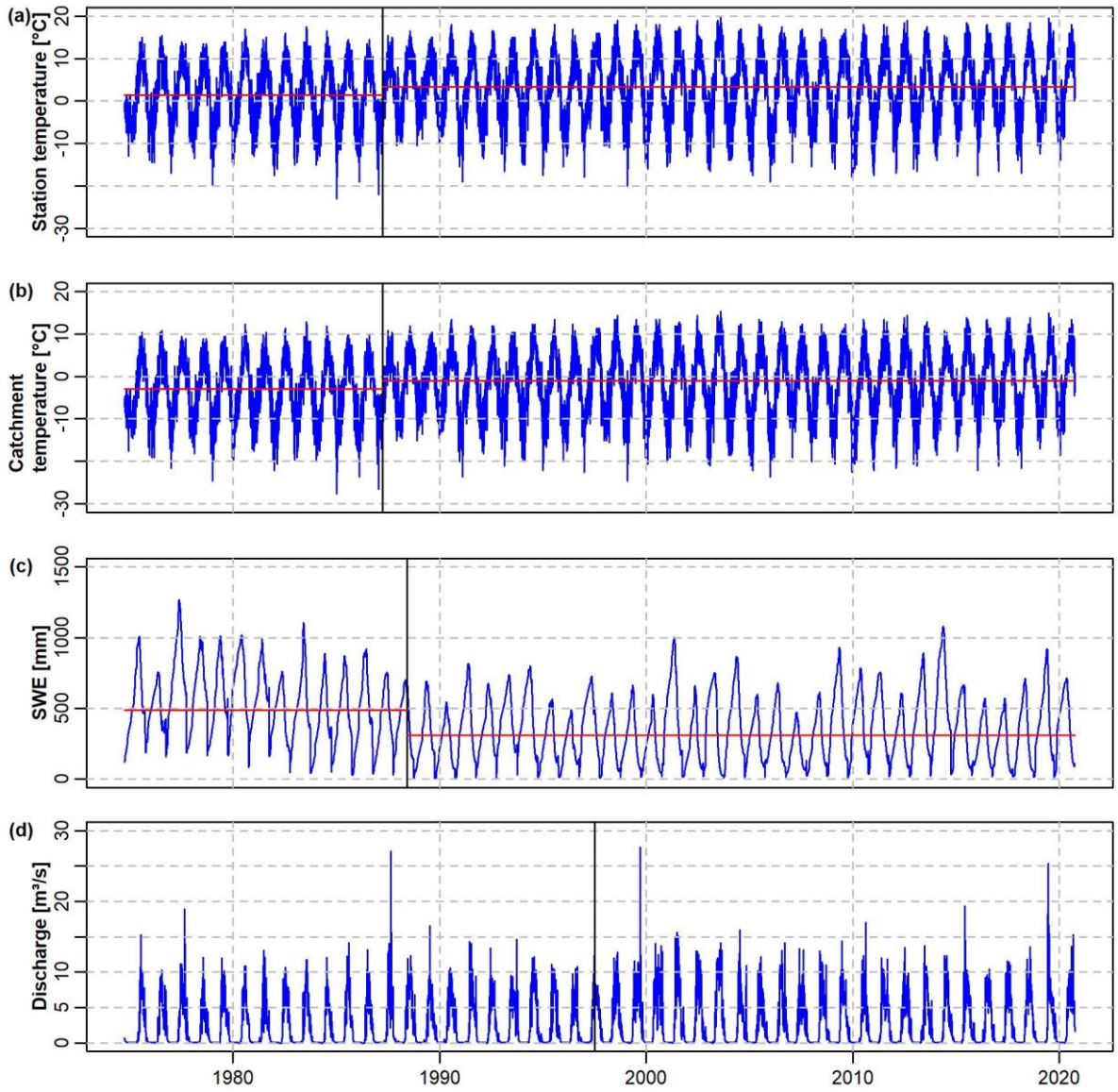


Figure 2-7: (a) Daily station temperature, (b) interpolated catchment temperature, (c) simulated snow water equivalent and (d) discharge relative to the model created using the original temperature time series of the Zutritt station. The vertical lines show the change points detected by the Bernier and by the Pettitt tests. The horizontal lines show the average of the time series before and after the change points.

We then created an adjusted temperature time series for the Zufritt station by adding 1.67°C, which corresponds to the change in the mean before and after the change point, to the daily temperature before the change point. We ran the model again with this adjusted time series and the obtained results are shown in Figure 2-8. We observe that, as expected, no change point was detected in the adjusted temperature time series and the interpolated catchment temperature. The correction of the inhomogeneity also shifted the SWE change point from 1988 to 1994 and reduced the difference in the mean before (SWE 391 mm) and after (SWE 352 mm) the change point. In contrast to the SWE simulated with the original temperature input, only a slight significant negative trend (SWE=391-0.02*m) was detected in modeled SWE. There is no trend or change point in monthly simulated discharge. The pdfs of maximum

yearly discharges have the same distribution (Kolmogorov-Smirnov test results p value = 0.6669) and slightly differ in the highest flood events, e.g., HQ100 27.4 m³/s for the original case and 29.7 m³/s with adjusted temperature time series. The difference in the high flow events can be explained in the case of uncorrected temperature before the 1988 change point by a larger share of solid precipitation, which is particularly important for the August 1987 flood event.

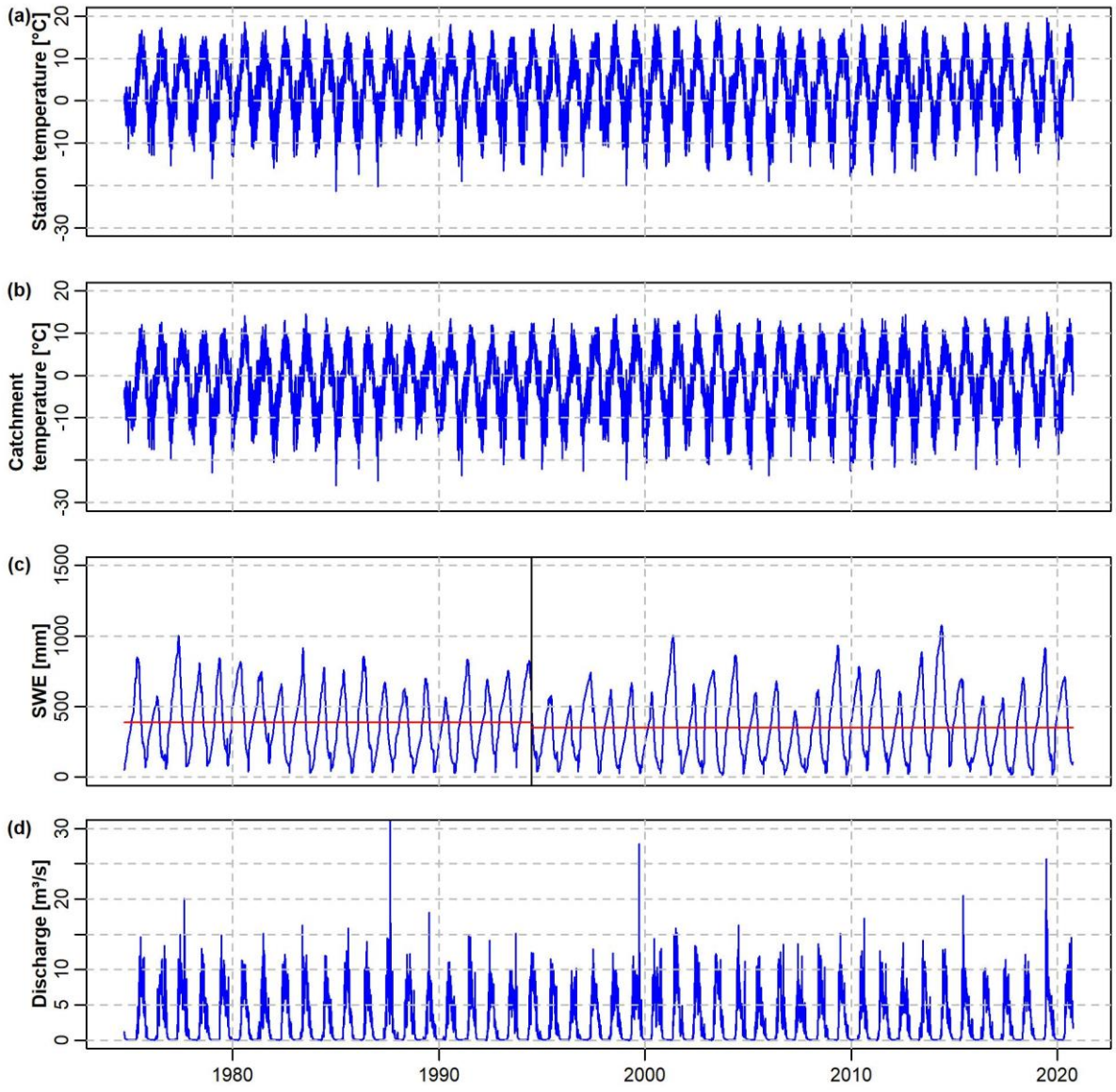


Figure 2-8: (a) Daily station temperature, (b) interpolated catchment temperature, (c) simulated snow water equivalent and (d) discharge relative to the model created using the adjusted temperature time series of the Zutritt station. The vertical lines show the change points detected by the Bernier and by Pettitt tests. The horizontal lines show the average of the time series before and after the change.

The introduction of an inhomogeneity in the temperature time series can also significantly affect the glacier mass balance, as illustrated in Figure A.1.4. In the original case, the glacier mass balance shows an increasing trend up to the end of the 1980s, followed by a decreasing trend, which slightly stabilized between 2005 and 2015. The glacier mass balance of the

adjusted case is characterized by a decreasing trend for the studied period, with a stable period between 2003 and 2015. During the European heat wave in 2003, both glacier mass balances have the largest losses. The relatively rapid glacier mass balance losses of post-1990 for the uncorrected temperature setup are a direct response to the sudden increase in temperature. Although the two simulated glacier mass balances reached a maximum absolute difference of about 4000 mm in 1988, it was reduced to about 1000 mm at the end of the simulation period in 2020.

The tendency to obtain higher simulated runoff in the original than in the adjusted model setup becomes clearer when comparing the seasonal runoff regimes before the 1988 change point (Figure A.1.5). The higher runoff is generated by both snowmelt and glacier melt. With a higher mean temperature, snowmelt starts earlier and is more intensive in spring. This results in larger snowmelt in June, which directly propagates to the runoff response. When the snowmelt rate drops in July and is exceeded by the original configuration, the larger glacier melt component provides more water for the runoff generation from August to October. The dominant runoff component before the 1988 change point is snowmelt for both cases with 932 mm for the original and 985 mm for the adjusted setup. The glacier component contributes 366 mm in the original and 505 mm in the adjusted configuration in the same period. After the 1988 change point, the snowmelt rates (Figure A.1.5b) are almost identical (original 931 mm, adjusted 911 mm) and slight differences can still be observed in the runoff from July to October, which is a consequence of a larger glacier melt contribution with the original temperature time series.

2.4. Discussion

2.4.1. Time series analysis – entire dataset

The analysis of the available time series for the two case studies in the Alps shows that the statistical tests considered in this work quite often lead to contrasting results. One reason for this is that different statistical tests often rely on different hypotheses and test different statistical properties of the time series. Statistical tests should be first chosen considering the properties of the time series and whether they satisfy the hypothesis of the test. In many cases, such hypotheses (e.g., absence of autocorrelation, normality) are often not satisfied by hydrological time series. In this case, a possible approach is to apply multiple tests on the time series, so that it can be made clearer whether some issues affect a time series even if we change the tested hypothesis.

However, as a general conclusion, we note that the results of the tests are influenced by the temporal resolution of the time series, with monthly time series being less prone to inhomogeneity than daily time series. This may be due to several reasons, such as a reduction in the significant autocorrelation for monthly time series in comparison to daily time series, the

reduced sample size of monthly time series in comparison to daily time series and the error compensation effects when data are averaged for longer temporal periods. Particularly relevant in our view is the role of autocorrelation that may lead to a higher probability of detection of inhomogeneity. In this study, the mean H0 non-rejection rate of daily homogeneity tests was 32 % and 66 % for monthly homogeneity tests, respectively.

Considering the results obtained for the different hydrometeorological variables, we observed that discharge and snow depth are the most subject to inhomogeneities, trend and change points. This is particularly relevant for hydrological modeling studies, since these two variables are commonly used for model calibration and validation. In contrast, precipitation measurements did not show particular statistical deficiencies although they are often associated with large uncertainties in Alpine areas (e.g., winter solid precipitation is systematically underestimated due to strong winds).

2.4.2. Time series analysis – Horlachalm precipitation and Zufritt temperature time series

For the Horlachalm precipitation, no inhomogeneity was detected in 66 % of the tests at a daily resolution, and only the trend and change point analysis highlighted possible problems for this station. The adjusted Mann-Kendall tests for short-term autocorrelation confirmed the presence of a significant trend in monthly precipitation sums. Only the Mann-Kendall test adjusted for long-term autocorrelations detected no significant trend (see Table 2-6). However, none of the nearby precipitation stations showed any trend in monthly precipitation sums. From this example we can draw two conclusions: (i) using multiple different tests is beneficial to evaluate the quality of a time series and (ii) the comparison with nearby stations allows the interpretation of detected anomalies.

This second conclusion is also relevant for the analysis of the temperature time series of Stausee Zufritt where the interpretation of the homogeneity, trend and change point results is more difficult due to the lack of metadata and a physical explanation for the behavior of the time series. The sudden increase in mean temperature and the simultaneous decrease in snow depth in the Alpine region at the end of the 1980s has already been identified in previous studies (Marty, 2008; Gobiet et al., 2014; Beniston et al., 2018; Olefs et al., 2020; Michael Matiu et al., 2021; Pepin et al., 2022). Several studies have examined the reason for the elevation dependent warming in the Alpine region at the end of the 1980s (Manara et al., 2015; Philipona, 2013; Tudoroiu et al., 2016; Zeng et al., 2015). One explanation is the distinct decline of anthropogenic aerosols in Europe since the mid-1980s that lead to solar brightening at low elevations (500 m.a.s.l.) whereas the inherently low aerosol concentrations at high elevations (2200 m.a.s.l.) led to only minor changes in solar radiation in the Alps (Manara et al., 2015; Philipona, 2013; Zeng et al., 2015). Since most of the meteorological stations with

long observational records (> 30 years) are located in lower elevation areas, incorrect conclusions for the higher-altitude areas (above 2200 m.a.s.l.) may be drawn from observed trends in the valleys. This especially concerns the temperature extrapolation from lower stations via fixed lapse rates.

2.4.3. Hydrological model results

The consideration or exclusion of suspicious time series is a difficult step, since it can make the setting up of a hydrological model impossible or lead to the loss of important spatial and temporal information. In this work, we presented two practical examples focusing on two main hydrometeorological variables: precipitation and temperature. Starting with precipitation, we can observe the highly non-linear impact that the exclusion of a station has on the hydrological cycle modeled with WaSiM. Although the Horlachalm precipitation time series displays a significant positive trend in monthly precipitation sums, the computed monthly SWE time series displays a significant negative trend. This negative trend is consistent with the observed decreasing snow depth trend at Leiter station in Horlachtal and the results of Matiu et al. (2021). The comparison between model results and observations shows that considering the precipitation of Horlachalm station is particularly important to simulate the spatio-temporal snow accumulation in this area (see Figure 2-5) and to avoid overestimation of snow accumulation. Similar to the simulated SWE, the simulated discharge time series contain a slight negative trend and a change point. Since the differences in the discharge CDF (Figure 2-5b) and peak flows are not particularly large, it can be assumed that the small-scale snow processes are more sensitive (Figure 2-5a) to changes in the precipitation input than discharge. We can explain this result by the fact that, although the glacier coverage is relatively low in Horlachtal, the glacier melt can partially compensate for the decrease in snowfall and snow melt in the south-oriented side valleys in August and September. This shows that due to the complexity of hydrological processes in high-elevation Alpine areas, discharge can only partially be used to validate a hydrological model and a multi-objective approach is more suitable to assess the correct representation of the complex interaction among the different hydrological compartments.

A similar conclusion can be reached considering the Martelltal case study in which a jump in the temperature time series is present. Moreover, in the Martelltal case, we discuss whether and how we should correct change points and what the consequences of this choice are. As we consider the use of temperature-index models, as expected, we observe important differences in snow and glacier dynamics between the Martelltal models. The change point in temperature and consequently in SWE did propagate to the simulated discharge as well, however, the change point in the discharge occurred with a delay of nine years (i.e., 1997). For the case with the adjusted temperature input, the change point in discharge is no longer

present and the one in SWE was shifted from 1988 to 1994. Trend detection in the model results is also affected by the correction of the change point present in the temperature time series. While the original simulated SWE and discharge showed a significant negative and positive trend, respectively, only SWE still shows a significant negative trend, although with a much gentler slope. Through the adjusted increase in mean temperature before 1988, the snowfall lines were shifted to higher altitudes, which led to a higher share of liquid precipitation and consequently to a higher peak discharge. Moreover, we should also consider that the presence of change points and trends in the available time series affects the spatial extrapolation of meteorological input to higher altitudes for the hydrological model. This is particularly relevant for Alpine catchments and simulations lasting several decades, where observations from low elevations often have to be transferred to higher elevations, through fixed lapse rates, even though stations at different elevations may have different change points and trends (e.g., Carturan et al., 2019). Again, we suggest that model results should be critically examined with the inclusion of additional observational products such as glacier mass balances, snow depth and coverage beside river discharge.

2.5. Conclusion

Many statistical tests are available to test hydrometeorological time series. Their results are often complementary, provide a broader view of the statistical properties and reliability of a time series and support its proper use in hydrological analysis and models. Unfortunately, contrasting results may emerge when using different tests. This complicates the interpretation of the results. Therefore, inhomogeneities, trends and change points should be critically evaluated also considering their physical interpretation and comparing them with the behavior of nearby correlated stations (e.g., relative homogeneity tests if available for the variable of interest).

We further observed the effect of autocorrelation on the results of many commonly used tests (e.g., Mann-Kendall). Although we removed the periodicity from each time series, autocorrelation is still pronounced at high temporal resolution (< month) and streamflow, temperature and snow depth tend to have higher autocorrelation than precipitation or wind speed in the two Alpine case studies. In general, the autocorrelation periodicity leads to a larger share of non-rejections of the homogeneity tests. Additionally, the standard Mann-Kendall test detected more significant trends than the tests adjusted to consider short- and long-range autocorrelation. Thus, we recommend not only removing the deterministic periodic component from the time series, but also using tests that can account for stochastic autocorrelation.

The high non-linearity of hydrological processes greatly complicates the judgment about the consequences of including or excluding suspicious or even erroneous time series from the setup of hydrological models and the evaluation of their results. This problem is particularly relevant in data-scarce regions or regions with high spatial variability in hydrometeorological variables, such as Alpine catchments. Excluding a time series can lead to a loss of valuable spatio-temporal information, with consequences for the simulation results. Therefore, maintaining high-quality monitoring networks and testing the available time series before using them in hydrological models is necessary to increase the reliability of their results and of the outcomes of subsequent interdisciplinary analysis (e.g., vegetation studies, sediment transport).

Finally, the analysis of hydrological model results based only on discharge time series does not lead to a broad understanding of the system dynamics since discharge integrates information on all hydrological processes. We recommend a multi-objective framework, in which at least snow and glacier dynamics are considered in modeling Alpine catchments, to better understand the response of the catchment towards changes in the climatic input time series.

3. Automated streamflow measurements in high-elevation Alpine catchments

Hofmeister, Rubens Venegas, et al., 2023[2]

Abstract

Salt dilution is a well-established streamflow measurement method in creeks, which works particularly well downstream of turbulent flow sections as the mixing of the salt tracer is enhanced. Usually, salt dilution measurements are performed manually, which considerably limits the observations of rare peak flow events. These events are particularly important for constructing robust rating curves and avoiding large uncertainties in the extrapolation of streamflow values. An additional challenge is the variability of the river cross-section, especially after larger discharge events, leading to non-stationary rating curves. Therefore, discharge measurements well distributed over time are needed to construct a reliable streamflow-water level relationship and to detect changes caused by erosion and deposition processes. To overcome these two issues, we used an automated streamflow measuring systems at three different sites with contrasting hydrological and hydraulic characteristics in the Alps. This system allowed us to measure discharge at nearly maximum flow of the observation period (2020-2021) at all three sites and to detect abrupt changes in the rating curve by performing event-based salt injections. The uncertainty in the measurements was quantified and the streamflow was compared with official gauging stations in the same catchment. Based on a very large data set of almost 300 measurements, we were able to evaluate the reliability of the system and identify the primary sources of uncertainty in the experimental setup. One key aspect was the site selection for the downstream electrical conductivity sensors, as measurement location strongly controls the signal-to-noise ratio (SNR) in the recorded breakthrough curves.

[2]Material from:
Hofmeister, F., Rubens Venegas, B., Sentlinger, G., Tarantik, M., Blume, T., Disse, M., & Chiogna, G. (2023). Automated streamflow measurements in high-elevation Alpine catchments. River Research and Applications, rra.4203. doi: 10.1002/rra.4203.

3.1. Introduction

Measuring and recording water level and streamflow of rivers is essential for the dimensioning of water management facilities and for the rational management of water supply as well as for the simulation of hydrological processes with the help of models. While these are basic hydrological measurements, they come with additional challenges in Alpine settings: extremely dynamic fluctuations in streamflow combined with often unstable stream cross-sections. Manual streamflow measurements are generally very limited in these catchments because they depend on the frequency of site visits, which are often sparse due to unfavourable weather and snow conditions. This is in direct conflict with the here especially pronounced need for high-resolution data resulting from the high variability of streamflow in space and time (Morgenschweis, 2018), the often short runoff concentration time (Mutzner, Tarolli, Sofia, Parlange, & Rinaldo, 2016; Simoni et al., 2011), and distinct seasonal characteristics (i.e., low flows in winter and high flows in summer) (Mutzner et al., 2015). Moreover, sub-daily discharge variations are typically caused by snow and glacier melt cycles in spring and summer (Mutzner et al., 2015; Weijs, Mutzner, & Parlange, 2013). Regular discharge measurements are furthermore needed to frequently validate the correctness of the rating curve (i.e., the relation between river stage and streamflow) since rating curves are affected by changes in the river cross-section caused by high sediment and coarse bedload transport of glacio-nival rivers (Comiti et al., 2019; Weijs et al., 2013). Morgenschweis (2018) recommends 10 to 12 discharge measurements per year for natural cross-sections with loose sediments. Additionally, regular and high-quality streamflow measurements are particularly important to quantify the uncertainty in rating curves and to understand better how the uncertainty propagates from the measurement to the rating curve (Kiang et al., 2018).

Tracer-based methods (e.g., salt) have been used for the derivation of rating curves in Alpine rivers for many decades (C. M. Allen & Taylor, 1923; Moore, 2005; Østrem, 1964). Although a range of water-soluble hydrological tracers can be used (Leibundgut et al., 2011; Morgenschweis, 2018), food-grade table salt (NaCl) is preferred as it is generally non-toxic at the concentrations typically involved in stream gauging (Morgenschweis, 2018; Resources Information Standards Committee (RISC), 2018; Sentlinger et al., 2019). Moreover, the salt injection method is a reliable and relatively cheap technique for measuring discharge in small streams (Gottardi, Maglionico, & Bolognesi, 2006). Although tracer-based methods are simple and relatively easy to use, two main requirements have to be met (i.e., complete mixing of the tracer and mass conservation). Limitations can occur in the application for example due to absorption of salt tracer in river sections with a lot of aquatic vegetation, stream water exfiltration to the groundwater and riparian zone, or delay of the breakthrough curve due to pools in the river section (Clow & Fleming, 2008; Moore, 2005). A detailed summary of the method and the individual sources of uncertainty can be found in Resources Information

Standards Committee (RISC) (2018). If the main requirements are fulfilled, the salt dilution is very accurate with a measurement uncertainty about five % (Richardson, Moore, et al., 2017; Richardson, Sentlinger, et al., 2017). Hauet (2020) developed a first complete framework for uncertainty quantification of salt dilution discharge measurements following the GUM (Guide to the expression of uncertainty in measurement) method that takes into account all uncertainty sources. Another limitation of the salt dilution method is the measurement of discharges larger than 10 m^3/s, since the required amount of salt (i.e., between five to 10 kg) can be difficult to be diluted and injected within the time of the event (Richardson, Sentlinger, et al., 2017).

While salt dilution is the discharge measurement method of choice for Alpine systems, we are still left with the challenge of monitoring a system with pronounced dynamics and fast responses under conditions of limited or difficult accessibility. Our proposed solution for this dilemma is the automatization of the salt dilution method. This would remove the need for continuous access to the field sites and would provide a much higher data density and thus, lower uncertainty of the stage-discharge curve than would be possible with manual measurements.

In this work, we test an automatic salt dilution system for the derivation of robust rating curves in three Alpine catchments characterized by different background electrical conductivities, glaciated areas, streambed gradients, and ranges of discharge. This innovative measurement device allowed us to perform event-based salt injections to capture rare events and to detect abrupt changes in the rating curve. To the best of our knowledge, this is the only commercial automatic salt injection system available. Moreover, we discuss how uncertainties in the measurements can be quantified and reduced.

3.2. Material and methods

3.2.1. Research areas and site descriptions

The research sites are located in three high-elevation Alpine catchments in Tyrol/Austria (Horlachtal and Kaunertal) and South Tyrol/Italy (Martelltal). In addition to the already existing stream gauges (triangles in Figure 3-1) operated by the Tiroler Wasserkraft AG (Kaunertal and Horlachtal) and the Hydrological Office of South Tyrol (Martelltal), we installed new gauges for estimating runoff contribution from side valleys (in the case of Kaunertal and Horlachtal) or at the outlet of the upper Martelltal (see Figure 3-1 for locations and setup). Each gauge consists of an OTT CTD probe inserted in a two-inch aluminium pipe attached to larger boulders in the creek. The measurement sites were selected at river sections with cross-sections that were assumed to be stable and not too turbulent flows to increase accuracy of transducer readings. Pool location behind larger rock is ideal to protect the probe from turbulences and debris (U.S. Environmental Protection Agency (EPA), 2014). The sensors log mean values of water level [m], water temperature T [°C] and temperature compensated electrical conductivity EC_T [µS/cm] to 25 °C at 15-minute resolution. The OTT CTD sensor accuracy of the actual electrical conductivity EC [µS/cm] is ±0.5 %, ±0.1°C for temperature and ±0.05 % full scale (FS) for water level (OTT, 2023). We installed an automated salt injection system (AutoSalt, Fathom Scientific Ltd.) at the same location as the OTT CTDs in order to collect streamflow measurements at different water levels by an automatic event-based configuration of the device. The AutoSalt in Kaunertal was in operation for one year, while the other sites collected data for two years (Table 3-1). The AutoSalt in Horlachtal was moved to another sub-catchment (Grastal) in 2022 (Figure 3-1c), where it was partly destroyed by an extreme event.

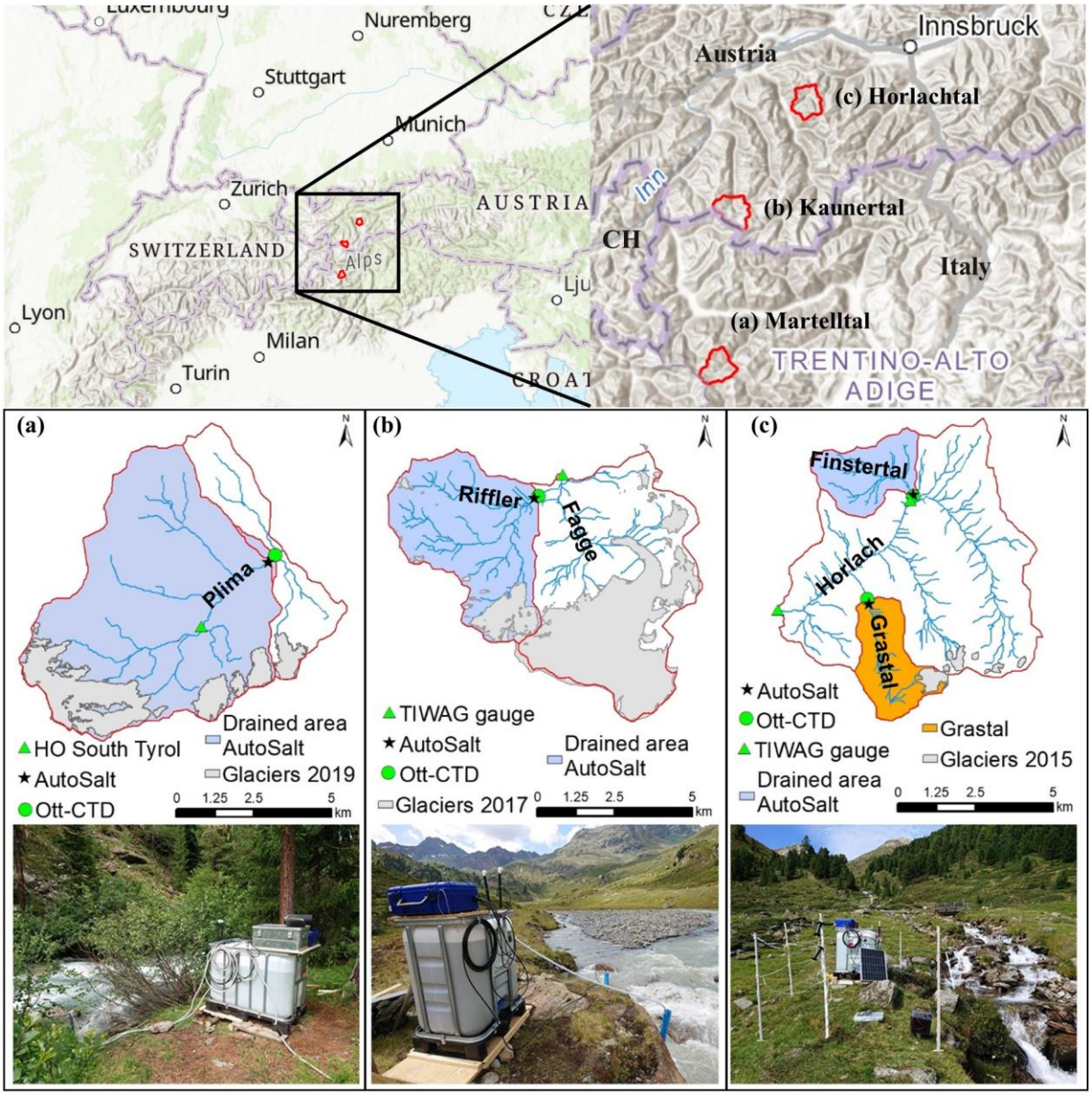


Figure 3-1: Overview maps of the three research areas, i.e., (a) Martelltal, (b) Kaunertal and (c) Horlachtal, with outer borders and delineation of the sub-catchments (red), as well as the drained area of the AutoSalt sites (blue). Green circles indicate locations of the Ott CTD probes and the green triangles already existing stream gauges. Lower panel contains corresponding pictures of the AutoSalt system. The orange area in (c) Horlachtal delineates the Grastal catchment where the AutoSalt was moved to from Finstertal (blue area) in 2022.

The main differences between the three monitored catchments are the size of the drained area, the extent of the glaciated area (Buckel & Otto, 2018; Knoflach et al., 2021), the streambed slope and the background EC_T (Table 3-1). While the drained area in Martelltal just before the inflow of the Plima into the Zufritt/Gioveretto reservoir is about 50 km², the site in Kaunertal at the Rifflerbach drains 20 km² and in Horlachtal it is a very small side valley (Finstertal) that drains about six km². Since the mean discharge depends directly on the drained area, higher amounts of salt are needed for one discharge measurement at Rifflerbach and Plima than at Finstertalbach to achieve a noticeable increase in stream EC_T. The streambed gradient of the streambed along the mixing length ranges from 3.9 % (Plima) to 4.5

% (Rifflerbach) and finally to 18 % (Finstertalbach). Further information about the site characteristics is listed in Table 3-1. The grading of streamflow measurements, as listed in Table 3-1, follows the Resources Information Standards Committee (RISC) (2018) hydrometric standards for streamflow measurements. Grade A is for measurements with an uncertainty of smaller 7 %, B for the range 7 % to 15 % and C for uncertainty larger than 15 %.

Table 3-1: Key features of the AutoSalt sites such as drained area [km²], glaciated area [km²], elevation of AutoSalt and mean elevation of basin [m a.s.l.], tank volume, distance between AutoSalt, ECT and OTT CTD probes, mean slope of the mixing reach, mean background ECT. Also included are metadata of the measurements such as total number of measurements, highest peak event, proportion of grades (A, B, C) according to Resources Information Standards Committee (RISC) (2018) and dose of the salt tracer [g NaCl per m³/s].

Valley	**Martelltal**	**Kaunertal**	**Horlachtal**
Stream	**Plima**	**Rifflerbach**	**Finstertalbach**
Drained area [km²]	50.4	19.9	6.2
Glaciated area [km²]	8.15 (2019)	2.45 (2017)	0.0 (2015)
Elevation of AutoSalt [m a.s.l.]	1917	2192	1982
Mean elevation of basin [m a.s.l.]	2837	2817	2512
Tank volume [l]	600	300	300
Distance between AutoSalt injection and EC_T sensors [m]	200 (right); 210 (left)	79 (right); 82 (left)	57 (right); 67 (left)
Distance between AutoSalt injection and OTT CTD probes [m]	181	51	41
Mean slope of the mixing reach [%]	3.9	4.5	18
Observation period AutoSalt	07.2020-10.2021	07.2021-10.2021	08.2020-10.2021
Mean background EC_T [µS/cm]	205	174	31
# of Q measurements	78	40	180
Highest measured Q [m³/s]	17.1	5.9	2
Lowest measured Q [m³/s]	1.1	0.5	0.1
Share grade A [δQ <7 %]	74.4 %	95 %	73.3 %
Share grade B [>7 % δQ <15 %]	25.6 %	5 %	26.6 %
Share grade C [δQ >15 %]	0 %	0 %	0 %
Dose [g NaCl per m³/s]	400	200	400

3.2.2. Salt dilution method

The salt dilution method is based on the point injection of a solution of NaCl and the measurement of the breakthrough curve of the electrical conductivity downstream (Moore,

2005). The method relies on two key assumptions: i) mass conservation of the salt tracer and ii) complete mixing of the salt tracer across the stream width at the location of the EC sensors. The stream discharge Q (m^3/s) can be computed with the following equation:

$$Q = \frac{M}{CF_T * A_{BC}} \quad 3\text{-}1$$

where M is the mass of salt injected (kg), CF_T [kg cm m^{-3}*μS^{-1}] is a conversion factor, which is the slope of the relation between salt mass concentration in the calibration solution and EC_T (Richardson, Sentlinger, et al., 2017), used for calculating the Q from the EC_T, and A_{BC} [s.µS/cm] is the area under the breakthrough curve commonly calculated as:

$$A_{BC} = \Delta t \sum_{i=1}^{n} [EC(t) - EC_{BG}] \quad 3\text{-}2$$

where Δt is the recording interval (s), EC(t) is EC_T as a function of time recorded downstream of the point of salt injection (µS/cm), EC_{BG} is the background EC_T of the stream water, and the summation is carried out over the duration of salt breakthrough curve. Additional information about the salt dilution method are available in many references such as Leibundgut, Maloszewski, & Külls (2011) or Moore (2004).

3.2.3. AutoSalt system

The automated salt injection system (AutoSalt) is an autonomous flow measurement system providing discharge data in turbulent watercourses with high temporal resolution. The system usually consists of a control module, a brine tank and stand, a creek pressure transducer (sensor accuracy ±0.1 % FS), two high-resolution electrical conductivity sensors (T-HRECS) one for each riverbank, and a salt injection system. EC and temperature (T) sensors downstream of the injection location log the EC, T, and the temperature compensated EC_T to 25°C. The EC_T records of the T-HRECS sensors are stored on individual SD cards. The measurements are also transmitted wirelessly using LoRa radios to the AutoSalt control module and processed to calculate the streamflow Q and its uncertainty. The uncertainty analysis is based on the general framework of the Guide to the expression of Uncertainty in Measurement (GUM) (JCGM, 2008) and is described in the supporting information. The detailed EC_T records (set to five second interval) is also stored in the SD Card on the AutoSalt control module for more detailed quality assurance and quality control (QA/QC) analysis. The brine is delivered from the tank to the creek by a pump through rigid piping with a mechanical flow meter providing feedback on the rate and total amount of brine delivered. The AutoSalt is programmed to trigger salt injections on the falling limb of the hydrograph, as the stage is generally more stable on the falling limb than on the rising limb. The AutoSalt internally calculates Q in real-time on the AutoSalt controller. This automated Q value is a good estimate of the measured discharge even though it does not have any QA/QC correcting, for example,

for changing background EC_T or missing data. Therefore, external post-processing of the discharge data and its uncertainty quantification is required (Sentlinger et al., 2019). Further information about the AutoSalt system can be found in the latest user manual (Fathom Scientific Ltd., 2020a).

The high sensitivity of the T-HRECS sensors (i.e., EC in 0.001 µS/cm and temperature in 0.001°C), the high temporal resolution (5 sec), and accuracy (i.e., 0.01 % of reading for both EC and temperature) allow to achieve a high signal-to-noise ratio (SNR), thereby requiring less salt solution than conventional sensors (Sentlinger et al., 2019). The uncertainty analysis of the AutoSalt system is based on standard equations for error propagation (Sentlinger et al., 2019). The internal grading system of the AutoSalt follows the Resources Information Standards Committee (RISC) (2018) hydrometric standards for discharge measurements, which was already introduced. The computed uncertainty corresponds to the 95 % confidence interval. Additionally, basic information about the uncertainty quantification of the AutoSalt and equations is provided in the supporting information.

Since we collected various observational data with different sensors at different temporal resolutions and experimental sites, it is beneficial to give an overview of the collected data before presenting the results (Figure 3-2). Moreover, the observation period of the various sensors differs. The AutoSalt was only in operation over the summer, whereas the Ott CTD recorded throughout the year. We set up a stage-discharge relationship with stage data from the AutoSalt and Ott CTD. Thus, there are two stage-discharge relationships per experimental site. We estimated the uncertainty in the Ott and AutoSalt water level measurements by considering the sensor accuracy and the mean hourly standard deviation of the water level measurement during steady flow periods. The equation for calculating the relative accuracy of the stage measurements is in the supporting information.

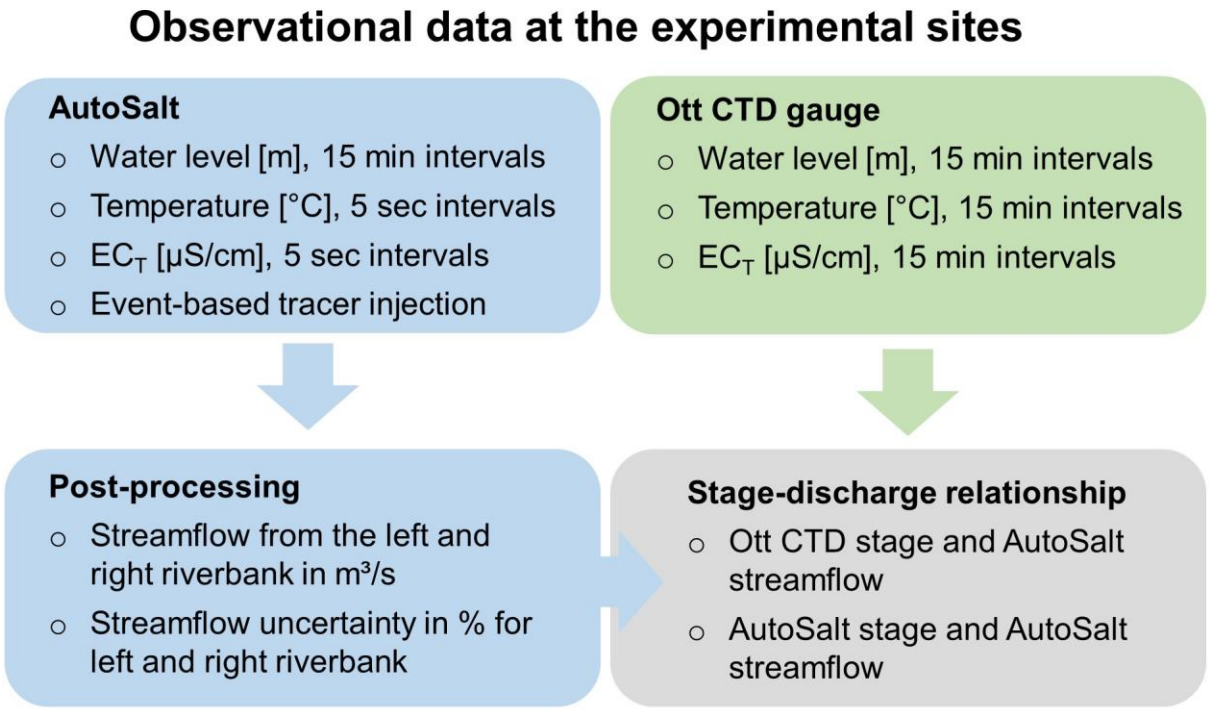


Figure 3-2: Overview of collected observational data, its post-processing, and analysis for each experimental site. The blue box represents the measurements collected with the AutoSalt, and the green box with the Ott CTD gauge.

Table 3-2 lists the quality check measures we used to evaluate the reliability and quality of the AutoSalt's measurements. With the EC_T measurements from left and right riverbank, we checked whether the salt tracer is completely mixed, which is one main prerequisite for a reliable salt tracer measurement. The equation to calculate the uncertainty due to incomplete mixing is given in the supporting information. Additionally, the comparison also allows for checking the influence of air bubbles (i.e., aeration) at the EC_T measurement sites. With an additional high-pass filter, it is possible to increase the SNR of the EC_T breakthrough curves. The redundant stage measurements allow us to quality check the water level measurements and to verify the stage-discharge relationships of each experimental site for possible inconsistencies, such as a change in the cross-section. In addition, we performed a plausibility check with the measurements from the AutoSalt by comparing the discharge measurements with data from official gauging stations and performing manual measurements during each site visit. Based on the continuous Ott CTD stage measurements, we assess the recorded peak events with respect to their probability of occurrence.

Table 3-2: Performed quality checks to test the reliability of the AutoSalt and quality of the collected automatic discharge measurement.

Quality checks	Implemented quality check measure
Test for incomplete mixing of salt tracer	Comparison of the measured values by two EC sensors on the left and right river bank
Test for change in the cross-section	Rank correlation of the measured river stage values at two different cross-sections within the mixing length
Test for noise in stage data caused by wave action	Rank correlation of the measured river stage values at two different cross-sections within the mixing length
Test for noise in EC data caused by aeration	Comparison of the measured values by two EC sensors on the left and right river bank
	Use of a high-pass filter on the breakthrough curve
Plausibility check of discharge values	Rank correlation with official gauging stations
	Comparison with manual measurements
Assessment of peak events	Comparison between continuous Ott CTD river stage measurements and event-based AutoSalt recordings

3.3. Results

3.3.1. Recorded events by the AutoSalt

The system performed 298 measurements in total, with discharges ranging from 0.1 m^3/s to 17.1 m^3/s (see also Table 3-1 for more details). The largest amount of discharge measurements (180) were generated at Finstertalbach since it has the smallest drainage area with a mean discharge of 0.7 m^3/s and hence requires the smallest volume of brine injection. The AutoSalt system at Rifflerbach was only in operation in 2021 and collected 40 discharge measurements with a mean discharge of 2.6 m^3/s. The Martelltal AutoSalt collected 78 discharge measurements with a mean discharge of 4.8 m^3/s. The average relative uncertainty of Q, estimated by the aforementioned uncertainty quantification framework, is 6.2 % at Plima, 4.1 % at Rifflerbach, and 5.4 % at Finstertalbach.

The default dose of the AutoSalt is 200 g per m^3/s. However, for a higher background EC_T or longer transient times, larger doses are recommended. For short transient times and lower background EC_T, the dose can be reduced. At Rifflerbach, we used the default dose of 200 g per m^3/s and increased the dose to 400 g per m^3/s at Finstertal and Plima. The increase of the dose at Plima was necessary due to the relatively long distance between injection and EC_T recording of about 200 m and the high mean background EC_T (205 µS/cm).

Despite the higher dose at Plima, the EC_T peaks were only about 10 µS/cm above the background EC_T as can be seen in Figure 3-3a. It was possible to derive discharge

measurements with lower uncertainty during the peak flow event that occurred end of August 2020 due to the clear EC_T signal (i.e., high SNR). However, the EC_T signal shows several small spikes during the rising limb of the breakthrough curve and a slight increase in the background EC_T (Figure 3-3b). The AutoSalt system performed six discharge measurements during this particular peak event (Figure 3-3a).

At Rifflerbach, we were able to collect 11 measurements during a peak event, which occurred at the end of July 2021 (Figure 3-3c). Despite the lower dose of 200 g per m³/s and a partly noisy background EC_T, the EC_T peaks are well recognizable. The mean EC_T at Rifflerbach was 174 µS/cm during the measurement period 2021. The snapshot shows a clear peak of about 30 µS/cm above the background EC_T with hardly any spikes in the signal (Figure 3-3d).

We can observe the highest relative change in EC_T (> 70 %) for injections at the Finstertal (Figure 3-3e) since the background EC_T is relatively low with 31 µS/cm (Figure 3-3f). Therefore, the EC_T peaks of the injections are clearly visible. However, the background EC_T signal starts to oscillate when the water level and discharge increase (Figure 3-3e). The EC_T peaks are still clearly visible, but the stronger oscillations make the processing of the measurements more difficult and lead to higher uncertainty. Therefore, we have decided to set the dose to 400 g per m³/s at Finstertal.

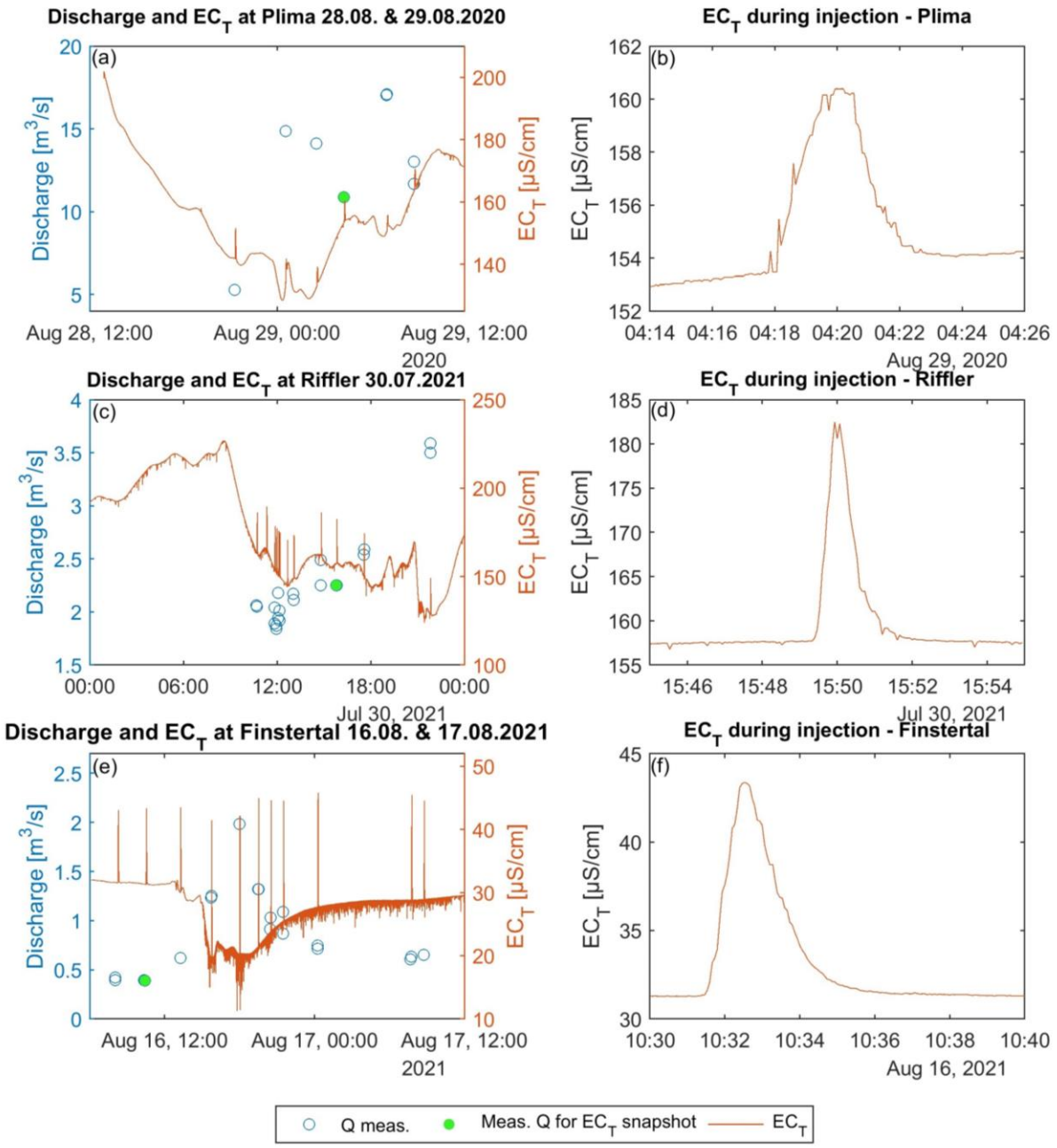


Figure 3-3: Snapshot of measured discharge (blue dots) and EC_T (orange line) at each site (a) Plima, (c) Rifflerbach, and (e) Finstertalbach, as well as a detailed snapshot of an exemplary EC_T breakthrough curve (green dots) of the corresponding sites.

At each experimental site, the AutoSalt pressure sensor at the site of the injection and the downstream Ott CTD sensor both record the water level at the same 15-minute time interval. Depending on the length of the mixing section and flow velocity, there may be a deviation between injection and recording at the downstream sensors of up to five minutes, as in the example of the Plima. Each streamflow measurement of the AutoSalt can be linked with two stage observations, one performed with the Ott CTD sensor, and the other with the AutoSalt sensor. We compared the relationship between the two stage measurements in the supporting information (Figure A.2.6), and we can observe that they are highly correlated (Kendall tau of 0.72 at Plima, 0.74 at Rifflerbach, and 0.81 at Finstertalbach). The Kendall rank coefficient is

a non-parametric test and, hence, does not rely on any assumptions concerning the distributions (Kendall, 1938).

The left panel of Figure 3-4 shows the stage-discharge relationship at all three sites, including the measurement uncertainties of discharge and water level at the downstream Ott CTD river section and the right panel at the AutoSalt section. In addition to the different channel geometry, the stage-discharge relationship also differs because of the aforementioned timing of the measurements and the prevalence of turbulence or wave action at the measurement site, which can lead to noise in the stage signal. The Ott CTD pressure transducer is only located at the deepest point of the cross-section at Finstertal. At the other sites, the sensor is not recording the lowest stages in winter. The AutoSalts were not in operation when performing some manual streamflow measurements. This is why a few manual streamflow measurements are only linked with the Ott CTD stage (left panel of Figure 3-4). Manual and automated measurements do not display particular anomalies and agree with each other. The uncertainty of water level and discharge increases during larger discharge events, especially at the Rifflerbach (Figure 3-4c). In addition, a jump in the stage-discharge relationship at 0.3 m water level can be seen in the case of Rifflerbach due to cross-section widening where the Ott CTD probe is located (Figure 3-4c). A change in the geometry of the control-section where the Ott CTD probe is located between 2020 and 2021 is already visible at mean discharges at Plima (Figure 3-4a). The cross-section at the AutoSalt is more stable and does not show a change (Figure 3-4b).

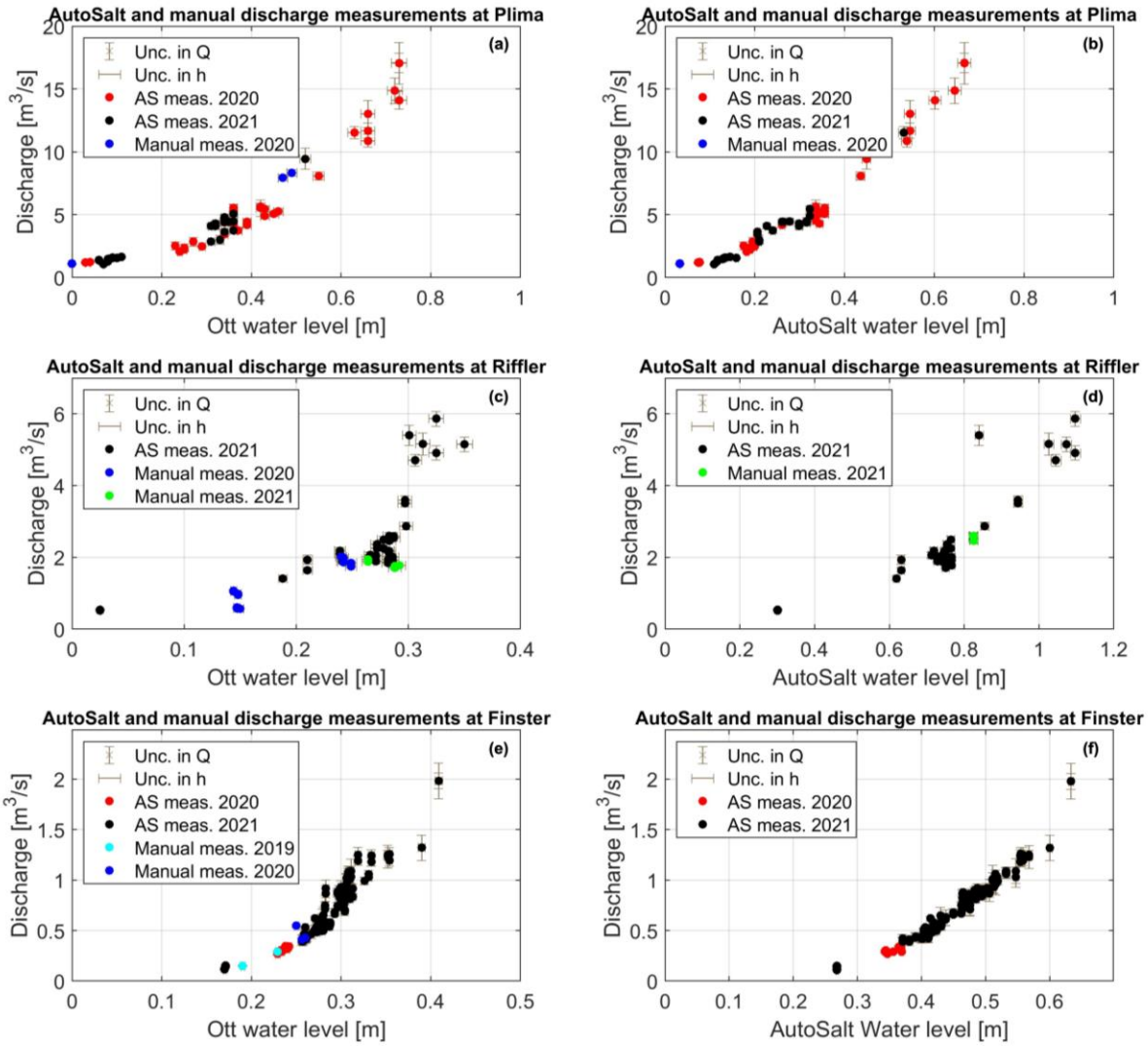


Figure 3-4: The left (right) panels show the stage-discharge relationship of the salt dilution measurements with stage recorded by Ott CTD (AutoSalt stage), including the measurement uncertainties at the three sites (a-b) Plima, (c-d) Rifflerbach and (e-f) Finstertalbach. AS - AutoSalt measurements.

In order to assess at which discharge condition the AutoSalt recorded the most streamflow measurements we analysed the percentage of measurements during frequent water levels and discharges, which corresponds to the 25 to 75 percentile of the continuous stage observation recorded by Ott CTDs (Figure 3-5). At Plima, most measurements (55 %) were collected during frequent water levels and discharges. Most measurements were collected during higher water levels at Rifflerbach (43 %) and Finstertalbach (28 %). Consequently, lower stages (<0.25 m) are underrepresented at Rifflerbach and Finstertalbach. The pronounced seasonality of runoff in high-elevation catchments is the reason for the underrepresentation of low flows, as the AutoSalt is not in operation during the winter period. The percentile of the recorded peak event is 0.99 at Plima, 0.98 at Rifflerbach, and 0.99 at Finstertalbach. Accordingly, the recorded streamflow peaks are slightly lower than the recorded maximum stage peaks in the respective observation period.

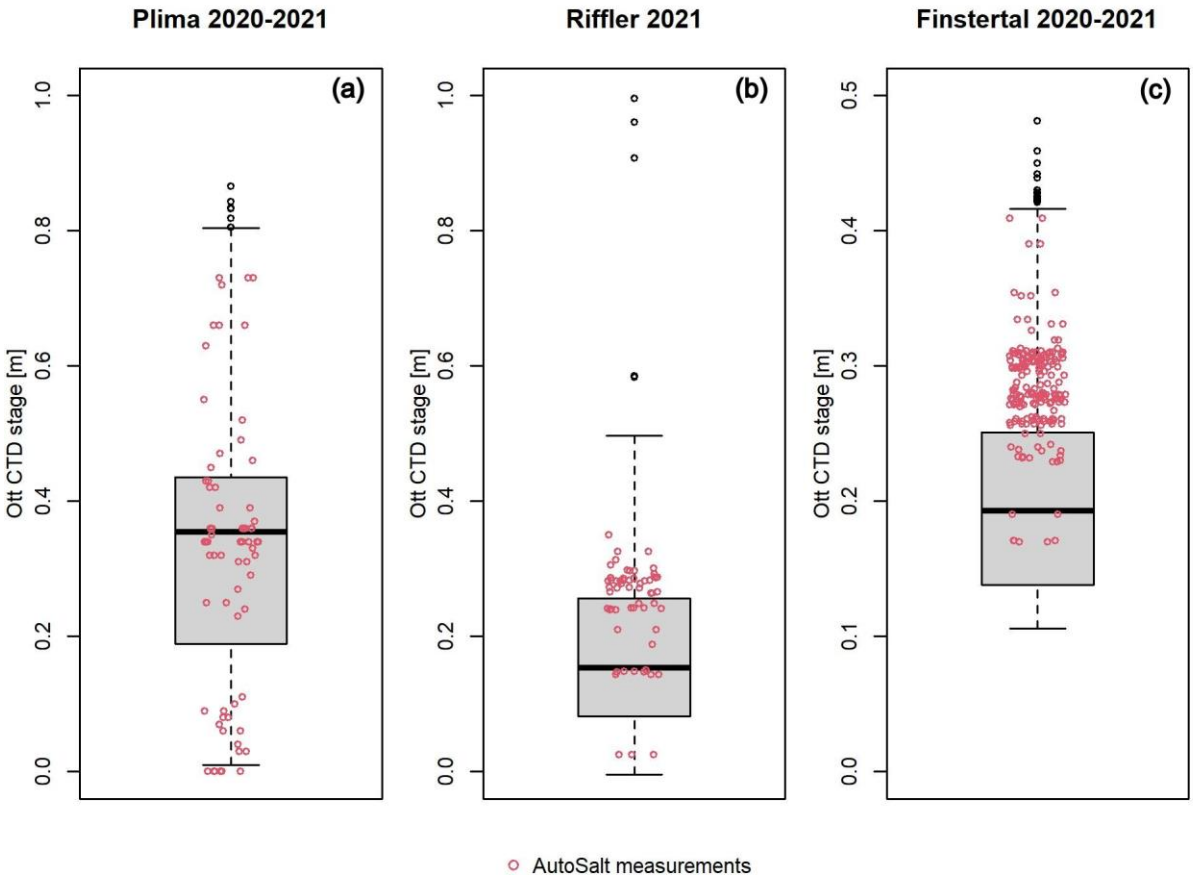


Figure 3-5: The Boxplot shows the Ott CTD stage during the observation period and the stages at which AutoSalt measurements (red dots) were taken at each site (a) Plima, (b) Rifflerbach and (c) Finstertalbach.

3.3.2. Quality of the AutoSalt measurements

Since the measurement uncertainty is available for each riverbank in addition to streamflow, we examined whether there is a systematic correlation between streamflow magnitude and uncertainty or a difference between the left and right bank. The left panels of Figure 3-6 illustrate the dependence of discharge uncertainty on discharge magnitude for both stream banks separately (i.e., the right or left bank from orographic view). Generally, the uncertainty is mostly below five %. Larger measurement uncertainty (> 10 %) occur sporadically at certain discharges and not always on both river banks. There is no systematic correlation between discharge magnitude and uncertainty (i.e., Kendall tau of 0.36 at Plima, 0.24 at Riffler, and 0.07 at Finstertal). Rather, the SNR at different water levels and the presence of turbulences is decisive for a low measurement uncertainty. The right panels of Figure 3-6 show the observed discharge by the already existing stream gauges and the discharge recorded by the AutoSalt. The discharges of the already existing gauges are not based on simultaneous measurements but are extracted from the discharge time series available at 15-minute resolution or 10-minute resolution in the case of Plima. Since we do not have the rating curves of the existing gauges, we cannot make any statement about the uncertainty regarding the discharge time series. The highest agreement (Kendall tau of 0.86) was found for the official

discharge observations at Horlach Fassung and AutoSalt at the Finstertal site (Figure 3-6f), which is located only about 65 meters upstream of the Horlach Fassung gauge. For the Plima, the official stream gauge is about 3.8 km upstream of the AutoSalt injection site. There is predominantly only good agreement at lower and medium discharges (< 7 m³/s) (Figure 3-6b). At higher discharges, the scatter increases. Nevertheless, a Kendall tau correlation coefficient of 0.8 is achieved. At the Riffler stream, the distance between the two discharge observation sites is smaller (about 1.3 km) than at the Plima. However, the discharge magnitudes recorded at the Gepatschalm gauge and the AutoSalt injection site differ significantly (Figure 3-6d), which results in a Kendall tau of 0.76. The reason for the discrepancy is the larger share of the glaciated area at the Gepatschalm gauge compared to the AutoSalt site.

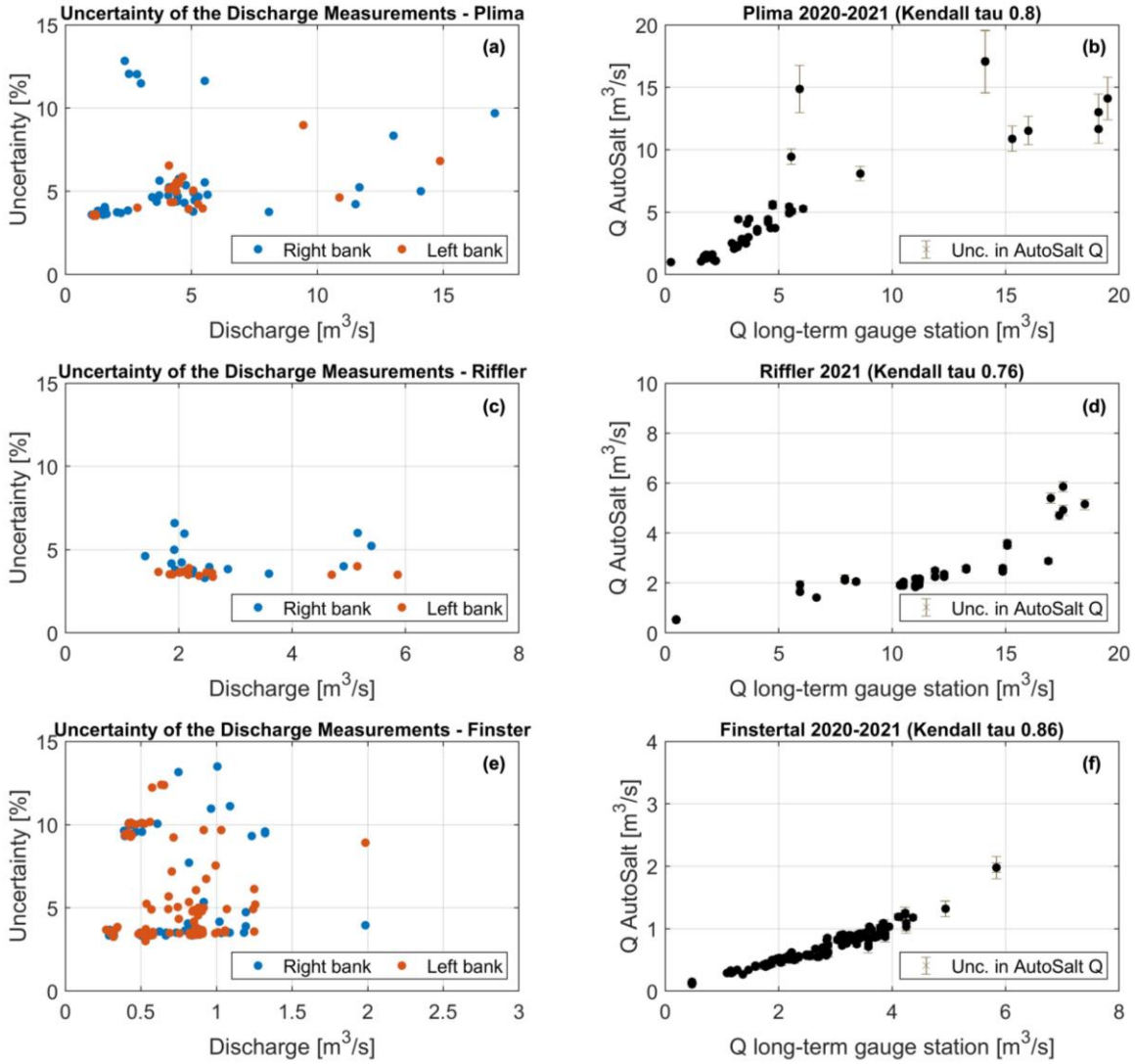


Figure 3-6: The left panels show the uncertainty of the discharge measurements of the AutoSalt observed at the right (blue dots) or left (red dots) bank at the respective experimental sites (a) Plima, (c) Rifflerbach, and (e) Finstertalbach. The right panels show the observed discharge by the already existing stream gauges vs. the discharge recorded by the AutoSalt at the respective sites and when the AutoSalt performed a measurement (b) Plima, (d) Rifflerbach, and (f) Finstertalbach.

We moved the Finstertal AutoSalt to another sub-catchment (Grastal) of Horlachtal in order to collect discharge measurements to efficiently set up a stage-discharge relationship at Grastalbach in 2022. The Grastal is characterized by steep topography and a high activity of debris flows (Rom, Haas, Heckmann, et al., 2023). The AutoSalt started to perform discharge measurements in mid of July 2022 (Figure 3-7). About one week after installation, the system was partly destroyed by an extreme event. A storm event has triggered not only a fast flood wave but also activated debris flow, which transported debris into the creek (Rom, Haas, Hofmeister, et al., 2023). The highest observed peak in the stage of the downstream Ott CTD probe was reached in less than two hours (i.e., time to peak). The AutoSalt made an injection at the highest peak, but at this point, the injection hose and the pressure transducer of the AutoSalt were torn away by the debris flow. Therefore, we cannot compute the discharge of the peak flow. Nonetheless, thanks to the continuous monitoring of the device, we can estimate the time of the change in the cross-section and assume that the change occurred during the peak flow of the first event of July 19.

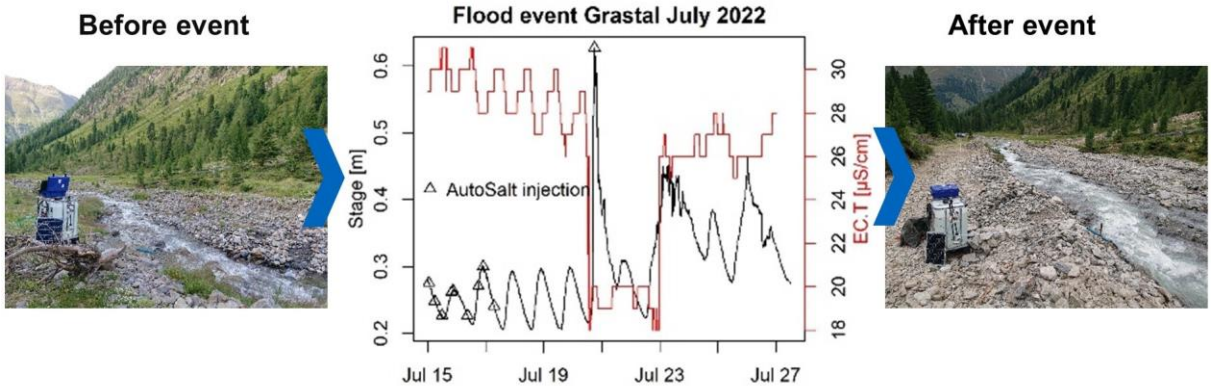


Figure 3-7: AutoSalt monitoring of a combined extreme event of flood and debris flow in a high-elevation catchment (Grastal) in July 2022. The graph shows the recorded downstream stage and ECT of an Ott CTD probe as well as the salt injections performed by the AutoSalt. The change in the river channel as a result of this event can be seen by comparing the two photos.

3.4. Discussion

3.4.1. Recording of peak flow

The AutoSalt system recorded close to the peak discharges at all experimental sites in the observational period. However, it is unclear how reliably the AutoSalt works at discharges > 20 m³/s. The extreme case at the Grastalbach demonstrates the risk associated with these events, which would make the manual acquisition of any data impossible, while the AutoSalt system was able to retrieve some important information, like the rate of increase in the water the level. The collected high discharge events are especially important for the construction of robust rating curves. The highest discharge measurements were made slightly below peak water levels since the AutoSalt performs the measurements on the falling limb of the

hydrograph. We observed some variability in the highest recorded discharge values at similar water levels, as seen, for example, at Rifflerbach (Figure 3-4c and d. In addition, the pronounced noise in the stage measurements at Finstertalbach (Figure A.2.7) also leads to scatter in the stage-discharge relationship (Figure 3-4e and f). The reason for the noisy stage signal at Finstertalbach is the relatively high streambed gradient (i.e., 18 %) that results in a highly turbulent flow with pronounced wave action at the measurement sites. The two other experimental sites have distinctive lower streambed gradients (i.e., 3.9 % at Plima and 4.5 % at Rifflerbach) that result in less turbulent flow (Figure A.2.7).

Another important aspect of measuring peak flow events with the AutoSalt is the time interval between two injections. The default interval between two consecutive injections is 1.5 h, which ensures that the tracer measurements do not interfere with each other. Under good mixing conditions (e.g., high flow velocity and turbulence), the time interval can be reduced to allow a higher number of measurements during a peak flow event. At Rifflerbach, for instance, the breakthrough time was about three minutes, as shown in Figure 3-3d. Therefore, we reduced the interval to 0.5 h. Although the system is autonomous and can be operated remotely, site visits should be at least monthly to ensure reliable operation of the system by checking the system's functionality and calibrate the system (e.g., brine concentration, flow meter, ECT sensors) if necessary.

3.4.2. Changes in cross-section

The large number of discharge measurements recorded by the AutoSalt allows us to identify changes in the cross-section from one year to another, as is the case at Plima (Figure 3-4a). This will enable us to quality-check the rating curve and have a valid rating curve for each observational year or before and after each relevant hydrological event that can modify the river cross-section. We could only observe the change in the cross-section at the downstream river section, where the Ott CTD is located, and not at the injection site (Figure 3-4b), showing the possibility of very local effects that can lead to uncertain streamflow observations. Even after the complete change of a cross-section, as was the case at Grastalbach, the AutoSalt systems enable a quick and efficient collection of discharge measurements for setting up a new stage-discharge relationship.

Moreover, we can detect segments within the stage-discharge relationship between mean and high flow conditions, specifically for the Rifflerbach at about 0.3 m stage (Figure 3-4c). The segments can be considered in the construction of a three-segment rating curve with an optimal rating curve fit for low, mean, and high flow conditions (Horner, Le Coz, Renard, Branger, & Lagouy, 2022).

3.4.3. Contrasting experimental sites

We obtained good results with the AutoSalt at all three experimental sites, which differ strongly concerning discharge and stage range, glaciated area, channel geometry, and background EC_T. As mentioned, the injection quantity depends directly on the expected streamflow quantity, which in turn depends on the catchment size. Thus, the AutoSalt is well suited for smaller catchments (<10 km^2) since a larger number of measurements are possible with one reservoir filling (i.e., 300 L). However, it is also possible to upgrade the system (e.g., 600 L), as we did at the Plima in 2021. Accordingly, it is possible to perform about 60 injections with a dose of 400 g per m^3/s for a high-elevation catchment with 50 km^2 and a mean discharge of about five m^3/s. The range of collected discharge measurements was largest at Plima (from 1.1 to 17.1 m^3/s), followed by Riffler (from 0.5 to 5.9 m^3/s) and Finstertal (from 0.1 to two m^3/s).

It turned out that the selection of the sites for the EC_T sensors is essential for controlling the SNR and hence the measurement uncertainty, as shown in the left panels of Figure 3-3. We installed the sensors as close as possible to the riverbed (but ensured that there was no sediment accumulation inside the sensor casing) to reduce the influence of aeration and prevent the sensors from falling dry at low flows. Nevertheless, it cannot be ruled out that the flow behavior changes depending on the water level, resulting in increased turbulence and consequently introducing noise in the data due to waves and aeration, as was the case at the Finstertal site with the steepest streambed gradient (i.e., 18 %). Depending on the background EC_T and turbulences at EC_T monitoring sites, the injection dose can be increased to achieve a higher SNR. In the end, the selection of the injection dose is a compromise between the efficient use of the salt brine and a sufficiently high SNR, which allows a relatively low measurement uncertainty (<7 %). Since unstable background EC_T, as seen in Figure 3-3b, can lead to higher uncertainties in the discharge measurements, we have considered its uncertainty in the post-processing (i.e., the deviation between pre and post background EC_T). This is particularly relevant when a measurement extends over several minutes while the background EC_T changes quickly, like at Plima (Figure 3-3a).

While we performed regular manual measurements at each site visit, these were often at very similar streamflow conditions (Figure 3-4). Nevertheless, they are very helpful in verifying the automated measurements since they have less uncertainty in the injected mass. In some cases, the manual measurements can deviate from the stage-discharge relationship, as visible at Rifflerbach (Figure 3-4c). At low discharges with low turbulence, there may be deviations because of incomplete lateral mixing of the salt tracer.

It is noticeable how EC_T drops abruptly during larger precipitation events at all three measuring sites (Figure 3-3a). The natural background EC_T in streams as a predictor for discharge was already observed in previous studies (Cano-Paoli, Chiogna, & Bellin, 2019; Chang, Mewes, &

Hartmann, 2022; Weijs et al., 2013), and therefore the collected continuous EC_T values of the Autosalt system and of the Ott CTD probe could be further used beyond the event-based observations to generate EC_T based gauging station.

3.5. Conclusion

By using the AutoSalt, it was possible to perform continuous and event-based streamflow measurements in high-elevation Alpine areas at very different locations. Rare peak discharges of almost 20 m^3/s were measured with the system. The mean measurement uncertainty of the almost 300 measurements is in the majority (81 %) below 7 %. Larger measurement uncertainties resulted from low SNR due to turbulence at the EC_T measurement sites. The low SNR occurred particularly at the experimental site with the largest streambed gradient (i.e., 18 %), where turbulence and wave action affected EC_T and stage measurements independent of the discharge magnitude. In addition, the noise in the stage observations transfers to the stage-discharge relationship. Consequently, we could only find a correlation between the measurement quality and the streambed gradient as the only site-specific characteristic. Moreover, we could not find a systematic correlation between discharge magnitude and measurement uncertainty. Therefore, the sensor locations should be thoroughly selected in river sections with a high streambed gradient to establish a reliable stage-discharge relationship at creeks in high-elevation catchments. Another benefit of the collected measurements of the system was the detection of non-stationarity in the cross-section, which is a particular challenge for establishing reliable stage-discharge relationships on creeks with natural cross-section.

4. Intercomparison of Sentinel-2 and modelled snow cover maps in a high-elevation Alpine catchment

Hofmeister, Arias-Rodriguez, et al., 2022[3]

Abstract

Modelling runoff generation in high-elevation Alpine catchments requires detailed knowledge on the spatio-temporal distribution of snow storage. With Sentinel-2 MultiSpectral Instrument (MSI), it is possible to map snow cover with a high temporal and spatial resolution. In contrast to the coarse MODIS data, Sentinel-2 MSI enables the investigation of small-scale differences in snow cover duration in complex terrains due to gravitational redistribution (slope), energy balance and wind-driven redistribution (aspect). In this study, we describe the generation of high-resolution spatial and temporal snow cover data sets from Sentinel-2 images for a high-elevation Alpine catchment and discuss how the data contribute to our understanding of the spatio-temporal snow cover distribution. The quality of snow and cloud detection is evaluated against in-situ snow observations and against other snow and cloud products. The main problem was in the false detection of snow in the presence of clouds and in topographically shaded areas. We then seek to explore the potential of the generated high-resolution snow cover maps in calibrating the gravitational snow redistribution module of a physically-based snow model, especially for an area with a very data-scarce point snow observation network. Generally, the calibrated snow model is able to simulate both the mean snow cover duration with a high F1 accuracy score of > 0.9 and the fractional snow-covered area with a correlation coefficient of 0.98. The snow model is also able to reproduce spatio-temporal variability in snow cover duration due to surface energy balance dynamics, wind and gravitational redistribution.

[3]Material from:
Hofmeister, F**.,** Arias-Rodriguez, L. F., Premier, V., Marin, C., Notarnicola, C., Disse, M., & Chiogna, G. (2022). Intercomparison of Sentinel-2 and modelled snow cover maps in a high-elevation Alpine catchment. *Journal of Hydrology X*, *15*, 100123. doi: 10.1016/j.hydroa.2022.100123.

4.1. Introduction

Alpine snow cover and its subsequent melt can dominate local to regional climate and hydrology in both high-latitude areas (Gascoin et al., 2019) and the world's mountainous regions (Dozier & Painter, 2004). In the European Alps, snow is the major driver of Alpine hydrology, storing water during the winter season and releasing it in the spring and summer, with impacts on water supply, agriculture and hydropower production (Matiu et al., 2021). Moreover, ongoing climate change in the Alps affects the abundance of snow (Matiu et al., 2021). Finally, snow accumulation and melting processes also depend on topographical parameters such as slope and aspect, which are linked to incoming solar radiation, gravitational transport and wind redistribution (López-Moreno & Stähli, 2008; Grünewald & Lehning, 2011; López-Moreno et al., 2014; Grünewald, Bühler, & Lehning, 2014; Pedersen et al., 2016; Gurung et al., 2017; Mott, Vionnet, & Grünewald, 2018; Saydi & Ding, 2020; Vionnet et al., 2021). Therefore, to calibrate and validate models capable of properly capturing snow dynamics in such a complex environment requires accurate observed snow cover maps.

The standard source of information regarding snow cover is a network of automatic and manual ground-based meteorological stations that perform daily or sub-daily observations, mostly of snow depth (Romanov, Gutman, & Csiszar, 2000). However, in-situ snow observations are generally insufficient for characterizing the high spatial variability of the snow pack in mountainous regions (Gascoin et al., 2019) and are often only suitable for making qualitative comparisons with hydrological model results (Tuo, Marcolini, Disse, & Chiogna, 2018). Moreover, field measurements are sparse in both space and time and they are subject to several sources of error (Rasmussen et al., 2012; Marcolini et al., 2019a). Furthermore, only few snow observation stations operate at altitudes above 2000 m, which complicates the investigation of altitudinal gradients (Matiu et al., 2019). Despite these limitations, snow depth ground observations are very valuable when it comes to evaluate snow detection quality using remote sensing products (e.g. Foppa et al., 2005; Gascoin et al., 2020; Barrou Dumont et al., 2021).

Since the launch of the first Landsat satellite in the 1970s, remote sensing has become a key tool for mapping snow cover and revealing snow properties at multiple spatial and temporal resolutions (Romanov et al., 2000; Dozier & Painter, 2004). Snow can be characterized by many variables, such as the snow-covered area (SCA), fractional area (fSCA), albedo, liquid water content, snow depth or snow water equivalent (Frei et al., 2012). In remote sensing applications, snow cover is mainly investigated using SCA and fSCA products on the scale of entire mountain ranges like the Alps or Pyrenees (Gascoin et al., 2019). Remote sensing products supply information about snow-covered areas across elevations, but they have problems with correct snow detection in cloudy conditions. Consequently, when it comes to

improving a snow detection algorithm, the main challenge is to reduce misclassification of cloud as snow (Gascoin et al., 2019). Another difficulty is in the detection of snow in dense forest areas, where the ground is obstructed by the canopy. This is particularly the case with evergreen conifer forests in Alpine regions (Di Marco et al. 2020). Besides the influence of forests, topographical features such as steep, shady slopes can impact the detection of snow by optical remote sensing products (Gascoin et al., 2019).

Moreover, there is a trade-off between spatial resolution and swath width (i.e. the acquisition strip of a satellite), which influences the observation frequency (Dozier & Painter, 2004). Despite these limitations, spatial snow information derived from satellite data has been used for decades for various purposes, including hydrology and water resource management (Rango, Wergin, & Erbe, 1996; Hall, Foster, DiGirolamo, & Riggs, 2012; Brown et al., 2014; Kääb et al., 2016; Fassnacht et al., 2017). The application of MODIS data to derive snow cover information on a daily or twice-daily basis, depending on the geographical location, is very widespread (Matiu et al., 2019), but their spatial resolution of 500 m is too coarse for hydrological applications in mountain regions, where snow cover properties can vary on scales of 10 m to 100 m (Blöschl, 1999; Gascoin et al., 2019; Vionnet et al., 2021). In fact, although Gurung et al. (2017) used snow cover maps made with MODIS to investigate the effect of topography on the SCA in very large basins (> 30,000 km²), Bouamri et al. (2021) demonstrated that MODIS is not capable of capturing the spatial heterogeneity of snow cover induced by solar radiation, because it does not capture spatial variability below 500 m. This limits the usage of MODIS snow cover maps for calibrating and evaluating spatially distributed snow models (Bouamri et al., 2021).

High-resolution (30 m) snow cover maps can be generated from Landsat images, but the low temporal revisit time of the Landsat mission (16 days) is a significant limitation to snow cover monitoring, and, moreover, data availability can be considerably reduced by cloudiness. The launch of the second Copernicus Sentinel-2 satellite in 2017 has made it possible to map the extent of snow cover at a 20 m resolution, with a revisit time of 5 days (Gascoin et al., 2019). The high-resolution spatial and temporal data collected by Sentinel-2 MultiSpectral Instrument (MSI) are important for enabling detailed investigations of snow cover and snow redistribution processes on the catchment scale (Foppa et al., 2005).

Sentinel-2 data can therefore not only be used to evaluate the parametrizations of snow melt and redistribution in physically-based models, but also to actually act or contribute to the objective function used for model calibration (Mott, Schirmer, Bavay, Grünewald, & Lehning, 2010). In general, the challenge with physically-based snow models - even without considering snow redistribution processes - is that with increasing catchment size and simultaneous high grid discretization, computation times increase significantly when, for example, multi-layer

snow models are applied for long periods, i.e. climate change studies. The computational effort increases even more when mechanistic snow redistribution approaches in physically-based models are applied at the catchment scale (Thornton et al., 2021). Hence, snow redistribution approaches are mainly based on empirical assumptions in hydrological models and have different levels of complexity. A research gap thus exists between the development of process-based snow redistribution models, i.e., based on a mechanistic snow redistribution routine (Freudiger et al., 2017), and their experimental validation (Warscher et al., 2013; Schöber et al., 2014; Frey & Holzmann, 2015). Recently, Vionnet et al. (2021) used the Canadian Hydrological Model (CHM) to simulate small-scale variabilities in snow accumulation caused by gravitational redistributions (avalanches) and blowing-snow transport (saltation and suspension), and among others factors. By driving the CHM with spatial high-resolution (50 m) wind fields, it was possible to account for the influence of topographical features due to wind speed and direction. The simulated snow accumulation was evaluated using high-resolution airborne light detection and ranging (lidar) snow depth data and snow persistence indexes derived from remotely sensed imagery. Although the CHM was able to simulate the small-scale variability of snow accumulation, there is still the need for optimization snowdrift-permitting models for large scale application (≥ 1000 km²), in particular the representation of subgrid topographic effects on snow transport (Vionnet et al., 2021). Besides the limitation of physically-based and fully distributed snow models on the catchment scale due to their heavy computational requirements (Thornton et al., 2021), detailed model input data (e.g. local wind fields) are needed when wind-driven snow redistribution is simulated with models based on semi-empirical parameterizations of the physics of snow transport (Essery, Li, & Pomeroy, 1999; Durand, Guyomarc'h, Mérindol, & Corripio, 2005; Liston et al., 2007; Pomeroy et al., 2007) or even with models resolving the 3D turbulent-diffusion equation (Gauer, 1998; Lehning, Löwe, Ryser, & Raderschall, 2008; Sauter et al., 2013; Schneiderbauer & Prokop, 2011; Vionnet et al., 2014) for blown snow particles in the atmosphere (Mott et al., 2018; Vionnet et al., 2021). Although SCA from satellite products have been used in various studies to calibrate spatially distributed snow models to improve model-internal consistency (Duethmann, Peters, Blume, Vorogushyn, & Güntner, 2014; Freudiger et al., 2017; Thornton et al., 2021), to the best of our knowledge, satellite data have rarely been used to calibrate snow redistribution routines, which are often only calibrated against observed discharge (Warscher et al., 2013; Frey & Holzmann, 2015). In a recent publication, Thornton et al. (2021) also considered parameter optimization of gravitational redistribution in a novel calibration approach for an energy balance-based snow model including snow cover maps derived from Landsat-8.

In this work, we aim to develop a high-resolution, spatial and temporal method of deriving snow cover and, further, to calibrate and evaluate the WaSiM (Water Flow and Balance Simulation

Model) snow module (Schulla, 2021). Hence, the objective is to perform a process-oriented analysis of gravitational snow redistribution and energy balance to accurately estimate the snow cover extent. Therefore, we generated two alternative high-resolution snow cover products from Sentinel-2 images for a high-elevation Alpine catchment with a very sparse snow observation network. The use of two products, allows us to acknowledge the uncertainty affecting also satellite products and to consider it in the calibration of the model. To accommodate cloud coverage, each of the two snow cover products used a different snow and cloud detection algorithm and we quantified the difference between the resulting snow cover properties. Cloud detection enables the inclusion of partially clouded Sentinel-2 images in the snow cover dataset by setting clouded areas to no data. Even with cloud cover, these images may contain valuable information beyond the cloud extent regarding the snow cover evolution during the melting season and during the onset of snow accumulation. Finally, snow cover maps with high spatial and temporal resolutions allow an accurate analysis of the effect of topographical features (elevation, slope and aspect) on the snow cover duration and provide useful information for assessing simulated snow cover maps from a physically-based hydrological model for a period of five years.

The main novelties of this paper consist in the application of two different, i.e. one unsupervised and one supervised, snow and cloud detection algorithms for Sentinel-2 images, their use for hydrological model calibration in a topographically complex region, including an assessment of the differences between the two products and an investigation of the effects of topography (elevation, slope and aspect) on mean snow cover duration.

4.2. Data and methods

4.2.1. Research area and in-situ snow observation

The upper Martell valley (Martelltal, Val Martello) is an Alpine valley with a SW-NE orientation located in the upper Adige catchment in South Tyrol (Italy), covering an area of 65 km². To evaluate the satellite products, the research area was slightly expanded to the west to include the snow observations collected at the Madritsch station located in the neighbouring Sulden valley. The upper Martell valley has an elevation range of 1840 m a.s.l. to 3760 m a.s.l. with a mean altitude of 2814 m a.s.l. (Sonny, 2017), and the dominant land cover classes are bare rock (41 %), sparse vegetation (34 %) and glaciers (19 %) (Environment European Agency CLC Corine 2018). Coniferous forest covers only 6 % of the catchment at elevations lower than 2370 m a.s.l.. Along the ridges, extremely steep slopes with inclinations of more than 60° (highlighted in red in the overview map Figure 4-1) can be found throughout the area. The main river of the upper Martell valley is the Plima, which flows into the Zufritt (Gioveretto) reservoir. The runoff regime is dominated by glaciers and snow melt, with high flows during

the spring and summer and low flows during winter (Puspitarini et al., 2020). There are also 107 lakes of proglacial origin in the catchment. Previous hydrological studies investigating the mass balance of the Langenferner glacier (Galos et al., 2017; Galos & Klug, 2015) and the impact of glacier shrinking on hydropower production (Puspitarini et al., 2020) also focused on this study area. In addition, the catastrophic flood event from 1987 was the subject of several vulnerability assessment studies (Totschnig & Fuchs, 2013; Papathoma-Köhle et al., 2015).

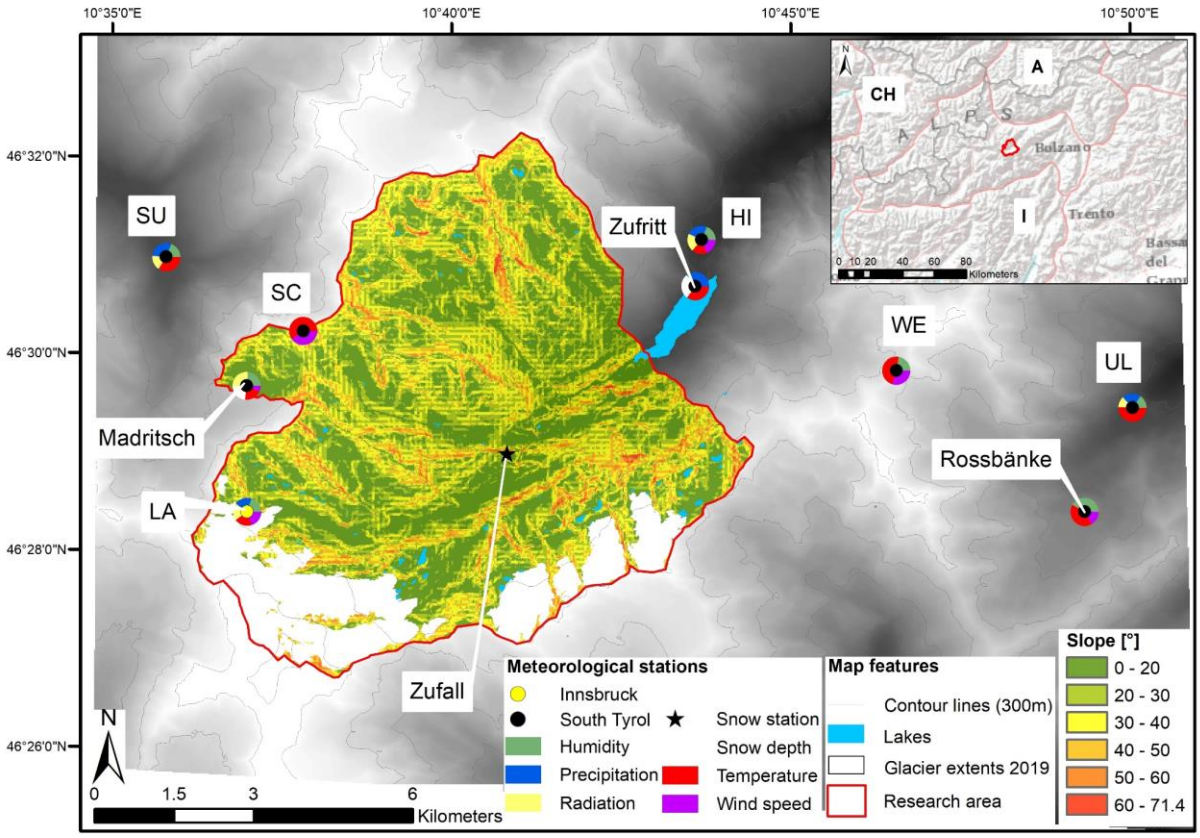


Figure 4-1: Map of the research area, the upper Martell valley in South Tyrol (Italy), showing in-situ snow observation stations (full names) and meteorological stations (abbreviations) (Source layer: Esri, Garmin, USGS, NPS, NOAA).

There is only a sparse observation network of meteorological stations covering the upper Martell valley. The station at the Langenferner (LA) glacier is the only one located inside the upper Martell valley. At the Zufall snow station, snow surveys are performed manually at weekly intervals during the winter period. Continuous snow depth observations are available from the Madritsch, Zufritt and Rossbaenke stations from various elevations and valleys. The investigation period of this study covers the period with most available station data from October 2014 to September 2020. Table 4-1 gives detailed information of the in-situ snow observations and meteorological stations and the latter are used as model forcing in the hydrological model WaSiM.

Table 4-1: Meteorological and snow observation stations. Measured variables: P = Precipitation, T = Temperature, WS = Wind speed, H = Humidity, R = Radiation, SD = Snow depth.

Station	Acronym	Elevation [m a.s.l.]	Latitude	Longitude	Measured variables	Resolution	Temporal coverage	Valley	Provider
Madrisch	-	2825	46.4938	10.6144	P, T, WS, H, R, SD	10-min	2000-2020	Sulden	1
Rossbänke	-	2255	46.469351	10.819436	T, H, WS, SD	10-min	2015-2020	Ulten	1
Zufall	-	2265	46.48129	10.67802	SD	Weekly (manual)	2004-2020	Martell	1
Zufritt	-	1851	46.509063	10.725072	P, T, SD	Daily	1980-2020	Martell	1
Hintermartell	HI	1720	46.5169	10.7269	P, T, WS, H, R	10-min	2009-2020	Martell	1
Langenferner-Felsköpfl	LA	2967	46.47245	10.61391	P, T, WS, H, R	10-min	2012-2020	Martell	2
Schöntauf-spitze	SC	3328	46.5029	10.6286	T, WS	10-min	1998-2020	Martell	1
Sulden	SU	1907	46.5159	10.5953	P, T, WS, H, R	10-min	1987-2020	Sulden	1
Ulten Weißbrunn	UL	1900	46.4868	10.8318	P, T, H, R	10-min	1987-2020	Ulten	1
Weißbrunn-spitze	WE	3252	46.494	10.774	T, H, WS	10-min	2012-2020	Ulten	1

Provider 1) Autonomous Province of Bozen/Bolzano - South Tyrol

Provider 2) Institute of Atmospheric and Cryospheric Sciences, University of Innsbruck

4.2.2. Methods

Two snow and cloud detection approaches, named TUM (unsupervised algorithm) and Eurac (supervised algorithm), respectively, are applied to identify the snow covered area in the catchment over the study period. These data sets are compared to the hydrological model WaSiM results with different parametrizations of snow redistribution (i.e. default and optimized parameters). Figure 4-2 gives a comprehensive overview of the individual processing steps used in the observed and simulated snow cover maps.

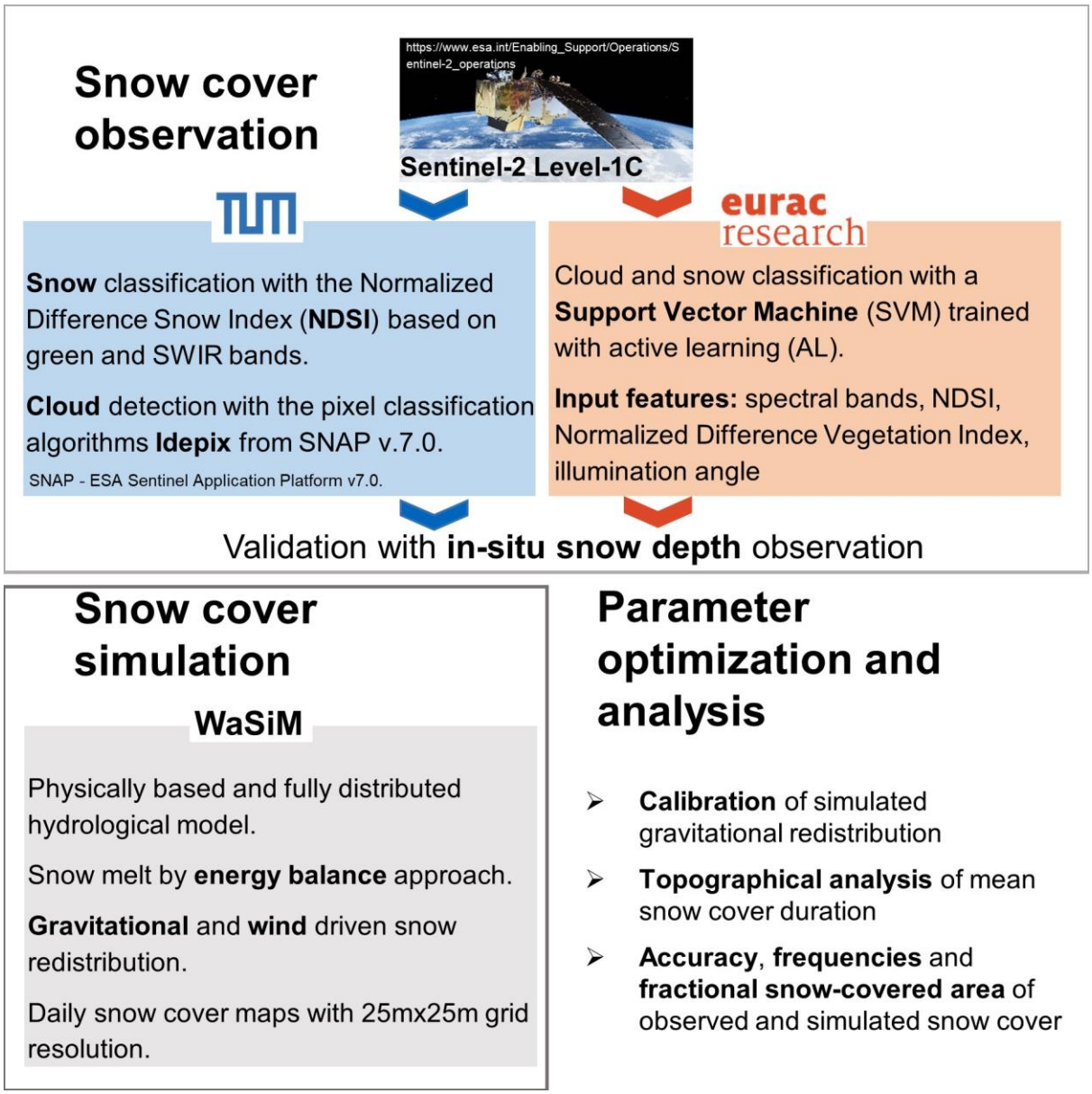


Figure 4-2: General workflows for observed and simulated snow cover maps.

4.2.2.1. TUM approach for Sentinel-2 snow cover map

Covering most of the period June 2015 to October 2020, 439 Sentinel-2 Level-1C images were downloaded with the R toolbox Sen2r (Ranghetti, Boschetti, Nutini, & Busetto, 2020) for the area of interest (AOI) including all seasons of the year. The Level-1C product type provides Top of the Atmosphere reflectance in cartographic geometry projected in UTM/WGS84 (European Space Agency, 2021). The spectral bands, central wavelength and spatial resolution of the Sentinel-2 MultiSpectral Instrument (MSI) are listed in the supplementary material of (Hofmeister, Notarnicola, et al., 2022). Using equation 4-1, the Normalized Difference Snow Index (NDSI) was calculated for each image, after resampling band 11 from 20 m to 10 m resolution (Cimpianu, 2018). The NDSI uses the distinct reflection signature of snow surfaces with very high reflectance in the visible spectrum and strong absorption in the short-wave infrared range (Gascoin et al., 2019).

$$NDSI = \frac{green - SWIR}{green + SWIR} \quad 4\text{-}1$$

where *green* corresponds to Sentinel-2 band 3 (wavelength: 0.560 µm) and *SWIR* is short-wave infrared represented by band 11 (wavelength: 1.610 µm). An NDSI pixel is classified as snow if the NDSI value is larger than 0.4 (Dozier, 1989). Larger NDSI values are classified as snow-free. The NDSI threshold was the same for all images. Several cloud detection approaches have been developed for Sentinel-2 images, such as Sen2Cor, Idepix, Fmask, MAJA and Sentinel Hub's Cloud Detector (Menekay, 2019). In this study, cloud detection was performed using cloud masks created by the Identification of Pixel Properties algorithms (Idepix) from the Sentinel Application Platform (SNAP v7.0) (European Space Agency, 2019) at 60 m spatial resolution. Idepix is a single-scene method and therefore suitable for cloud detection over changing terrain conditions, such as snow accumulation and melt. Since it is available as an SNAP plugin, it is very user-friendly, fast and effective at providing cloud masks. Idepix combines cloud masks with sun geometry to search regions of maximum probability for cloud-shadow pixels. In the projected region of potential cloud shadow, the cloud mask is shifted along the illumination path towards the surface reflectance minimum. Idepix can identify different types of pixels, such as semi-transparent clouds, fully opaque clouds, cirrus clouds, cloud shadows and even mountain shadows. This study employs Idepix_Cloud classification, which includes cloud pixels identified with full confidence as well as pixels identified with uncertainty as clouds. Clouded pixels were set to no data.

The snow and cloud maps were masked to the extent of the area of interest and resampled with the nearest neighbour method to the same spatial resolution as that of the hydrological model (25 m). We excluded highly clouded (> 80 %) images from further analysis, which reduced the total number of images to 366. Following an initial assessment of the cloud maps, we observed that the cloud maps generated with Idepix tended to overestimate cloudiness due to mountain shadows, since low illuminated areas with low spectral reflectance are misclassified as clouds. Therefore, only cloud covered areas for which no snow was detected by the NDSI were set to no data. This enables the snow detection under clouds thanks to semi-transparency in SWIR, which would otherwise be masked by the Idepix cloud mask. In addition, the cloud masks are very conservative, since they are calculated with a coarser resolution (60 m) compared to the snow maps (20 m), which would mask out too many pixels, especially in the peripheral areas of the clouds, for which the snow detection might be correct (Serco Italia SPA, 2017). Water bodies were also set to no data, since their reflectance is often very similar to that of a snow surface (Dozier, 1989).

To test the utility of atmospheric correction over satellite imagery for snow cover detection, a set of atmospherically corrected Sentinel-2 images (September 2018 - October 2019) was analysed and compared against the methods used. The correction was applied using 'Second

Simulation of the Satellite Signal in the Solar Spectrum' (6S), as developed by Vermote et al. (1997), which uses Radiative Transfer Models (RTMs) to simulate the passage of solar radiation across the atmosphere. The 6S algorithm is adapted to a Python (Py6S) interface (Wilson, 2012) and was implemented recently for use with the Google Earth Engine (Murphy, 2018) via a Python API and Docker container.

4.2.2.2. Eurac approach for Sentinel-2 snow cover map

The Eurac snow algorithm also takes the Sentinel-2 Level-1C data as its input. As pre-processing step, the Sentinel-2 images are scaled from digital number (DN) to reflectance values using the quantification value provided in the Sentinel-2 metadata (i.e., equal to 10,000 for all the considered images). Thus, all bands are reprojected and resampled to the final model resolution of 25 m using cubic interpolation. The classification algorithm consists of two steps: i) cloud detection; and ii) snow detection. In both cases, a support vector machine (SVM) classifier is trained with an active learning (AL) procedure. By dividing the procedure into two steps, it is possible to exploit the most representative features of each of the two classification problems. The AL procedure allows us to speed up the learning curve of the classification by asking the user to specify the label of the most uncertain pixels. For cloud detection, when a pixel is classified as cloud or no-cloud, all spectral bands of the Sentinel-2 are considered except for those at 60 m resolution (i.e. bands 1, 9 and 10). Moreover, the Eurac approach uses additional features, such as the normalized difference vegetation index (NDVI), calculated as the normalized difference between near the infrared (NIR) and red bands in addition to the NDSI, which have been shown to introduce benefits into the classification (Tarrio et al., 2020).

For snow detection, there are three possible classes: snow, snow-free and hard shadow. The difference between shadow and hard shadows is defined in terms of energy recorded by the sensor. If the recorded energy is too low to distinguish between snow and snow-free areas, i.e. the signal-to-noise ratio (SNR) of the sensor is too low, we call it hard shadow. This is generally the case when the sun is low on the horizon, which is from approximately mid-November to mid-February at the latitude of the Martell valley, and the terrain is particularly steep. In order to define a threshold under which the recorded energy by the sensor is too low, we selected samples for which the human photo-interpreter cannot distinguish whether the pixel is snow or snow free. Low reflectance situations, generally associated with dark areas e.g., turbid lake, shadow by cloud or terrain, flooded areas, etc., are excluded from snow detection from state of the art and operational algorithms applying thresholds on the visible bands e.g., MODIS snow cover product version 6 (Riggs & Hall, 2015). The shadow detection is performed with a SVM classifier together with the snow detection after masking out the clouds. In addition to the spectral bands and features also used for cloud detection, the

classifier includes illumination angle as a further input feature. The illumination angle is calculated from the solar zenith and the solar azimuth angle (Riano, Chuvieco, Salas, & Aguado, 2003) and enables better mitigation of the effect of the differences in the solar illumination during the year.

The SVM has been trained in order to emphasize the presence of snow also in mixed conditions. This is done by assigning the class snow to those pixels whose spectral characteristics can be attributed to the presence of snow by a careful visual inspection. For both SVM models (snow and cloud), we used a radial basis function kernel and performed a model parameters selection according to a grid search strategy to identify the regularization parameter C and the kernel coefficient gamma. The grid is initialized with a user-defined range. The model selection start with a coarse grid and then given the obtained results is refined around the values of C and gamma that perform the best. The best values are selected by evaluating the mean and standard deviation of the over accuracy calculated in a cross-validation strategy with k folder (k=5). For the cloud detection, we collected 173 training points (97 “cloud free” samples and 76 “cloud” samples). The selected parameters are C equal to 9.885 and gamma equal to 0.896. The number of selected support vectors is 98. The accuracy from the cross validation is 0.82 and the standard deviation is 0.1 by considering the 5-folds. The low accuracy during the training shows the difficulty in the cloud discrimination during winter condition with the spectral bands provided by Sentinel-2 sensor. For the snow detection, we considered three classes of training samples. We collected 177 “snow free” samples, 264 “snow” samples and 138 “shadow” samples, for a total amount of 579 samples. The selected C value is 30.018 and gamma is 0.0003. The number of support vectors was 445 and the accuracy from the cross validation is 0.98 with a standard deviation of 0.1. The samples are collected initially from two scenes and then we adopted an iterative active learning procedure to collect new samples. The active learning procedure ask the user the labels of the most uncertain samples, i.e., the ones with computed probabilities for the assigned class are low. The probability was calculated according to Lin et al. (2007).

It should be mentioned that at the time of the analyses for this publication, the Eurac approach is still in the final stage of the development but it was applied already once in a previous study (Ebner et al., 2021). The Eurac approach was chosen to verify the TUM snow detection results because the supervised algorithm should be more accurate than the unsupervised TUM product. As in the TUM method, water bodies were set to no data.

4.2.2.3. Snow cover simulation with WaSiM

The snow module of the physically-based hydrological model WaSiM (version 10.04.07) (Schulla, 2021) was used as an example for demonstrating the benefits of high-resolution snow cover maps in calibrating simulated snow redistribution. For snow melt simulation, the energy

balance approach, including gravitational snow redistribution, was selected, as originally implemented by Warscher et al. (2013). We also tested the effect of wind-driven snow redistribution (Warscher et al., 2013; Schulla, 2019). Canopy snow interception was not considered in this study because of the small proportion of coniferous forest (6 %). The multi-layer snow model was deactivated to keep the computational time reasonable, limit the number of calibration parameters and avoid the need to define multiple soil and snow parameters. The topography-dependent adjustment of radiation and air temperature follows the scheme devised by Oke (2002). The parametrization of the snow accumulation and gravitational snow redistribution Table 4-2 originates in part from the WaSiM user manual (Schulla, 2021) and Förster et al. (2018), whereas the correction factor for incoming long-wave radiation (*LWINcorr*) was slightly increased from 1.0 to 1.1 considering the observations available for May and June. The equations for energy balance approach, *Psnow*, *LWINcorr* and *LWOUTcorr* can be found in the WaSiM user manual, which is available in the supplementary material (Hofmeister, Notarnicola, et al., 2022). To account for solid precipitation undercatch of rain gauges, we used a wind-dependent snow correction factor (*Snowb*) of 0.1 s/m, as also used by Förster et al. (2018) and confirmed by Kochendorfer et al. (2016).

Gravitational snow redistribution is based on a computationally efficient, mass-conserving algorithm that parameterizes the gravitational transport and deposition developed by Gruber (2007) and later implemented in WaSiM by Warscher et al. (2013). Four parameter grids that specify the sliding fractions for each of the cardinal directions (N, E, S, W) were generated using the WaSiM preprocessing tool Tanalys. Total outflow from the gravitational model (*Mout*) is determined by equation 4-2:

$$Mout = \begin{bmatrix} ferosion \times Min - Dgrav, & if\ i \geq ierosion\ and\ SWE > 0 \\ Min - Dgrav, & if\ i\ \geq ierosion\ and\ SWE = 0 \\ Min - Dgrav, & if\ i\ < ierosion \end{bmatrix} \quad 4\text{-}2$$

where *Min* is the inflowing mass from all other neighbouring cells [mm], *SWE* is the snow water equivalent in the current cell [mm], *ferosion* is the erosion factor, which depends on the time step, *i* is the local slope, *ierosion* is the lower inclination limit for snow erosion [°] depending on the spatial resolution and *Dgrav* is the deposition in the current cell [mm]. The current snow deposition in a cell *Dgrav* is defined by equations 4-3 and 4-4:

$$Dgrav = \begin{Bmatrix} Min, & if\ Min < Dmax, grav \\ Dmax, grav, & if\ Min\ \geq Dmax, grav \end{Bmatrix} \quad 4\text{-}3$$

$$Dmax, grav = \begin{Bmatrix} \left(1 - \frac{i}{ilim} \times Dlim\right), & if\ i < ilim \\ 0, & if\ \geq ilim \end{Bmatrix} \quad 4\text{-}4$$

where *Dlim* is the upper deposition mass limit [mm], e.g. the maximum amount of snow that would be deposited on horizontal terrain, and *ilim* is the upper slope limit [°] at which all

inflowing masses will be transported to the next downslope cell(s). In this work, we show how *ierosion* and *ferosion* can be calibrated to improve the model's performance using Sentinel-2 snow cover data.

Wind-driven snow redistribution can also be simulated with WaSiM. However, the approach implemented is in fact a kind of snow precipitation correction, by which areas sheltered from the main wind direction receive an increased snow fall and areas exposed to the main wind direction receive a reduced snow fall amount. Which cells are exposed or sheltered is estimated using the directed sky view factor (*SVFdir*). For this study, the main wind direction was set to south west (180° to 270°) which corresponds with the main wind direction of the Langenferner meteorological station (mean wind direction 229° ± 54°) for the snow accumulation period (October to May). The snow precipitation is corrected by equation 4-5:

$$Psnow = Psnow + Cwind \times Psnow \quad 4\text{-}5$$

where *Psnow* is the solid precipitation and *Cwind* is the correction factor, which is determined by equation 4-6:

$$Cwind = e \times (Dmax \times (1 - SVFdir) - 1) + cmin \quad 4\text{-}6$$

where *e* is a linear elevation weighting factor, *SVFdir* is the directed sky view factor, *Dmax* is the maximum possible deposition, and *cmin* is the minimum correction factor for shifting *cwind* to a more or less solid precipitation correction. The parameter *e* ranges from 0 at the lowest elevated pixel to 1 at the highest pixel and linearly scales the amount of snow redistribution (Warscher, 2014). The impact on the mean snow cover duration was tested on different spatial scales by performing different model simulations with and without activated wind-driven snow redistribution.

Table 4-2: WaSiM parameters for snow accumulation, gravitational slides, ablation and wind driven redistribution.

Process	WaSiM Parameter	Description	Values
Snow accumulation	T0R	Temperature limit for rain (°C)	0
	Ttrans	½ of temperature-transient zone for rain-snow (°C)	0.5
	Snowb	Wind-dependent snow precipitation correction (s m^{-1})	0.1
Gravitational redistribution	i_{lim}	Maximum deposition slope (°)	55
	D_{lim}	Scaling for maximum deposition (mm)	2
	$i_{erosion}$	Minimum slope for creating slides (°)	50*
	$f_{erosion}$	Fraction of snow pack that forms the slide (-)	0.002*
Snow ablation	LWINcorr	Correction factor for incoming long-wave radiation (-)	1.1
	LWOUTcorr	Correction factor for outgoing long-wave radiation (-)	1.0
Wind redistribution	start azimuth	1st quantile of wind direction (°)	180
	end azimuth	3rd quantile of wind direction (°)	270
	cmin	Minimum correction factor (-)	0.3

*These parameters have been optimized. See Section 3.3 for more information.

The time series of all meteorological stations, illustrated in Figure 4-1 and listed in Table 4-1, are spatially interpolated with an elevation-dependent regression (i.e. temperature, wind speed and humidity) and inverse distance weighting (IDW) for precipitation and solar radiation. The simulation period covers six years from 2014/10/01 to 2020/09/30 in hourly time steps. However, the simulated snow maps were only saved as daily means. The selection of the spatial resolution (25 m) was a compromise between the level of detail and the computational demand. Previous hydrological applications of WaSiM used a 50 m spatial and 1h temporal resolution for Alpine catchments (Kraller et al., 2012; Warscher et al., 2013; Förster et al., 2018) or even 25 m in the most recent application (Thornton et al., 2021). Two different exceedance threshold values (0 mm and 5 mm SWE) for classification as either snow or no snow were tested.

To ensure consistency in the comparison with the observed snow maps, clouded areas (from the TUM product) and water bodies were set to no data in the snow cover maps produced by WaSiM. Since the dynamic glacier model of WaSiM requires a comprehensive calibration, which increases the model complexity, it was not activated, and glaciered areas were set to constantly snow covered in the simulated snow maps. Although glaciers are a key contributor to runoff generation in this area, the focus of this study is to examine the spatio-temporal variability of the snowpack. Glaciered areas were derived for each year from the cloudless snow cover map for August or September taken from the Eurac product, since the TUM product showed an tendency of false snow detection on some very steep (> 60°) north-facing slopes, which would partly lead to wrong glacier delineation.

4.2.2.4. Analysis of observed snow cover maps

The quality of snow/no snow detection is evaluated against the in-situ snow observation by calculating the accuracy score F1 (equation 4-8) based on the confusion matrix in Table 4-3. The F1 score divides the sum of all true positive (TP) and true negative (TN) matches by the total population (n). Accuracy was only computed at three snow observation stations with continuous snow depth recordings.

Table 4-3: Confusion matrix for evaluating the detected and simulated snow cover.

	Observed snow	Observed no snow
Detected/simulated snow	True positive (TP)	False positive (FP)
Detected/simulated no snow	False negative (FN)	True negative (TN)

4.2.2.5. Analysis of topographical feature on observed and simulated snow cover

The influence of topographic characteristics, e.g. elevation, slope or aspect, on the observed and simulated mean snow cover duration is analysed such that all grid cells with the same topographical characteristics (e.g. elevation, slope and exposition) are aggregated to mean snow cover duration over the entire period of 6 years following equation 4-7. The snow cover duration SCD [d] for each pixel of the domain in an hydrological year was computed following Dietz et al. (2012):

$$SCD = \frac{365.25}{N}\sum_{i=1}^{N}(si) \qquad 4\text{-}7$$

where *N* is the number of days with Sentinel-2 recorded data, beginning with 1 October and ending with 30 September of the next year. On average, the length of a year is 365.25 days considering leap years. *si* refers to the cloud-free daily snow cover data set recorded to values one for snow and zero for snow-free area. Accordingly, we calculated the mean and standard deviation of snow cover duration of all pixels of the 366 snow cover maps with respect to the three topographical feature classes (i.e. aggregation of elevation in ten meter classes, slope and exposition in three degree classes).

4.2.2.6. Comparison of observed and simulated snow cover maps

A pixel by pixel spatial analysis was performed to evaluate the simulated snow cover maps against the two different observed snow maps at the catchment scale. Two additional performance measures are thus introduced: the F2 score (equation 4-9) and the F3 score (equation 4-10), which are often used to evaluate binary classifications (Aronica, Bates, & Horritt, 2002; Warscher et al., 2013; Thornton et al., 2021). The range of the F1 and F2 scores is from zero to one while F3 ranges from -∞ to one. All scores are 1 if the simulated snow cover perfectly matches the observed one. The F1 score tends to show the highest performance

values, since it takes only true positives and true negatives into account, and these are usually high during winter and summer. F2 and F3 exclude true no snow pixels and are therefore more sensitive to differences in snow cover (Warscher et al., 2013). F3 is even more sensitive to snow extent than F2 due to the subtraction of the false positives in the numerator.

F1 score $$F1 = \frac{\sum_{i=1}^{n} TP + \sum_{i=1}^{n} TN}{n} \quad (4\text{-}8)$$

F2 score $$F2 = \frac{\sum_{i=1}^{n} TP}{\sum_{i=1}^{n} TP + \sum_{i=1}^{n} FP + \sum_{i=1}^{n} FN} \quad (4\text{-}9)$$

F3 score $$F3 = \frac{\sum_{i=1}^{n} TP - \sum_{i=1}^{n} FP}{\sum_{i=1}^{n} TP + \sum_{i=1}^{n} FP + \sum_{i=1}^{n} FN} \quad (4\text{-}10)$$

For a further spatial comparison of observed and simulated snow products, the fractional snow-covered area (fSCA) is also considered. fSCA is defined as the ratio between the number of pixels classified as snow-covered and the total number of pixels excluding clouds and lakes (Di Marco et al., 2020b), as shown in equation 4-11:

$$fSCA = \frac{Nsnow}{Ntot - Nclouds} \quad (4\text{-}11)$$

where *Nsnow* is the number of snow cover pixels according to the Sentinel-2 dataset or WaSiM, *Ntot* is the total number of pixels representing the overall catchment area, and *Nclouds* are the pixels classified as cloud and water bodies.

4.3. Results

4.3.1. Comparison of cloud detection by TUM and Eurac

In terms of the empirical cumulative cloud detection frequency, Eurac detects a higher proportion of cloud within the range of 15 % to 80 % areal coverage than the TUM product (Figure 4-3a). However, both products are able to detect complete cloud cover and clear sky situations equally well. Besides cirrus (high-altitude clouds) and other clouds, the image processing algorithms also include topographic shadows, which are of particular interest in complex terrains such as the upper Martell valley. The ability to detect topographic shadows is especially important in mid-latitude areas during the winter period, where the solar elevation is very low, typically below 20° (Gascoin et al., 2020). Differences between the two approaches and their limitations in terms of cloud and shadow detection are illustrated in an example for January 27 2017, when no clouds were present in the AOI (Figure 4-3b). However, the low inclination angle of the sun produces extensive shading in the southern and eastern parts of the AOI. Cloud detection by TUM (Figure 4-3c) falsely classifies some snow-free ridges or

snow-free steep slopes as clouds. On the other hand, the Eurac method (Figure 4-3d) detects more shaded areas in the south eastern area. However, both approaches fail to fully detect the topographic shadows on the forested southern valley bottom, which are classified as snow-covered regions.

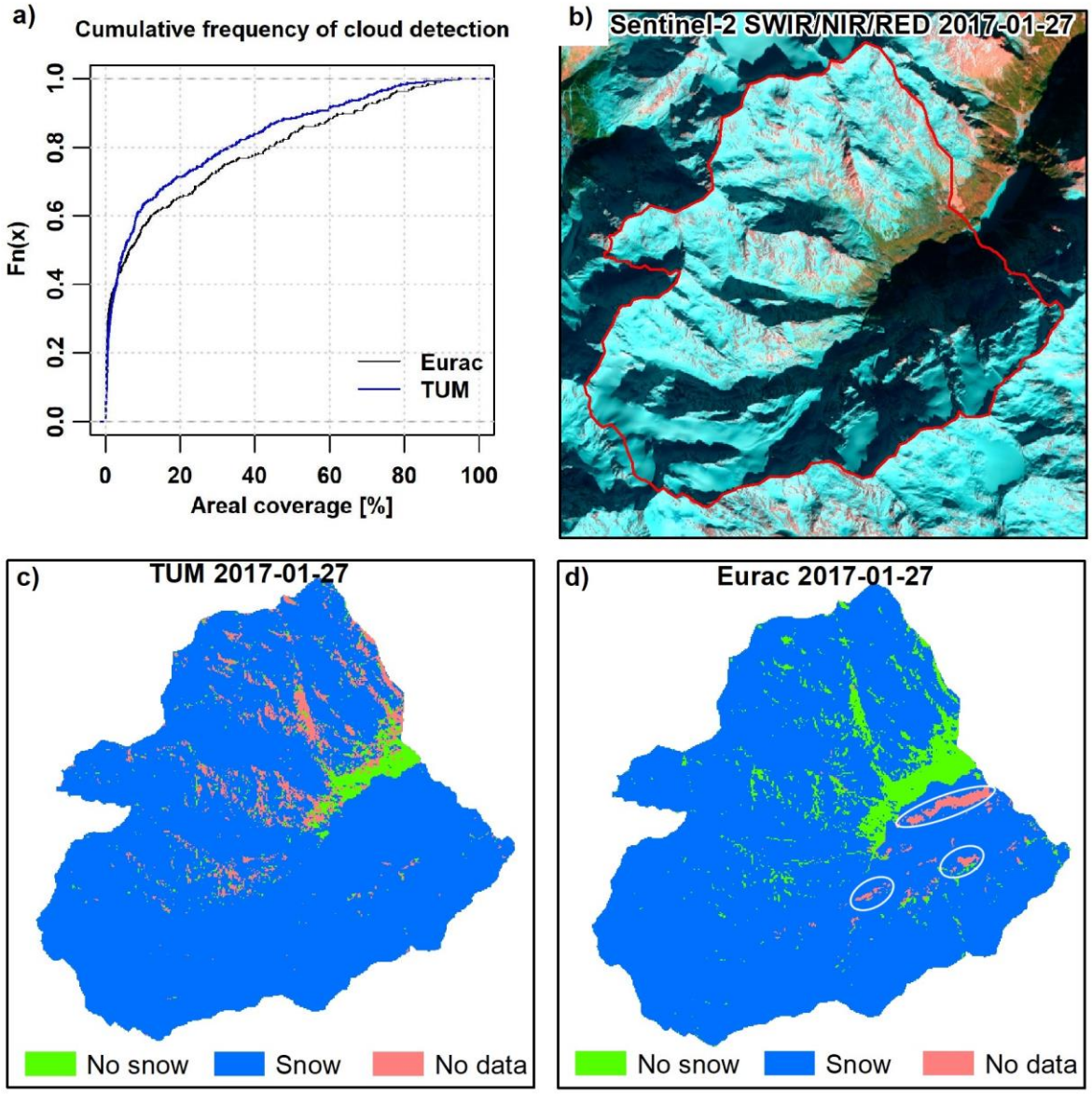


Figure 4-3: Empirical cumulative frequency of cloud detection by TUM and Eurac for the whole data set (a), Sentinel-2 false colour composition (SWIR/NIR/RED) in 20 m resolution for one case study on January 27 2017, showing the AOI boundary (red line) (b) and the cloud and snow detection by TUM (c) and by Eurac (d) in 25 m resolution, for the said date. No data pixels were considered as cloud or shadow by the algorithms. White circles in panel d indicate main differences between TUM and Eurac in detecting topographic shadows.

4.3.2. In-situ comparisons of snow cover detection by TUM and Eurac

The quality of snow detection was tested against three snow observation stations for various snow depth thresholds that differentiate between snow and no snow. The detection accuracy was calculated by the F1 score (equation 4-8). Snow detection varies not only according to the

chosen detection method (TUM or Eurac) but also according to the observation sites (Figure 4-4 and supplementary table A.3.7 in Hofmeister, Notarnicola, et al. (2022)). The most accurate score (F1 > 0.97) was obtained with both products for snow depth thresholds ranging between 4 cm and 15 cm for the Rossbaenke station, which is surrounded by grassland. Both TUM and Eurac attained a similar detection accuracy at the Madritsch station, however decreasing with increasing snow depth threshold. The highest accuracy of 0.96 (TUM) and 0.97 (Eurac) is reached at a snow depth threshold of 1 cm. Due to the station's relatively high-altitude (2825 m a.s.l.), the land cover of the surrounding area is mainly composed of gravel and debris. Therefore, it can be assumed that there is no interference between the spectral characteristics of the snow cover and those of the vegetation. The main differences between the snow detection products are observed for the lowest snow observation station of Zufritt, which is surrounded by forest and grassland. Whereas the TUM product has a detection accuracy larger than 0.91 for snow depth thresholds between one and ten cm, the detection accuracy of the Eurac product does not exceed 0.83 for the same threshold range. The lowest performance is probably due to the presence of permanently mixed pixels, such as forest and grassland. In fact, the spectral characteristics of these pixels do not show a snow presence greater than 50 % (see 4.2.2.2). In this case, a canopy correction is required to prevent error. The optimum snow detection accuracy of 0.95 for TUM and 0.92 for Eurac based on all stations was found at a 4 cm threshold, showing that, in general, a larger snow quantity, i.e. a closed snow cover with a few centimetres depth (~4 cm), is needed on the ground to enable better identification of snow from Sentinel-2 images.

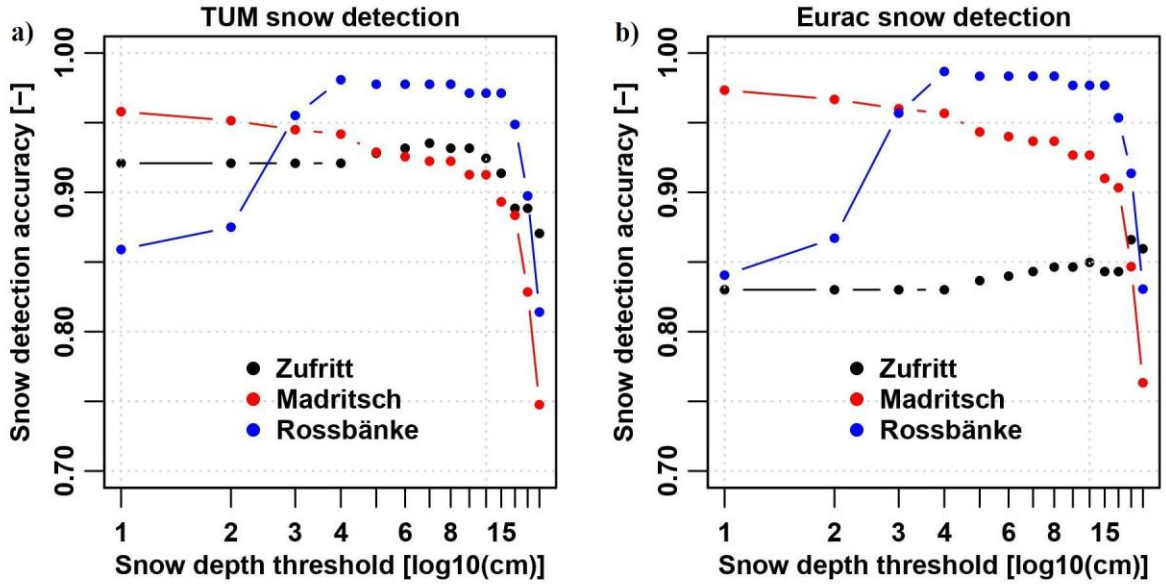


Figure 4-4: Snow detection accuracy obtained with TUM (a) and Eurac (b) at three snow observation stations for different snow depth thresholds, shown on a log scale.

We also tested two different exceedance thresholds for simulated SWE (0 mm and 5 mm) in order to convert SWE to binary snow cover maps. The commonly used exceedance threshold

of 5 mm (Warscher et al., 2013; Schulla, 2019; Thornton et al., 2021) attained a slightly higher F1 prediction accuracy of 0.95, compared to 0.92 for 0 mm at the Madritsch snow observation station. However, since only one snow observation station with continuous snow depth recording is located in the research area, it is not possible to make any general statements about the robustness of the exceedance threshold.

4.3.3. Calibration and analysis of snow redistribution as modelled by WaSiM

Topographical analysis not only allows the impact of topographic features such as elevation, slope and aspect on the mean snow cover duration to be investigated, but also enables a process-orientated evaluation of the simulation results and calibration of the gravitational redistribution simulated by WaSiM with respect to different slope gradients. A total of four simulation results were analysed with different snow redistributions (Table 4-4). Simulations A and C were performed with the default WaSiM parametrization of *ierosion* (minimum slope for creating slides) and *ferosion* (fraction of snow pack that forms a slide), while simulations B and C used optimized parameters. For this analysis, all pixel values of all 366 snow cover maps were aggregated by each topographic feature class (elevation, slope and exposure) and the mean and average standard deviation of the snow cover duration (SCD) was estimated for each class (Figure 4-5), as explained in section 4.2.2.5. Although the elevation gradient of the mean snow cover duration and the standard deviation bands (quantified for one mean standard deviation) are very similar for both observation and simulation in the range 2300 to 3500 m a.s.l., distinct differences can be observed in the lower (< 2300 m a.s.l.) elevation zones. The Eurac snow maps distinctly show lower mean snow cover duration values than TUM and WaSiM, accompanied by broadening of standard deviation band at the valley bottom (< 2300 m a.s.l.) due to the presence of forest and a higher mountain shadow detection (supplementary Figure A.3.6 in Hofmeister et al. (2022)). Although the TUM product also displays higher variance in the low elevation range, it shows better agreement with the simulated snow cover duration. In the elevation range from 2800 to 3400 m a.s.l., relatively wide standard deviation bands indicate a large spatial heterogeneity in mean SCD. Mean SCD varies only slightly for the highest elevation bands (> 3500 m a.s.l.).

The impact of *ierosion* on the mean SCD is visible on the slope feature class in Figure 4-5. We can observe that an *ierosion* value of 45° does not reproduce the turning point in the mean SCD, as it is present in both the TUM and Eurac products. Hence, an analysis of the satellite data enables us to determine the value of *ierosion* by applying a grid search approach to find the minimum distance between the peak of mean SCD at slopes between 40° to 50° in observation and model results. The optimum *ierosion* value was found at 50°. The *ferosion* parameter was lowered in 0.001 steps from the default value (0.007) to 0.001 and the best

parameter (0.002) determined taking into account the best combination of mean Pearson correlation (0.73) and RMSE (mean SCD of 25.7 [d]) with *ierosion* set to 50° (see supplementary Figure A.3.8 in Hofmeister, Notarnicola, et al. (2022)). Wind-driven snow redistribution was activated for the simulation runs C and D.

Table 4-4: WaSiM configurations for topographical analysis.

WaSiM Configuration	**Simulation A**	**Simulation B**	**Simulation C**	**Simulation D**
$i_{erosion}$	45°	50°	45°	50°
$f_{erosion}$	0.007	0.002	0.007	0.002
Wind redistribution	No	No	Yes	Yes

The small differences in performance of the simulated snow cover duration in the highest elevation zones (> 3500 m a.s.l.) can be explained by the steeper topography, in which more snow is redistributed by gravitational slides. The greatest simulated gravitational redistribution occurs with the default parametrization of *ferosion* (Figure 4-5a) in this elevation zone. In contrast, the optimized gravitational redistribution (Figure 4-5b) shows better agreement with the observed mean SCD. The activation of wind-driven redistribution leads to a distinct reduction in the mean snow cover duration in the range from 2800 to 3500 m a.s.l. (Figure 4-5b and d).

The effect of gravitational redistribution is most clearly visible on slope gradients steeper than 30°. With the default parametrization of *ierosion* (45°) and *ferosion* (0.007), WaSiM relocates snow from steep slopes (> 45°) to flatter areas, which results in very low snow cover durations for slopes above 45°, accompanied by a sudden change in the standard deviation (Figure 4-5a and c). The wind-driven redistribution has barely any detectable effect on the simulated mean snow cover duration on the slope feature class, with the exception of the steepest slopes (> 60°), for which a small increase can be observed (Figure 4-5b and d).

The final topographical characteristic to be evaluated is aspect (exposure). The mean SCD on areas of different exposure depends mainly on the incoming solar radiation (topographical shadowing) and on wind-driven snow redistribution. The influence of exposure on the mean SCD is clearly visible in Figure 4-5, with the highest values (> 250 d) on north- and north east-facing slopes and the lowest (< 200 d) on south-facing slopes. As with the slopes, the relatively large standard deviation bands indicate a high variability of SCD in areas with similar exposure. Although the mean SCD shape is very similar for the TUM and Eurac products, a systematic offset between both products (overall mean of SCD by TUM 241 d and Eurac 226 d) is present, which is more pronounced with the northern exposure. This can be explained by the differences in cloud and mountain shadow detection by the products. While the Eurac cloud product shows no influence of aspect with a constant overall mean cloud coverage of around

20 %, the TUM product detects more clouds on south-facing areas (overall mean cloud coverage of 20 %) than on north-facing ones (overall mean cloud coverage of 14 %) (supplementary Figure A.3.6 in Hofmeister, Notarnicola, et al. (2022)). Activated wind-driven redistribution leads to better agreement between the observed and simulated mean snow cover durations on areas from east to west exposure (Figure 4-5d). In the following comparison, we will refer only to the best WaSiM setup with the optimized gravitational snow and activated wind redistribution (configuration D).

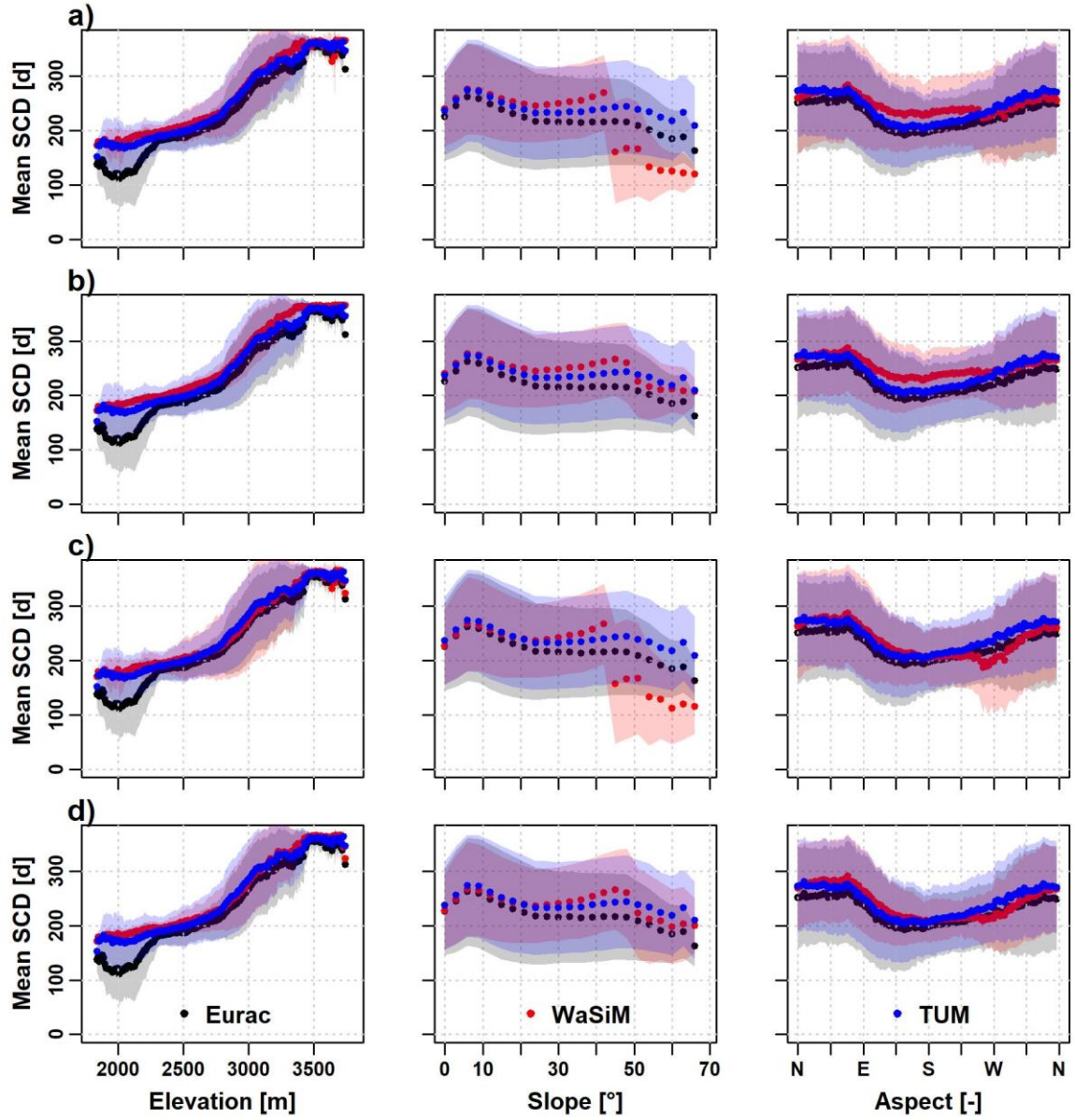


Figure 4-5: Topographical analysis of observed and simulated mean snow cover durations with default parametrization of gravitational redistribution and without wind redistribution (a), with optimized parametrization of gravitational redistribution and without wind redistribution (b), with default parametrization of gravitational redistribution and activated wind redistribution (c) and with optimized parametrization of gravitational redistribution and activated wind redistribution (d) over elevation, slope and aspect plus corresponding mean standard deviation.

4.3.4. Comparison and accuracy of observed and simulated snow cover products

The cumulative snow and no snow frequency curves show a strong similarity in the snow detection and simulated values (Figure 4-6a). The cumulative no snow frequency curves diverge with no snow coverage < 10 % in the lower elevation ranges of the AOI. Although the frequency response with snow detection is quite similar, there is a systematic offset between the snow detected by TUM and Eurac from 10 % to 90 % areal coverage. Eurac tends to detect a lower snow cover probability than TUM. This is because the Eurac classification was trained with the majority of mixed pixels being considered as snow-free. The simulated snow cover frequencies follow the same response as the detections, but the model has a slight tendency to overestimate the snow covered area within the 60 % to 90 % areal coverage range.

Overall accuracy scores for both TUM and Eurac snow detection are very high (F1 = 0.97, F2 = 0.91, F3 = 0.9), only dropping in August to 0.8 for F2 and F3 respectively, as can be seen in Figure 4-6b. The overall accuracy scores of TUM-WaSiM (F1 = 0.93, F2 = 0.83, F3 = 0.75) and Eurac-WaSiM (F1 = 0.93, F2 = 0.83, F3 = 0.71) provide further evidence that the model is able to simulate the spatial snow distribution in the research area well. Comparison of the mean scores on a monthly scale reveals the part of the season in which the model has a reduced accuracy in the snow cover distribution simulation. The critical months are the beginning of the accumulation periods (October, with an overall mean fSCA of 45 % (Eurac), 57 % (TUM) and 56 % (WaSiM)) and ablation (June, with an overall mean fSCA of 52 % (Eurac), 55 % (TUM) and 58 % (WaSiM)), although they still have a mean F1 score of above 0.8. The F2 and F3 scores exclude the no snow pixels and accordingly result in lower scores (Warscher et al., 2013).

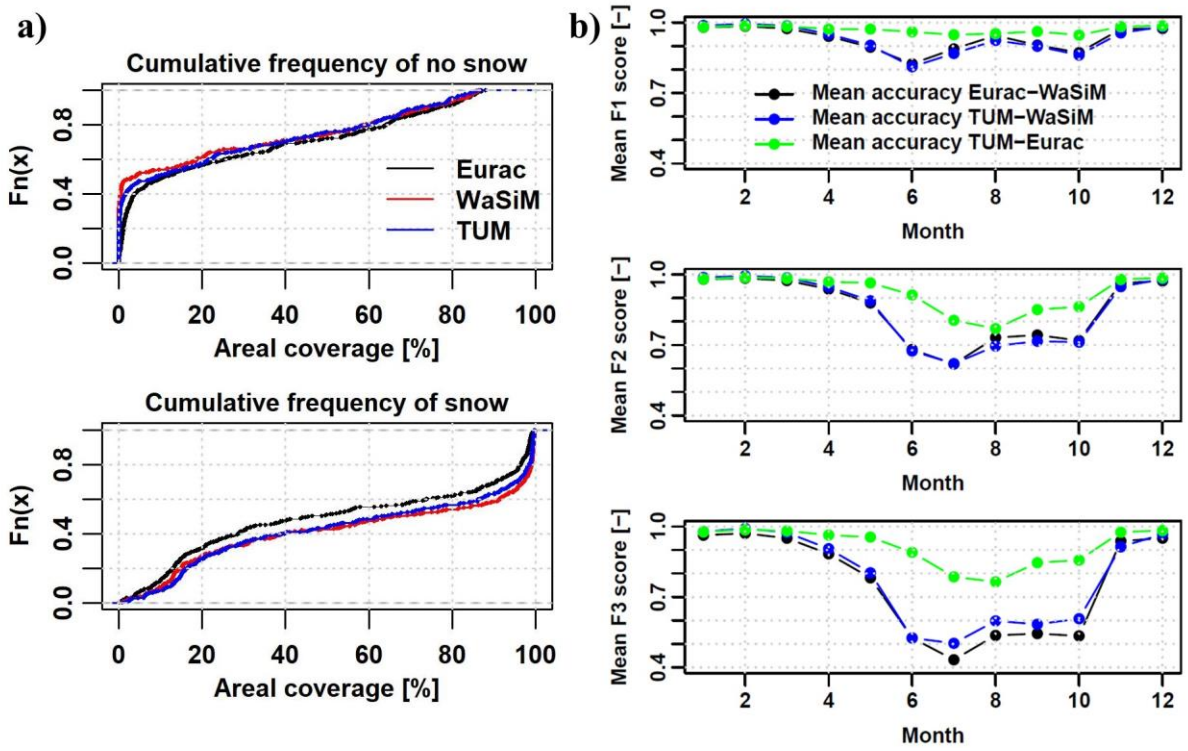


Figure 4-6: Empirical cumulative frequencies of no snow and snow (a) and mean monthly accuracy of observed and simulated snow cover areas (b).

4.3.5. Comparison of snow cover frequencies and fractional snow-covered area (fSCA)

Differences between observed and simulated snow cover days (SCDs) at the pixel scale were investigated in more detail for each hydrological year (1 Oct. - 30 Sept.), as illustrated in Figure 4-7. In general, the hydrological years 2016 and 2017 had a shorter snow coverage period than the subsequent three years (2018, 2019 and 2020). The lower snow coverage in 2016 and 2017 results in better visual agreement between the two snow observation products. However, the relatively high standard deviations (> 74 d) given in Table 4-5 indicate a greater variance in the snow cover days within the AOI. WaSiM simulates a higher snow cover duration for both years, which also results in a slightly lower standard deviation of SCDs (< 73 d). A distinct tendency towards overestimation of the SCDs can be observed, especially for the hydrological year 2016.

For the snow rich years 2018, 2019 and 2020, TUM shows a higher frequency of mean SCDs in the range above 250 d than Eurac. In terms of the observed and simulated SCDs, the WaSiM results are more similar to those of TUM snow detection for the corresponding period than they are to those of Eurac. With increasing mean SCDs, the standard deviation decreases, indicating a lower heterogeneity in the snow cover duration. The overall mean gives the mean and standard deviation of each snow product for the whole investigation period 2016-2020. The overall mean SCDs of 236 d confirms the tendency of Eurac to underestimate the snow cover duration as compared with TUM (251 d) and WaSiM (258 d) and also shows the highest

variability in snow cover duration, with a mean standard deviation of 64 d. The standard deviation shows that the snow cover duration varies greatly from year to year and declines in snowy winters with long snow cover durations (such as in 2018/2019). Moreover, the standard deviation of SCDs obtained from WaSiM is systematically lower than the one observed without taking canopy interference into consideration.

Table 4-5: List of means and standard deviations (SD) of snow cover days (SCDs) for the different snow products and for each hydrological year.

Hydrological year	Mean SCDs Eurac	Mean SCDs TUM	Mean SCDs WaSiM	SD SCDs Eurac	SD SCDs TUM	SD SCDs WaSiM
2016	221.16	233.20	261.94	77.04	75.99	72.20
2017	206.89	222.55	230.27	74.10	74.33	66.39
2018	218.20	241.58	240.49	63.57	61.59	63.73
2019	275.96	288.79	289.06	50.57	43.62	40.17
2020	256.23	269.78	270.42	55.67	51.44	47.83
Overall mean	235.69	251.18	258.44	64.19	61.40	58.07

Figure 4-7 shows the comparison among the frequencies of snow cover days and reveals the tendency of WaSiM to underestimate the snow cover duration in the 360 SCDs bin and to overestimate it in the following bin (> 360 SCDs) corresponding to perpetual snow cover. Moreover, WaSiM simulation results in a longer snow cover duration than the two observation products for the classes 260 d and 270 d in the extreme winter 2018/2019.

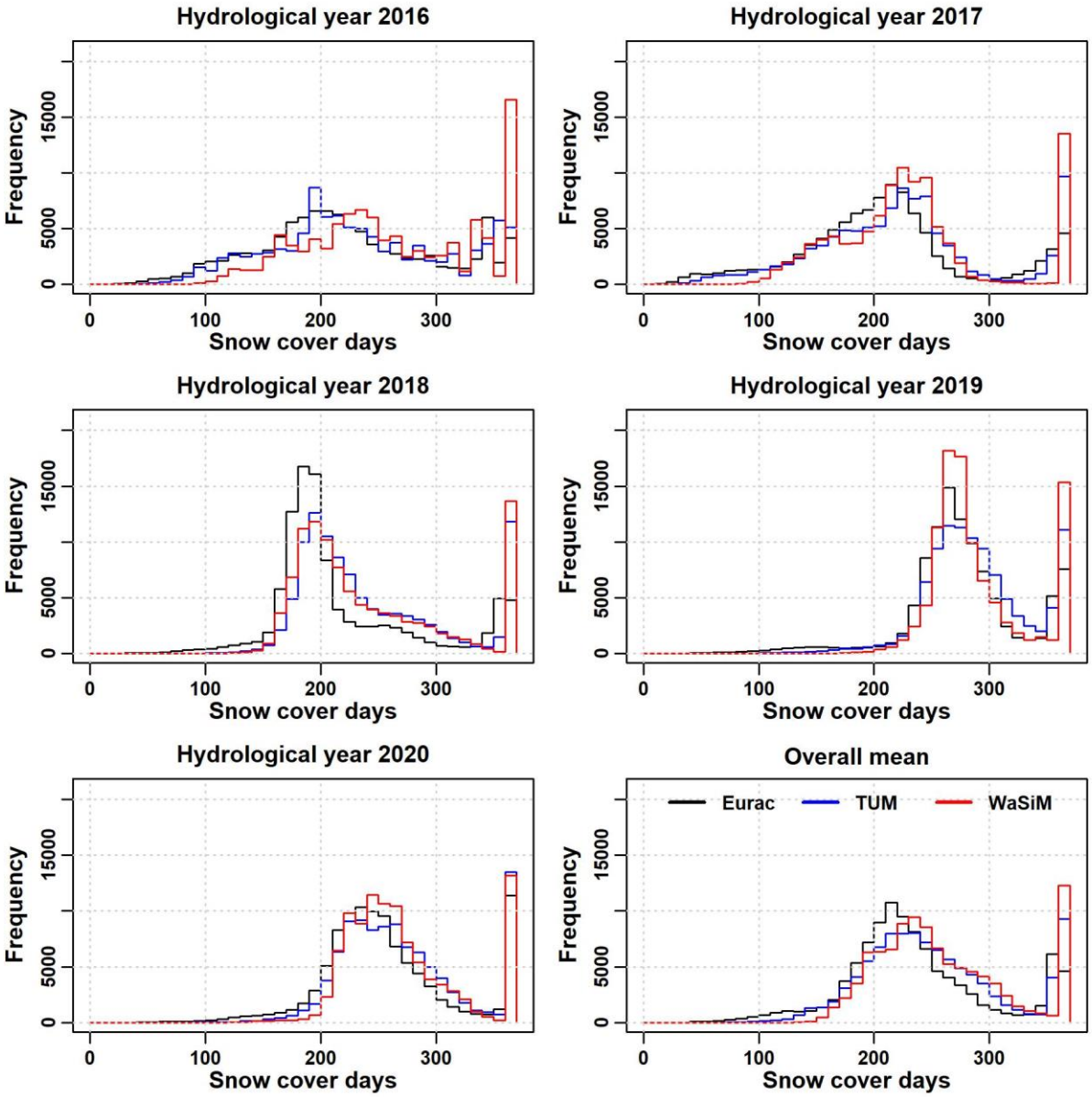


Figure 4-7: Frequencies of observed and simulated snow cover days (SCDs) in bins of 10 SCDs for each hydrological year and as average means.

Given that the frequency of snow cover days does not allow any spatial comparison between the two approaches and the simulated maps, Figure 4-8 compares case studies for a hydrological year with (2020) and without (2019) good agreement. Eurac shows differences in snow detection at the valley bottom in forested areas and on steep slopes for both winters (2019 and 2020), these often being set to no data due to the presence of shadows. Moreover, the presence of clouds impacts snow and ice detection on the glaciered areas. The TUM snow cover product mostly has problems with overestimations of snow cover at the bottom of steep north-facing slopes. Topographical features such as ridges and slopes are highly visible in both products, as is interference from vegetation, especially with respect to the year 2020. As snow cover duration increases, topographical features become less visible, as is the case for 2019. However, Eurac also underestimates the snow cover days on the forested valley bottom

in the snow rich year 2019 when compared with TUM. The hydrological model WaSiM is well able to simulate the spatial distribution of snow cover days for both seasons, but overestimates the snow cover duration at the valley bottom and to some extent at the highest elevations (> 3000 m a.s.l.) for the snow rich year 2019. The main topographical features (ridges and slopes) are also represented by the model. Activated wind redistribution leads to a longer snow cover duration on north east-facing slopes, which are also partially present in the observation products. In contrast, WaSiM underestimates the snow cover duration in west-facing areas due to wind-driven snow erosion, as is the case for the Madritsch snow station and the surrounding area. Canopy snow interception was not considered in the model and is therefore not discernible in the valley bottom.

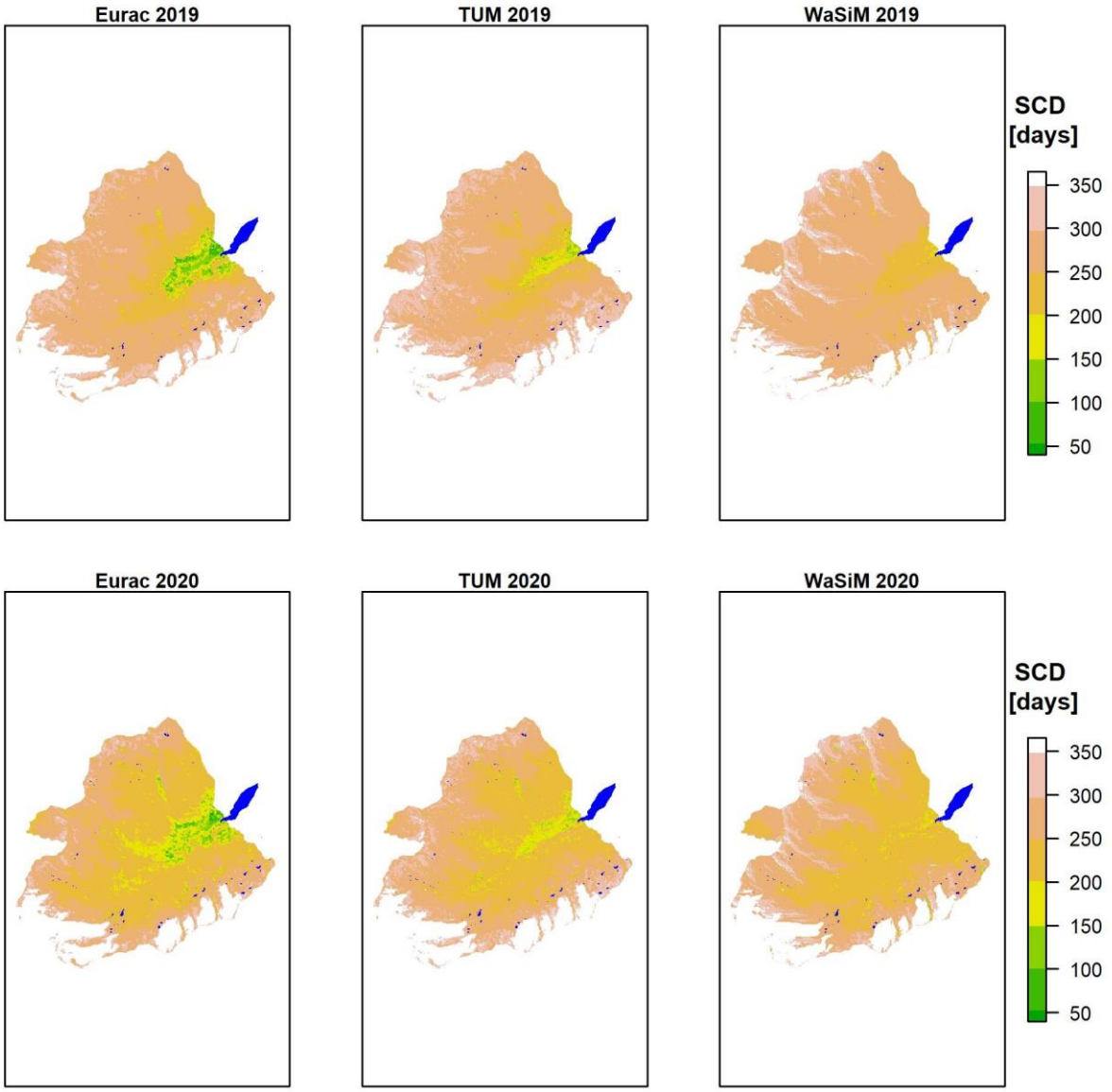


Figure 4-8: Observed and simulated snow cover maps for 2019 (above) and 2020 (below).

The fractional snow-covered area (fSCA) of the two observation products and the simulation results were evaluated for a more quantitative comparison. The fSCA shows strong seasonality

in all products (Figure 4-9). Although the simulated snow coverage given by WaSiM is in line with that observed by Eurac, there are larger fSCA in the ablation period (June and July) of the years 2016 (+14.2 %), 2017 (+3.2 %) and 2020 (+8.6 %) but also for new snow events in August and September of the years 2017 (+8.5 %) and to some extent 2018 (+5.5 %). This pattern is mainly present when comparing WaSiM with TUM for some dates of the years 2016 and 2017. However, these differences are not as pronounced as for the Eurac product. As an overall mean, WaSiM overestimates the Eurac fSCA by 5.1 % and slightly underestimates the TUM fSCA by 0.6 % during the ablation period (June and July). The overall Pearson correlation of fractional snow-covered areas is 0.98 for Eurac and WaSiM and 0.98 for TUM and WaSiM.

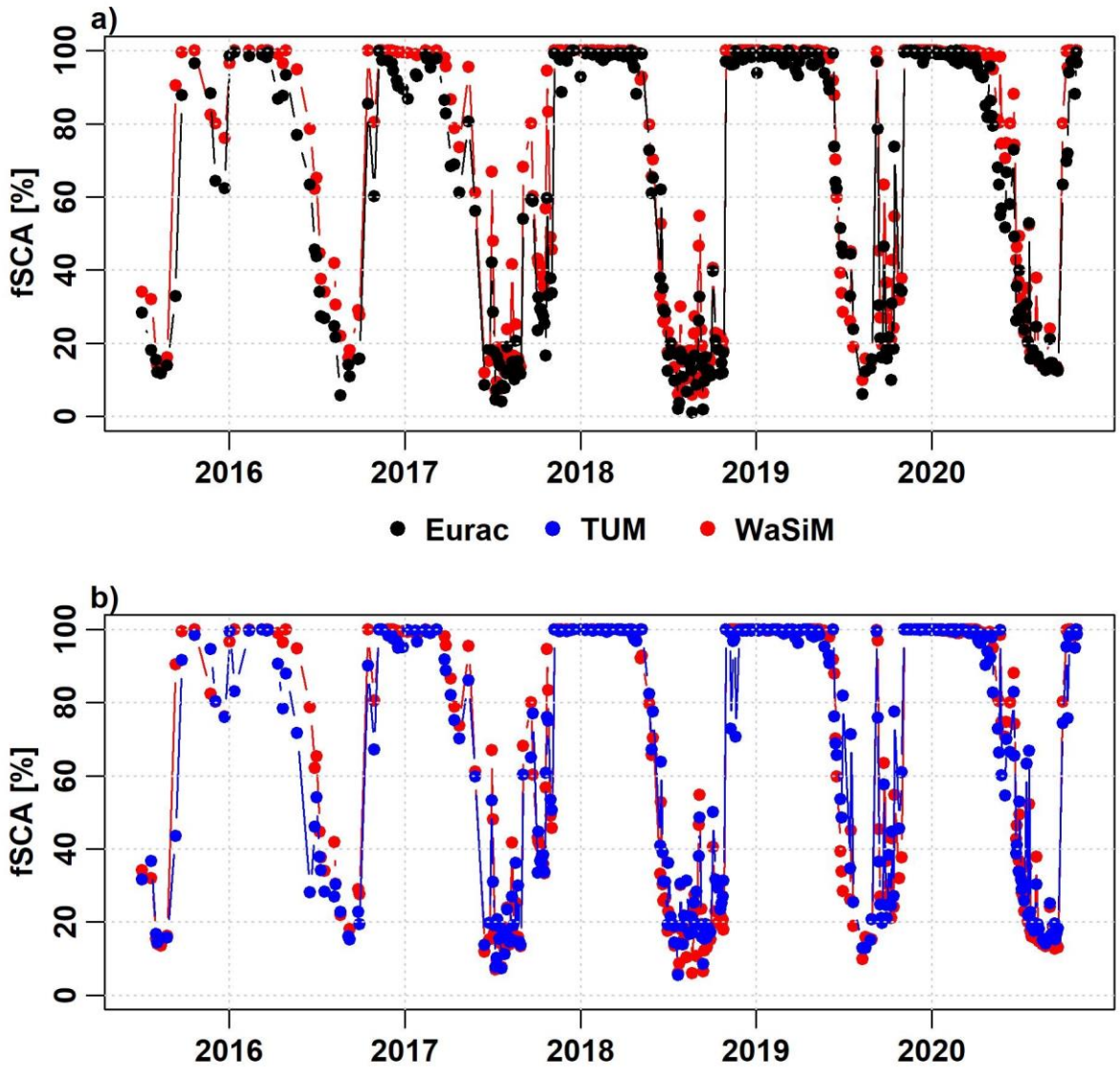


Figure 4-9: Fractional snow-covered area (fSCA) for each image and snow product, comparing Eurac and WaSiM (a) and TUM and WaSiM (b).

4.4. Discussion

4.4.1. Cloud detection by TUM and Eurac

In this work, we tested the extent to which a simple snow and cloud detection approach (TUM) performs in a complex terrain and determined what limitations might occur. Previous evaluations have found that Idepix overclassifies possible cloud shadow pixels (Alvera-Azcárate et al., 2021). Although Idepix has been tested for a wide selection of regions, it still has limitations and weaknesses in cloud detection that have not been completely resolved. The main difficulties are in its ability to distinguish between cloud and snow/ice and to detect optically very thin clouds (Brockmann Consult GmbH, 2017). In a visual inspection for the years 2015 and 2016 based on Idepix cloud products, we observed that snow was systematically falsely classified as cloud in clear sky situations during the winter due to the presence of topographical shadows (see supplementary Figure A.3.9 in Hofmeister et al. (2022)), which would lead to an underestimation of snow cover if the erroneously detected clouds are set to no data. Snow detection by NDSI was therefore given a higher confidence rating than cloud detection by Idepix and clouded areas were only set to no data when the binary snow detection value was zero. This resulted in false snow detection for a few pixels on steep shaded slopes with low illumination. This effect is also shown by TUM on the slope feature class (Figure 4-5), where the variance in the mean snow cover duration increased for very steep slopes (> 60°) when compared to Eurac. The complementation by the Idepix Mountain_Shadow layer might reduce false snow and cloud detection on steep and shaded slopes.

The Eurac algorithm is based on an ML approach that uses multi-temporal training samples. Even if the samples are collected over a large variety of scenes, including different illumination conditions, a single model is used and the quality of the classification still depends on the conditions of the scene. This is mainly due to the limited spectral information provided by Sentinel-2, which is insufficient for solving all ambiguities. Unlike the single-scene cloud detection of Idepix, the Eurac product is able to classify mountain shadows on north-facing slopes, as can be seen in the case study on cloud detection in Figure 4-3d. Additionally, it has the highest snow detection confidence, even in steep terrain, as can be concluded from the narrow uncertainty band in Figure 4-5. Since it uses an active learning approach, the cloud and snow detection quality relies on the user selecting the appropriate training pixels. It is sometimes difficult to visually assign the correct class to pixels showing mixed characteristics, such as shadowed or forested areas. In these pixels, the classifier usually returns an uncertain probability (i.e. around 50 %) of the pixel belonging to the snow class. In general, cloud detection by both Idepix and ML requires deeper investigation over forested areas in Alpine catchments. Moreover, a comparison with other available remote sensing snow cover products (e.g. Gascoin et al., 2019; Di Marco et al., 2020) may be beneficial for the future.

4.4.2. Validation of snow detection quality through in-situ snow observation

The individual selection of a snow depth threshold is recommended when comparing remote sensing and ground snow cover information. In fact, each snow observation site has its own characteristics, such as elevation, aspect, exposure to wind, and land cover, that influence this comparison. Due to the additional interference between vegetation and snow cover detection, validation by in-situ snow observation is more accurate for sites located above the forest line. Moreover, the mixed pixel issue of different spectral signatures can impact the binary snow and no snow information. For small snow depth thresholds (< 4 cm), a high snow detection sensitivity can be observed at the Rossbaenke site due to interference of grass cover with the snow (Figure 4-4). One way of reducing this error would be to filter the snow depth observations to avoid noise caused by the ground (i.e. grass). Nevertheless, a good level of accuracy can be achieved for lower lying sites that are surrounded by coniferous forest and grassland, as is the case for the TUM snow product. Barrou Dumont et al. (2021) found an optimal threshold value even at 1 cm snow depth by analysing a very large data set of 1764 in-situ snow observation stations. Differences in the snow cover detection accuracy by Sentinel-2 was also noted to be dependent on the land cover type. It was also found that the differences in the accuracy of snow cover detection by Sentinel-2 depend on the type of land cover. In particular, lower accuracies were obtained in closed forests and near water bodies (Barrou Dumont et al., 2021). Nevertheless, systematic false detection, such as in topographic shadows, can only be detected by comparing with another snow and cloud detection method, since most snow observation stations are located in flat terrain and are not shaded.

Comparison between simulated SWE and in-situ snow observation at Madritsch undermined the use of the commonly used SWE threshold of 5 mm to distinguish between snow and no snow. However, this only represents a point comparison at a single station. It might well be that the threshold sensitivity varies from site to site.

In a small case study (September 2018 - October 2019), we tested the utility and effect of atmospheric correction on snow detection with Sentinel-2. From this evaluation, atmospherically (AC) and non-atmospherically (nAC) corrected results mostly differ on lower elevated areas. Using the AC dataset, snow detection accuracy increased for the Madritsch observation station (from 0.85 to 0.94 at a 4 cm threshold) but decreased for the lower lying Zufritt station (from 0.78 to 0.74 at a 4 cm threshold). The full figure is available in the supplementary material (A.3.10 in Hofmeister, Notarnicola, et al. (2022)). Major differences can be seen in the topographical analysis, where the use of the AC dataset led to underestimation of snow at lower valley locations. In this elevation range, AC shows a stronger similarity to the Eurac product (supplementary Figure A.3.11 in Hofmeister, Notarnicola, et al.

(2022)). In contrast, AC overestimates the snow cover duration in the 2500 to 3500 m elevation range. Moreover, the performance in southern to north western areas no longer agrees with Eurac.

For this analysis, AC is useful for improving snow detection on the point scale, as is the case for the Madritsch station, but leads to underestimation of the snow cover duration at lower elevations (< 2500 m) with weaker illumination and in areas with a southern to western aspect, such as the Zufritt station. Comparing the gains and losses in accuracy for both stations, we conclude that it is important to critically evaluate the area studied in order to decide whether to use an AC. The application of these and other AC algorithms may significantly improve accuracy, although, in some instances, its use does not appear to offer sufficiently significant improvements to justify the larger time requirements and higher computational demand for detecting snow cover.

4.4.3. Topographical analysis and optimized WaSiM setup

The topographical analysis revealed differences between the snow cover observations from Sentinel-2 and simulation with the standard WaSiM setup for snow redistribution (Figure 4-5). Both snow observation products and the snow model are able to reproduce the increase in mean snow cover duration with increasing altitude. Larger differences can be observed in particular for lower lying areas (< 2200 m a.s.l.). The significant underestimation of snow cover duration by Eurac is due to its difficulties in detecting snow in forested areas compared to TUM and WaSiM. The large standard deviation of the observation products undermine the large variance in snow detection for the corresponding elevation zone, as one elevation zone can include pixels with different slopes, aspects and land cover.

The overestimation of gravitational redistribution can be reduced by optimizing the erosion factor (*ferosion*) (Figure 4-5a and b) for the corresponding elevation zone. The simulation of gravitational redistribution is particularly important for avoiding so called “snow towers” at the peaks and ridges. Since the *ferosion* parameter depends on the model time step, it must be recalibrated if, for instance, the model is run with a daily time step. In general, changes in the erosion factor within the range 0.001 to 0.007 with an hourly model time step significantly affected the model results on slope gradients steeper than 30°. Besides the erosion factor, the *ierosion* parameter, which determines the threshold for commencing gravitational redistribution was adjusted from 45° to 50° (Figure 4-5a and b). There is no change in the slight overestimation of the simulated mean snow cover duration on slopes from 40° to 50°, even after parameter optimization. It is therefore assumed that with larger snow accumulations, the snowmelt is delayed by the surface energy balance model. The activation of wind-driven snow redistribution impacts mainly the mean snow cover duration on steep slopes with inclines of

over 60° due to the intense snow accumulation in this class (Figure 4-5d). This shows the limitation of a global erosion factor that is constant in space and time.

Moving to a larger scale, the deficits on the slope feature class are levelled out when comparing the mean snow cover duration against aspect. Several previous studies have determined the effect of aspect on snow cover duration. South-facing slopes receive more solar radiation, which leads to more rapid snow melt and a shorter snow cover duration than on north-facing slopes in the northern hemisphere (López-Moreno & Stähli, 2008; Bouamri et al., 2021). The two observation products mainly differ for north and north west exposed areas, where TUM tends to detect more snow. This is partly due to false positives in shaded areas and lower cloud detection in north-facing areas (supplementary Figure A.3.6 in Hofmeister, Notarnicola, et al. (2022)). Although WaSiM was able to adjust the radiation and temperature with respect to the topography, it was determined that the mean snow cover duration was overestimated for south-facing areas (Figure 4-5a and b). This overestimation can be reduced by taking into consideration the wind-driven redistribution with a constant south-westerly wind direction (Figure 4-5d). Generally, north-west to north-east located areas indicate the largest spatial variability of mean SCD. Though the simulated mean snow cover duration improved on the catchment scale, this simple approach is unable to consider minor or major differences in wind fields and speeds. Comparing the dispersion of standard deviation of all topographical features, it can be seen that the features with the largest spatial variability are slope and aspect.

4.4.4. Accuracy and frequencies of observed and simulated snow cover and fSCA

Slight differences in the cumulative frequencies of no snow detection occurred mainly for the lower elevation zones, where TUM detects more snow areas than Eurac. It may be that Eurac underestimates snow cover in forest areas (Figure 4-4b) and TUM falsely detects snow in areas with low illumination. Since the canopy snow interception model of WaSiM was not activated, the model tends to overestimate mean snow cover durations when compared to observation products at the same elevations (Figure 4-5). Besides the systematic offset in cumulative snow detection frequencies between TUM and Eurac from 10 % to 90 % areal coverage, both approaches agree very well on the presence or absence of full snow coverage (Figure 4-6a). The simulated snow covered area approaches the Eurac product for low snow coverage since the glacier extents originate from this product. WaSiM tends to overestimate snow coverages of between 50 % and 95 % when compared to the observation products. It follows from the distinct decrease in all three accuracy scores in Figure 4-6b that WaSiM has difficulties in simulating the snow cover recession in the snow melt period (from May to July). The delayed snow cover recession results in an overestimation of the snow cover for these

months, which is also evident from an analysis of the fractional snow-covered area (Figure 4-9). This delay can have different reasons. For instance, the long-wave incoming and outgoing parameters (*LWINcorr* and *LWOUTcorr*) need a fine tuning based on the snow cover recession during the ablation period (May to July). Consideration of the snow-covered area in a multi-objective calibration approach would improve the model performance on the catchment scale. In addition, the use of a multi-layer snow model could enhance the simulation of snowmelt by additionally considering heat transfer in the snow pack, especially for snow rich winters (e.g. 2018/2019). However, the multi-layer snow model requires additional calibration parameters to characterize generally unknown soil properties, as well as a new calibration, as a compact snow pack without a layered snow model reacts quite differently to warming or cooling than a layered snow pack (Schulla, 2021).

4.5. Conclusion

This study presented the potential and limitations using Sentinel-2 images for observing snow cover with a high temporal and spatial resolution in a complex mountainous terrain. We also highlighted the additional benefit of using observation data in the process-oriented calibration of a physically-based snow model. To increase the available data, Sentinel-2 images with up to 80% cloud cover were also considered, as they can contain valuable spatial information regarding snow distribution. This requires comprehensive cloud detection to avoid false detection of either snow or no snow. Cloud detection was performed using the Idepix pixel classification algorithm in SNAP for the TUM product. We tested the snow and cloud detection quality against in-situ snow observation on the point scale and also against Eurac, a second snow and cloud product that is based on a supervised algorithm. Both observation products generally attained a very high overall F1 accuracy score (> 0.9) with respect to the in-situ station data. The advantages of using two different approaches are that it makes it possible both to perform an intercomparison and validation of the respective results and to identify the limitations of the two approaches. Moreover, the use of two products enables us to estimate the range of uncertainty in snow mapping with high-resolution optical remote sensing data. Although both observation products provide consistent estimates of the mean snow cover above the tree line (> 2300 m a.s.l.), distinct differences were observed for the slope feature class. The Idepix tool in particular tends to make false detections of snow and cloud on steep north-facing slopes (> 60°) and to falsely classify snow as cloud in some clear sky situations in winter. In contrast, the active learning approach of Eurac underestimates snow cover in evergreen forest areas unless explicit training pixels are selected. The limitation of optical remote sensing products for snow detection under the canopy is a generally well known problem, which is why only the snow cover of high-altitude regions is usually analysed

(Gascoin et al., 2019). Future research will be on assessing fractional snow cover in forested areas using optical remote sensing products (Gascoin et al., 2020).

Unlike the coarse MODIS data (Bouamri et al., 2021), Sentinel-2 enables the investigation of small-scale differences in snow cover duration in complex terrains due to gravitational redistribution (slope), energy balance and wind-driven redistribution (aspect). This makes it possible to calibrate and validate the physically-based snow model of WaSiM in a process-oriented manner on different scales (point, slope and catchment scale) with high spatial resolution (25 m). Depending on the scale and elevation, each snow detection approach has its advantages; for instance, TUM performs better at the forested valley bottom, while Eurac is more reliable at detecting snow on very steep slopes (> 60°). Nevertheless, the two snow detection approaches should be validated with other observational data (e.g. LiDAR) and compared to other algorithms to further investigate the robustness of these approaches.

The spatial comparison of observed and simulated snow cover durations revealed limited WaSiM capabilities on steep slopes (> 25°) with the default parametrization, which cloud be partially overcome by calibrating the gravitational redistribution. Moreover, it was shown that the mean snow cover duration computed with WaSiM at different exposures is not only dependent on the incoming solar radiation but also on the wind-driven snow redistribution. The optimized WaSiM model was able to simulate both the mean snow cover duration with a high F1 accuracy score of > 0.9 and the fractional snow-covered area with a very high correlation coefficient of 0.98. Although hydrological analysis would greatly profit from spatially distributed and highly resolved in time information about SWE, the approach that we propose allows us to constrain two important model parameters for the WaSiM model by considering only snow cover information. Our work therefore provides a robust methodology with which it is possible to collect spatio-temporal snow cover information to calibrate empirical gravitational snow redistribution models, and hence enables multi-objective calibration and validation for hydrological model applications in high-elevation Alpine catchments for further studies. We can hypothesis that calibrated WaSiM model results will also provide more accurate SWE estimation and consequently discharge predictions. Such hypothesis should be tested in a multi-objective optimization framework and under the consideration of further observational data, such as SWE and discharge time series. Moreover, snow cover model results can find application beyond the field of hydrology and are of interest for ecological studies (Qi, Wang, Ma, Zhang, & Yang, 2021; Wan et al., 2014) as well as for the evaluation of the sustainability of winter tourism (Ebner et al., 2021). Although we did not include glaciers in this study, it is possible to derive further cryospheric products from Sentinel-2 (e.g. glacier extents, snow line on glacier and glacier albedo) to calibrate and validate also glacier modules in hydrological models.

5. Coupling support vector machine and physically-based hydrological modeling for reducing the computational time in climate change studies

Hofmeister, Spadina, & Chiogna, 2022[4]

Abstract

Hydrological modeling of Alpine catchments is particularly challenging due to the high variability of hydrological processes in space and time. Although physically-based and fully-distributed hydrological models, such as WaSiM, are able to simulate these small-scale variabilities, the computational time for running a model on hourly time step and 25 m grid resolution in mesoscale catchments (10-100 km²) is significantly high. This becomes particularly relevant when large time periods (> 30 years) are to be simulated for climate change studies. Therefore, we applied Support Vector Regression (SVR) to reproduce the results of a high-resolution WaSiM model (25 m grid, hourly time step) using as an input a coarse spatial (100 m grid) and temporal (daily) resolution of the model and hourly meteorological time series. As a result, the computational time was reduced by 93 % for the model setup with hourly time step and 25 m grid resolution. The quality of the SVR results was quantified through different indicators: Root Mean Squared Error (RMSE), Standard deviation Ratio of RMSE (RSR), Nash-Sutcliffe Efficiency (NSE) and logarithmic NSE (logNSE). Additionally, the SVR results were compared with the flow duration curve. All indicators show an excellent performance (e.g., NSE=0.89) of the SVR in reproducing WaSiM results. We tested the robustness of the SVR also considering different data, such as meteorological inputs from different stations and simulated discharges of sub-catchments. Except for the cases of small sub-catchments with little glacier contribution, very good performance levels were achieved.

[4]Material from:
Hofmeister, F., Spadina, A., & Chiogna, G. (2022). Coupling Support Vector Machine and physically-based hydrological modeling for reducing the computational time in climate change studies. *Proceedings of the 39th IAHR World Congress*, 4827–4836. International Association for Hydro-Environment Engineering and Research (IAHR). doi: 10.3850/IAHR-39WC252171192022902.

5.1. Introduction

The consideration of physical processes at high spatial and temporal resolution in the field of hydrological modeling is still limited by the computational resources, especially when it comes to long simulation runs (> 30 years) on the catchment scale. This is one of the reasons why data-driven machine learning (ML) approaches have been a popular method for runoff modeling in recent years (Mohammadi, 2021). Most of the ML applications (e.g. support vector regression SVR) in hydrological research aim to directly reproduce or predict discharge. Parisouj et al. (2020) employed three different ML algorithms (SVR, Artificial Neural Network with backpropagation (ANN-BP), and Extreme Learning Machine (ELM)) for streamflow predictions at different rivers and concluded that SVR had the best performance among the three models for both daily and monthly scales. Already previous studies mentioned that SVR achieves much better simulation accuracy than ANN in hydrological predictions (Lafdani, Nia, & Ahmadi, 2013; M. Liu & Lu, 2014; Maity, Bhagwat, & Bhatnagar, 2010; Parisouj et al., 2020; Wang, Chau, Cheng, & Qiu, 2009). Liu et al. (2015) coupled a discrete wavelet transform with a SVR model for more reliable daily stream flow predictions. Moreover, SVR was used to forecast reservoir inflows during typhoon periods (Lin, Wang, & Chen, 2016). SVR can also be used to reproduce hydropeaking in response to energy price fluctuations and day of the week, which allows to mimic the management strategy of the reservoir without a specific knowledge of the management rules (Chiogna et al., 2018). Chiang et al. (2022) compared rainfall-runoff model results of a physically-based hydrological model with the results of a SVR approach. The SVR model estimated the rainfall-runoff relationship reasonably well independently of the time series length of the inputs. In contrast, the physically-based model needed more parameter optimization and inference processes to achieve the same performance level as the SVR model (Chiang et al., 2022).

However, to our knowledge, no SVR method has been applied as surrogate model to reproduce the model results of a physically-based hydrological model with high spatial (25 m) and temporal (1 h) resolution to reduce the computational time. As inputs for the SVR, we used the simulation results of the hydrological model WaSiM (Schulla, 2021) with a coarser spatial (100 m) and temporal (daily) resolution and hourly meteorological time series. Compared to the original computational time when running WaSiM with the high spatial resolution of 25 m for a 54 km² large catchment, it was possible to reduce the computational time by 93 % while achieving very good goodness of fit criteria.

5.2. Materials and methods

5.2.1. Research area

The Kaunertal is part of the Oetztaler Alps and located in the central Alps in Tyrol/Austria. The valley has a south-north orientation with an altitudinal range from 1895 m.a.s.l. to 3576 m.a.s.l. and is drained by the Fagge river, a tributary of the Upper Inn River system. The research area of this study covers the upper Kaunertal south of the Gepatsch reservoir closed at the gauge Gepatschalm with an approximated size of 54 km². The runoff regime of the Fagge has a glacio-nival characteristic with a long low flow period in winter (November to May) and a shorter period of larger discharge fluctuations in summer (June to October). Mean discharge of the Fagge at gauge Gepatschalm was 3 m³/s over the period 2009/11 to 2019/12 and maximum hourly discharge was 49.18 m³/s on 22.08.2011. The Gepatsch glacier, the second largest glacier of Austria, covers together with the Weißseeferner approximately 37 % of the catchment (Förster et al., 2016). Besides of multiple hydrological (Förster et al., 2016; Rogger et al., 2017) and cryospheric investigations (Fey et al., 2019; Fleischer et al., 2021; Schattan et al., 2020; Schöber et al., 2014; Wagner et al., 2020, 2016), a number of other processes were studied in the Kaunertal, such as changes in the sediment transport (Baewert & Morche, 2014; Heckmann et al., 2012; Schöber & Hofer, 2018) and in the morphodynamics of this high Alpine area (Altmann et al., 2020; Haas et al., 2012).

The study area with glacier coverage from 2009 and the corresponding WaSiM sub-catchments are shown in Figure 5-1 for two different model discretization 25 m and a coarser of 100 m. Glacier coverage and digital elevation model (DEM) were generated within the PROSA project (High-resolution measurements of morphodynamics in rapidly changing PROglacial Systems of the Alps). The DEM is based on multi-temporal airborne laser scanning data (ALS data), which have been generated during flights in 2012.

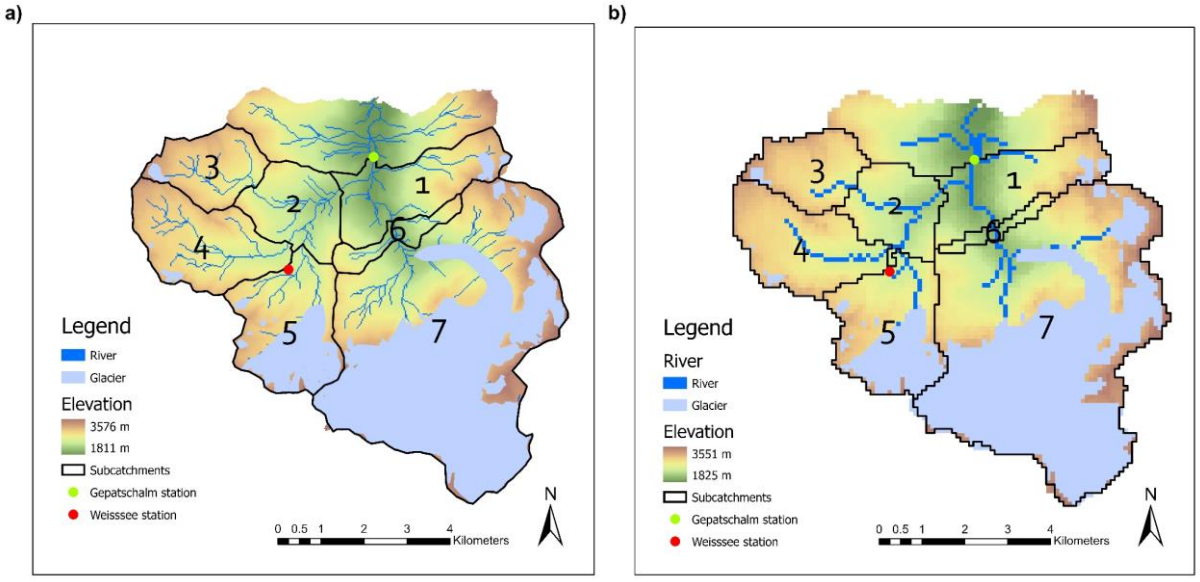


Figure 5-1: The research area upper Kaunertal with different spatial model resolutions 25 m (a) and 100 m (b) and the corresponding sub-catchments implemented in WaSiM. Colored points indicate the meteorological stations Weißsee (red) and Gepatschalm (green), where the stream gauge is also located. The meteorological station Dammfuß is located about 6.4 km north of Gepatschalm and is therefore not visible on the map. All stations are operated by TIWAG.

5.2.2. Support Vector Regression

Before building a Machine Learning model (ML), it is necessary to collect and preprocess the input data, which are afterwards subdivided in training x_i and testing data y_i. Support Vector Regression (SVR) deals with regression problems. The problem consists in finding the best function that maps the m-dimensional input $X \in \boldsymbol{R}^m$ into the output $Y \in \boldsymbol{R}$. The input-output relationship $f(X)$ of SVR can be either linear or non-linear. In this latter case, the training inputs x_i are mapped into a feature space F where the problem has a linear nature. This is done by means of a map function $\Phi(X) \rightarrow F$; in this way, the problem has a linear nature and can be solved as for the linear case (T.-M. Huang, Kecman, & Kopriva, 2006; Smola & Schölkopf, 2004).

The most commonly applied kernel functions are polynomial and Gaussian. We used the latter because of its ability to shorten the computational time and to improve the generalization performance (Dibike, Velickov, Solomatine, & Abbott, 2001; Maity et al., 2010). For solving the SVR, we used two different approaches, which are implemented and available in MATLAB (MATLAB r2021a). The first approach fits an exact SVR model (fitrsvm) by applying a Gaussian kernel function linear or polynomial while the second approach fits a Gaussian kernel regression model (fitrkernel) using random feature expansion, which results in an approximation of the kernel function.

The SVR requires further to define the model hyperparameters. According to Huang et al. (2006), we calibrated the most important parameters ε, C and the kernel scale, i.e. the variance of the Gaussian kernel, which control the smoothness of f(X).

5.2.3. Hydrological model WaSiM

The grid-based deterministic Water Flow and Balance Simulation Model WaSiM (Schulla, 2021) was already applied in various studies for hydrological investigations in Alpine or pre-Alpine catchments (Förster et al., 2018; Kraller et al., 2012; Strasser et al., 2019; Thornton et al., 2021; Warscher et al., 2013). WaSiM has a modular structure and offers several algorithms for the simulation of specific processes depending on the application requirements and data availability. Main model components, which were applied in this study (WaSiM version 10.04.07), are the simulation of the soil water flows in the unsaturated zone based on the process-orientated Richards approach, interpolation of meteorological input with inverse distance weighting (IDW) for precipitation and radiation and with elevation dependent regression for temperature, wind speed and humidity. Moreover, evapotranspiration was simulated with Penman-Monteith, snow and glacier melt with empirical degree-day-factor approach and consideration of gravitational snow redistribution. The WaSiM parameters are mainly based on default values from the WaSiM user manual (Schulla, 2021), while the vegetation and soil parametrization are derived from previous studies within the framework of the PROSA project. We considered a two years warm up period (2009 to 2011) as initialization phase and the hydrological year from October 1 to September 30 for all evaluations.

5.2.4. Data preprocessing

The input data for the SVR are the hourly and daily meteorological time series (temperature, precipitation, wind speed, radiation and humidity) and the daily simulated WaSiM discharge computed using a 100 m grid resolution. The training period of SVR was from 01.10.2011 to 30.09.2016 and testing from 01.10.2016 to 30.09.2018. Since WaSiM interpolates missing data in the meteorological input based on other station data, data gaps had to be interpolated in advanced with the same interpolation approaches to ensure consistency in the input data between WaSiM and SVR.

5.2.5. Input selection

In order to select which parameter should be used to train the SVR, the correlation coefficients between different meteorological inputs from different stations were computed. Considering location, data quality and availability of the time series, we selected meteorological input from the Weißsee station. Besides the daily simulated discharge of WaSiM and the hourly and daily WaSiM inputs, other inputs relevant for hydrological processes such as average temperature and radiation of the day before controlling snowmelt and the precipitation in the mean runoff concentration time of the catchment were considered. In addition, the month number was

included because of the strong seasonality of the hydrological processes in Alpine catchments. Different combinations of these parameters were used to train the support vectors while changing the three hyperparameters, one at a time. Afterwards, the performance of trained SVRs was tested in the validation time and the most promising combination of inputs was selected. The performance of the SVRs was measured with the root mean square error (RMSE) and the results for the hyperparameter kernel scale, which tuned out to be the most sensitive hyperparameter, are listed in Table 5-1 for fitrsvm and Table 5-2 for fitrkernel. As observed value in the RMSE function, we used the hourly discharge generated with the 25 m grid WaSiM.

Table 5-1: RMSE [m³/s] values for the inputs selection for the fitrsvm function in training and validation period. *Qd* is daily discharge from WaSiM in 100 m grid resolution, *P* precipitation, *T* temperature, *conc t* and *d before* mean value of the input in the concentration time and in the day before respectively.

Combination	**Inputs**	**Dammfuss**		**Gepatschalm**		**Weißsee**	
		Train.	**Val.**	**Train.**	**Val.**	**Train.**	**Val.**
1	*Qd, P, Pconc t, T, Tconc t, month*	1.28	1.15	1.20	1.11	1.02	0.99
2	*Qd, P, Pconc t, T, Td before, month*	1.28	1.20	1.22	1.13	1.00	1.01
3	*Qd, P, T, Td before, month*	1.35	1.23	1.31	1.18	1.14	1.08
4	*Qd, P, T, month*	1.39	1.25	1.36	1.23	1.19	1.11
5	*Qd, P, T*	1.33	1.22	1.28	1.16	1.09	1.05
6	*Qd, P, T, Td before*	1.41	1.37	1.37	1.37	1.21	1.26
7	*Qd, T, Td before*	1.41	1.24	1.36	1.19	1.21	1.07
8	*Qd, P, Pconc t*	1.83	1.78	1.82	1.79	1.81	1.76
9	*Qd, rad, radd before*	-	-	-	-	0.56	3.26

Best input combination for fitrsvm is number four with daily discharge, precipitation and temperature measured at the Weißsee station, and month number. The motivation in this choice lies in the small change of RMSE between training and validation period and a low number of inputs, as well as on the results obtained by changing the other two hyperparameters (results not shown). Thus, the RMSE difference between training and validation for the C hyperparameter is greater for the combination five (RMSE 1.08) than for combination four (RMSE 0.43).

Table 5-2: RMSE [m³/s] values for the inputs selection for the fitrkernel function in training and validation period. *Qd* is daily discharge from WaSiM in 100 m grid resolution, *P* precipitation, *T* temperature, *conc t* and *d before* mean value of the input in the concentration time and in the day before respectively.

Combination	Inputs	Dammfuss		Gepatschalm		Weißsee	
		Train.	Val.	Train.	Val.	Train.	Val.
1	*Qd, P, Pconc t, T, Tconc t, month*	1.27	1.14	1.20	1.09	1.03	0.99
2	*Qd, P, Pconc t, T, Td before, month*	1.28	1.17	1.22	1.12	1.00	0.97
3	*Qd, P, T, Td before, month*	1.28	1.19	1.25	1.15	1.06	1.03
4	*Qd, P, T, month*	1.41	1.25	1.39	1.23	1.23	1.10
5	*Qd, P, Pconct, T, Td before, month*	1.38	1.21	1.34	1.14	1.16	1.04
6	*Qd, P, T*	1.46	1.25	1.42	1.22	1.29	1.11
7	*Qd, P, T, Td before*	1.45	1.22	1.40	1.16	1.26	1.06
8	*Qd, T, Td before*	1.47	1.26	1.42	1.19	1.31	1.12
9	*Qd, P, Pconc t*	1.90	1.69	1.90	1.70	1.89	1.69

5.2.6. Training and testing of SVR

We applied the gridsearch and k-fold cross validation as training strategies for the two SVR models in a five year long training period (2011-2016). Gridsearch consists in trying possible combinations of the three hyperparameters, ε, C and kernel scale, and pick the best performing combination (Hsu, Chang, & Lin, 2016) using a grid for the definition of the parameter values. The values of the hyperparameters where chosen in a way that for each of them there was an exponentially growing sequence covering the entire range of the parameters. The best combination was used as an indicator of the "best" region of the grid and on that region a finer grid search was conducted. These steps were repeated until the performance of the SVR converged.

For each combination of hyperparameters, the SVR is trained with k-fold cross validation in order to prevent overfitting (Hsu et al., 2016). In k-fold cross validation, the dataset is used to produce *k* subsets, *k* - 1 subsets are used to train the SVR and the remaining one is used to test its performance. This step is done *k* times, keeping fixed hyperparameters and changing the subset used for testing, in a way that at the end each subset is used once to test the performance. The chosen objective function is RMSE. The dataset of the research area Kaunertal was subdivided in five parts, each corresponding to one year of training period. In order to assess the performance on the entire dataset the *k* RMSE values are averaged. After

several refinements of the grid the optimal hyperparameters were found, these are listed in Table 5-3 for both fitrsvm and fitrkernel.

Table 5-3: Optimal hyperparameters of the trained SVR models.

Hyperparameter	fitrsvm	fitrkernel
ε	$0.0226*10^{0.775}$	$0.2262*10^{-0.01748}$
C	$10^{3.4}$	$10^{3.20508}$
kernal Scale	$10^{1.775}$	$10^{1.41408}$

5.2.7. Baseflow correction

In order to avoid low flow oscillations generated by the two SVR models, we applied a constant baseflow correction. The discharges were substituted with a constant in the low flow periods (winter), which was based on the average of the WaSiM simulated discharge in the low flow periods during the training period and was then substituted to the baseflow computed by SVR machines in both training and validation period.

5.3. Results and discussion

5.3.1. Validation of the SVR models

For validating the SVR models, we computed multiple goodness of fit criteria (GoF) such as the RMSE, its standardized version RSR, Nash-Sutcliffe efficiency (NSE) and the logNSE over the entire period from 2009/11 to 2019/12. In order to compute the GoF, the simulated hourly discharge of WaSiM in 25 m grid resolution is considered as observed value and the hourly discharge reproduced by the SVR as simulated value. The calculated GoF are listed in Table 5-4.

Table 5-4: Performance of the SVR models reproducing hourly discharge by WaSiM in 25 m grid resolution.

GoF	fitrsvm 2009/11-2019/12	fitrkernel 2009/11-2019/12
RMSE [m³/s]	1.23	1.17
RSR [-]	0.34	0.33
NSE [-]	0.88	0.89
logNSE [-]	0.92	0.92

According to the GoF, the SVR model built with the fitrkernel function achieved slightly better model performance than the fitrsvm function. The SVR model based on the fitrkernel function has a better representation of higher discharge values as can be seen for some peak flows in Figure 5-2 for the training period. The difference in predicting high flow events between the two functions is hardly visible, yet present, in the cumulative frequency function (CDF) plots of Figure 5-3.

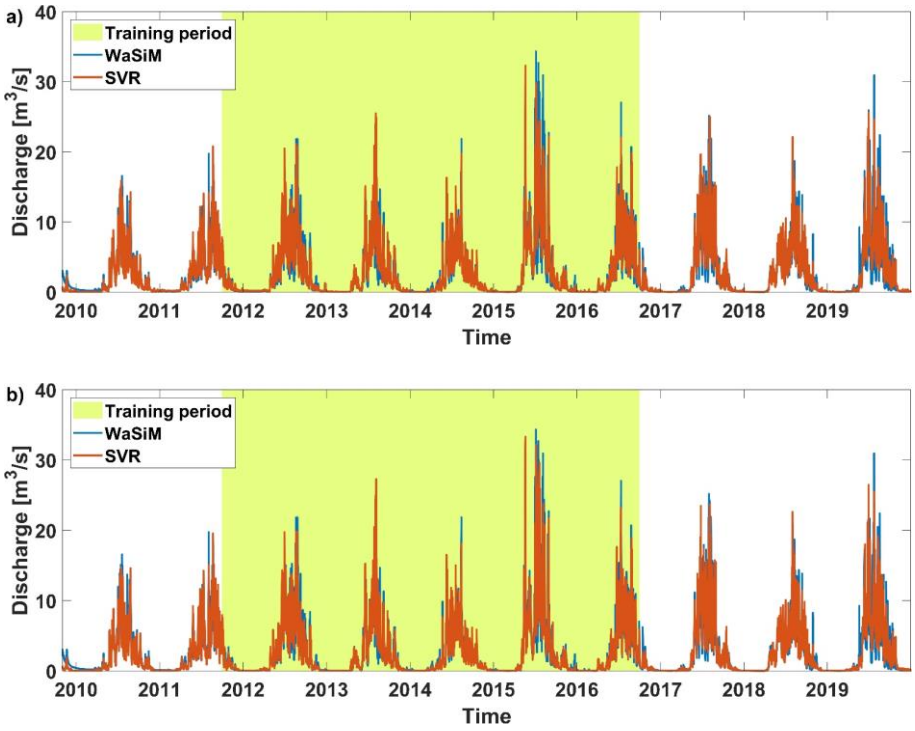


Figure 5-2: Hydrographs of simulated hourly discharge by WaSiM (blue line) and the one reproduced by the SVR models fitrsvm (subplot a) and fitrkernel (subplot b). The yellow area represents the training period of the SVR models.

In the training period, the maximum discharge value simulated by WaSiM is 34.32 m³/s and the maximum values computed by fitrsvm and fitrkernel are 32.31 m³/s and 33.31 m³/s respectively. The reason for both the SVR models underestimating the highest peak might be that only relatively few extremely high discharge values were available to train the SVR machine in the five years of the training period (01.10.2011 to 30.09.2016), in which only three high flow events (05.07.2015, 15.07.2015 and 05.08.2015) with mean hourly discharges larger than 30 m³/s were simulated by WaSiM.

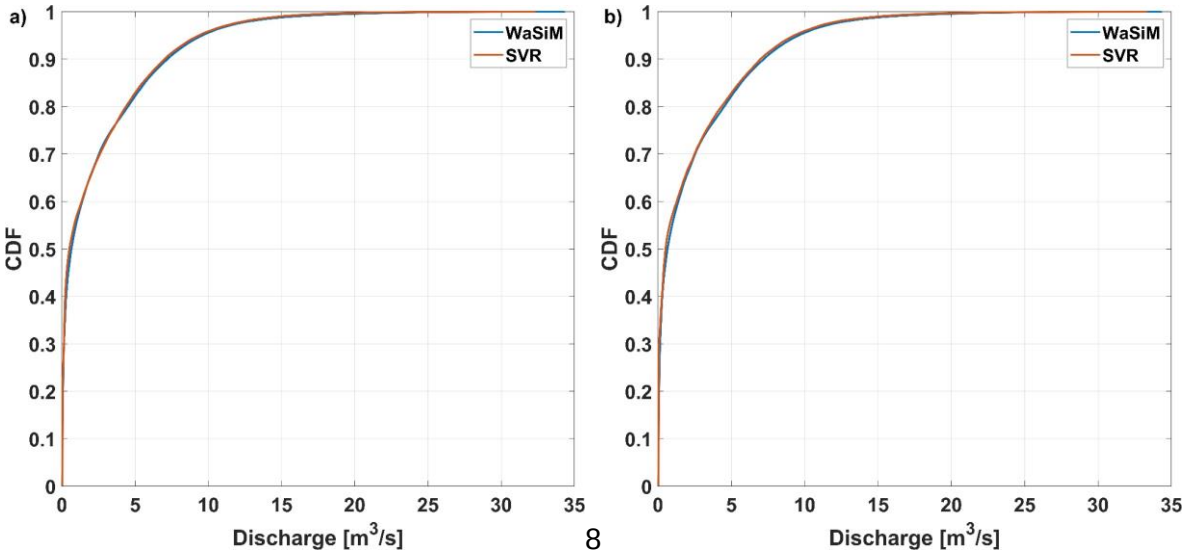


Figure 5-3: Cumulative distribution function (CDF) of hourly discharge by WaSiM and the SVR models with fitrsvm (a) and fitrkernel (b).

The magnitude-squared wavelet coherence is a measure of the correlation between signals in the time-frequency plane and allows the analysis of nonstationary signals. The SVR models are highly coherent with the WaSiM model simulations during the melting period, while the coherence decrease for scales smaller than one week in the winter period (Figure 5-4). However, the wavelet analysis of both the simulated discharge by WaSiM and the reproduced one by the SVR models reviled that none of them presents periodicities at these scales and this observation can explain the lack of coherence observed in Figure 5-4.

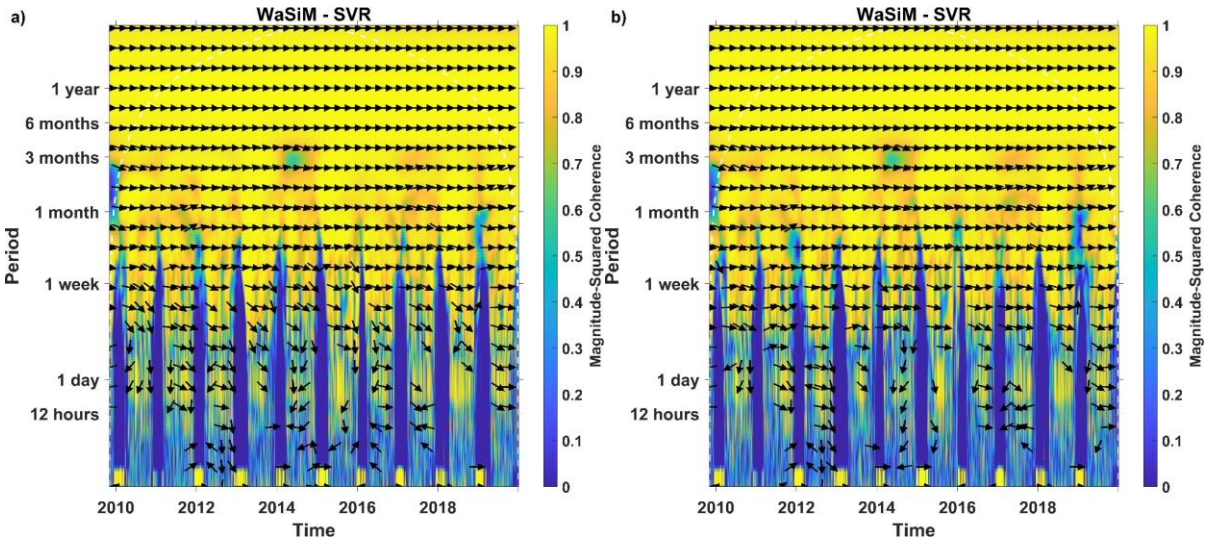


Figure 5-4: Wavelet coherence between the WaSiM simulated discharge and the discharge reproduced by the SVR models fitrsvm (a) and fitrkernel (b) in hourly resolution. The direction of arrows corresponds to the phase lag on the unit circle.

5.3.2. Reduction of computational time

The finer spatial grid resolution of the 25 m WaSiM setup has a significant higher spatial discretization than the 100 m grid version. Consequently, the computational time for running

WaSiM with 25 m spatial and one hour temporal resolution increases by factor of nine in comparison to the WaSiM simulations, which consider daily temporal resolution and 100 m grid. The SVR models aim at reproducing the hourly discharge simulated by WaSiM with 25 m spatial resolution using as input the variables reported in Table 5-1 and Table 5-2. The SVR simulations allows to reproduce the highly resolved WaSiM model with a computational time reduction of 93 % (Table 5-5). We used an ordinary computer (Intel(R) Core(TM) i7-8565U CPU @ 1.80GHz 1.99 GHz) to estimate the computational time listed in Table 5-5.

Table 5-5: Comparison of computational times for one year simulation of hourly discharge and 25 m grid resolution with WaSiM.

WaSiM configuration		**Computational time [s]**	**Time reduction including daily WaSiM with 100 m grid [%]**
25 m grid resolution	WaSiM	17,259	-
	fitrsvm	2.1	-93
	fitrkernel	0.2	-93

5.3.3. Meteorological input from other stations

Although the meteorological observations of the Weißsee station have proven to be the best input for the SVR models (see section 5.2.5), we tested the sensitivity of the SVR model performance to meteorological inputs from other stations. The performance of the SVR trained with temperature and precipitation from the Dammfuß station (1651 m.a.s.l.) resulted in values slightly worse than those produced by Weißsee station, while the SVR trained with temperature and precipitation from the Gepatschalm station (1950 m.a.s.l.) is slightly better than for the Dammfuss. This can be explained by the higher correlation coefficient between observed precipitation at Gepatschalm and Weißsee station (0.81) than at Dammfuß and Weißsee with a correlation coefficient of only 0.61. Nevertheless, this demonstrates the adaptability of the SVR models to different meteorological inputs and the relative low influence on the reproduced discharge.

5.3.4. Discharge of sub-catchments

While we have only evaluated the SVR model results for the outlet of the catchment in the previous sections, we also wanted to test the abilities of the SVR models to reproduce the simulated discharge by WaSiM for sub-catchments (shown in Figure 5-1). For this, we trained the SVR models with the daily discharge of each of the seven sub-catchments computed using the 100 m resolution model as input and gave the hourly 25 m resolution discharge of the corresponding sub-catchment as output. Generally, we observed a better model performance

for larger sub-catchments and in particularly when they contained a significant share of glacier coverage.

5.4. Conclusion

The aim of this study was to reduce the computational time required to obtain hourly discharge values at high spatial resolution based on daily simulated discharge at coarse spatial resolution by WaSiM. Thus, we applied two different Support Vector Regression (SVR) functions, which are able to achieve this goal by reducing the computational time for one year of hourly discharge from 4 hours and 48 minutes (WaSiM at an hourly scale and 25 m grid resolution) to less than 30 minutes with both fitrsvm and fitrkernel (see Table 5-5). This time includes the time needed to run one year of WaSiM simulation at a daily time scale and 100 m grid resolution.

All goodness of fit (GoF) criteria indicate a very good ability of both SVR functions to reproduce hourly discharge by WaSiM (see Table 5-4). However, when comparing the performance of the two functions in more detail, fitrkernel achieved slightly better GoF criteria. Furthermore, the shorter computational time of fitrkernel had a substantial impact on the time required for performing k-fold and gridsearch in the calibration of the hyperparameters. Additionally, the SVR models are not only able to adapt to slightly different meteorological inputs, but also to successfully reproduce the discharge values simulated with the hydrological model with finer spatial grid (25 m).

Difficulties consist in the reproduction of baseflow by the SVR models during the winter period, for which they produced oscillations in the discharge. We have overcome this issue by a constant baseflow for the winter period. Moreover, the SVR models underestimated the highest discharges, probably due to the low representation of peak events in the training period. The low and high flow reproduction can be addressed by training two different support vector regressions one for low flow periods and one for high flow periods. This strategy might lead to a better representation of the different dynamics of the two periods. In addition, the appropriate objective functions have to be selected depending on the season, i.e. logNSE for low flow and NSE, RSR and RSME for summer.

Based on the very good abilities of SVR models to reproduce hourly discharge, further research has to show whether SVR models are also able to reproduce hourly values of other hydrological processes such as snow melt or soil moisture.

6. Conclusion

Although, the topics of the research articles are varied, they are linked and address important research questions that help to test the overall research hypothesis of this dissertation, which is that *a holistic approach of high-resolution spatio-temporal data, comprehensive quality assessment of observation data, and machine learning constitute an advantage for reliable and efficient process-based hydrological modeling in high-elevation Alpine catchments.* In the studies of this dissertation we showed, on one hand, the importance of quality control of observation time series and, on the other, the benefits of high-resolution snow cover maps for obtaining reliable model results. In order to increase the model reliably, a process-based multi-objective optimization is highly recommended for high-elevation catchments. Besides the snow cover information, high-resolution and event-based streamflow observations are essential in a multi-objective optimization framework. Innovative and automated measurement systems can provide streamflow observations with low measurement uncertainty (< 7 %) in turbulent creeks of high-elevation catchments. Consequently, both datasets of collected observational data (i.e., event-based streamflow and snow cover maps) address one of the twenty-three unsolved questions in hydrology (Blöschl et al., 2019), which is: *How can we use innovative technologies to measure surface and subsurface properties, states and fluxes at a range of spatial and temporal scales?* To overcome the limitation of physically-based distributed hydrological models, regarding high computational requirements, we used machine learning as a surrogate model to speed up simulation times significantly while achieving high reliability of the predicted results (NSE=0.89). Thus, we can conclude that model reliability and efficiency are not contradictory and can be increased by different methodical approaches. In the following section, first, a summary and conclusion of the key findings is given followed by an outlook to future research.

6.1. Summary

Hydrometeorological observation data is still the most important data basis for the investigation of the hydrological system. However, especially for longer time series that cover years or even decades, inhomogeneities can occur for different reasons (e.g., change of the location or measurement instrument) (McMillan, Westerberg, & Krueger, 2018). In the first article of this dissertation, we analyzed the quality of hydrometeorological time series from two high-elevation Alpine catchments and investigated how inhomogeneity can affect the results of a hydrological model. As it turned out, around 50 % of the time series are suspicious and 27 % corrupted. Only 20 % of the time series were found to be reliable. Although, we removed the periodicity as a main component of deterministic autocorrelation, short- and long-autocorrelation can still influence the statistical test result of rank correlation methods, as it is

the case for the often applied Mann-Kendall trend test. Another requirement for parametric homogeneity tests is that the time series is normally distributed, which is not the case for all studied hydrometeorological variables. With the two case studies of an inhomogeneous precipitation and temperature time series, we were able to show that the inhomogeneities propagate directly to main components of the cryosphere (i.e., snow and glacier accumulation and melt) but only indirectly to the runoff generation when using them as model forcing, since the runoff generation includes superposition effects from the nonlinear hydrological system. Thus, the answer to the first research question – *What is the impact of inhomogeneous time series on hydrological model results?* – is that the impact of inhomogeneous time series differs highly between the hydrological components. Hence, the quality of the time series needs to be checked before the modeling to reduce the uncertainty in the model results. Based on the discussion in article one, it is beneficial to consider both absolute and relative homogeneity tests in the quality control of the hydrometeorological time series whenever possible.

Continuous and high-quality discharge time series are essential for water management, design of infrastructure and research. In high-elevation areas, the operation of stream gauges and construction of reliable rating curves is particularly challenging due to difficult measuring conditions (e.g., natural cross-section, sediment and bedrock transport, ice jam). Using the AutoSalt system, we were able to collect about 300 measurements in different Alpine catchments with heterogeneous characteristics (e.g., glaciated area, discharge range, background EC_T) within two years. Through an accurate installation and calibration of the system, mean measurement uncertainty was below 7 %. However, we found limitations in the experimental setup, especially regarding the installation of the EC_T probes. In highly turbulent flow sections (i.e., wave action and air bubbles), as it was the case for the river section with the highest streambed gradient (i.e., 18 %), a low SNR resulted in higher measurement uncertainty. Since turbulences in Alpine creeks vary depending on the water level, it is challenging to find sites with a stable EC_T signal, as water level fluctuates also greatly (approx. 1 m to 1.5 m) within measurement periods. Consequently, there is a predicament between turbulent flow, which enhances mixing of the tracer, and the choice of suitable measurement locations for installing the measurement probes with the least possible disturbance (i.e., turbulence and wave action). Redundant measurements of water level and EC_T have been benefitial to reduce aleatory uncertainty caused by aforementioned interference. Furthermore, one of the AutoSalt systems was partly destroyed during an extreme event. This shows the limitation of the system in gauging peak flows when they have a high percentage of debris flow. However, the AutoSalt also allows to quickly re-establish a new stage-discharge relation when the measurement site has been destroyed and the cross-section completely changed. Moreover, the use of the AutoSalt system enables us to detect non-stationarity in the cross-section from year to year or before and after peak events. Although the installation and

operation of the AutoSalt is quite time-consuming, the benefit of the system measuring rare peak events with a high level of accuracy outweighs the costs and efforts. Therefore, the second research question – *how beneficial and reliable are automated stream discharge measurements in high-elevation areas to reduce data scarcity?* – can be answered positively because the collected streamflow data provides valuable information and reduces data scarcity in high-elevation areas.

For process-based model calibration and validation in high-elevation Alpine catchments, snow information is particularly important. Snow observations from stations give valuable information of snow depth at the point scale. However, the station information does not consider the high heterogeneity of snow amount and cover duration in complex terrain, which limits the spatio-temporal validation of the distributed model output. In the third article, we were able to bridge the gap between the point and catchment scale regarding snow cover observations by means of optical remote sensing. Although the idea of snow observations from space is not new, the innovation was in the use of two different algorithms to derive continuous high-resolution snow cover maps in one case study for a period of five years. This allowed us to validate the two products against each other and a couple of snow observation stations. We found main constrains in the detection of snow cover in the canopy and in topographic shadows during low illumination conditions (i.e., winter period). In addition, the observed snow cover maps were used to calibrate and validate the snow cover simulation of a physically-based and distributed snow model. We achieved a high accuracy (i.e., F1 score of > 0.9) with the calibrated snow model, which predicted the fractional snow-covered area well with a correlation coefficient of 0.98. We were thus able to show that, on one hand, Sentinel 2 can observe spatio-temporal variability in snow cover duration due to energy balance dynamics, wind and gravitational redistribution and, on the other, WaSiM can predict the corresponding spatio-temporal variability. Thus, it can be concluded that we can reduce the data scarcity in high-elevation catchments, but there is no single remote sensing product that provides all necessary snow information (e.g., snow extant, SWE, liquid and solid fraction) in high temporal and spatial resolution. Therefore, the answer to the third research question – *can we reduce data scarcity of spatial snow information in high-elevation Alpine catchments by remote sensing data?* – is yes followed by a *but*, as the binary snow cover maps are an important component for reliable small-scale snow observations that includes the limitations mentioned before.

Although the computational capacities constantly increase, the high computational demand of fully distributed and physically-based hydrological models is still one of the main limitations for long simulation runs (> 30 years) with a high spatio-temporal resolution. An increase in computational efficiency of hydrological models can be achieved either by code parallelization

or by reduced order modeling (i.e., ROM). Since the code parallelization of the non-linear hydrological processes with their complex feedback effects is not trivial and high-performance clusters are not always available, reduced order models (i.e., low-dimensional surrogate models) can be used to replace the full system model (Paniconi & Putti, 2015). By coupling WaSiM with a Support Vector Machine, as discussed in article four, we were able to show that the high-resolution WaSiM (i.e., 25 m and hourly time step) results can be predicted with a high degree of reliability (NSE=0.89) and a significant reduction in computational time of 93 % based on a coarser WaSiM model discretization (i.e., 100 m and daily time step). Thus, both parts of the fourth research question – *how fast and reliable are coupled machine learning and physically-based hydrological models?* – can be answered positively. The results are quite promising because one of the main limitations of fully distributed hydrological models could be significantly reduced without a strong decrease in the reliability of the prediction results. Difficulties consist, on one hand, in the runoff prediction of smaller sub-catchments with a low proportion of glaciated area and, on the other, in the prediction of base flow in the winter period. In addition, it is unclear how reliable the predictions are over a longer simulation period (e.g., 30 years). Since the prediction of rare extreme events depends on the selection of the training data extent, the chosen five-year period is rather short in order to train the SVM for long simulation runs (> 30 years). Nevertheless, article four is a good foundation for further investigations and tests in other areas and for longer time periods.

6.2. Outlook

The individual studies address important research questions in the field of Alpine hydrology, but of course are limited in their scope and cannot address all open research questions. Therefore, an outlook for further research is given in the following section. Further research can also arise from further elaboration of the addressed research questions of the individual articles.

A general research question for further studies is the spatio-temporal generalization of the presented methods to other catchments, areas or regions and to other periods and temporal durations. For instance, the comparison of observed and modeled snow cover maps for other high-elevation areas with similar characteristics (e.g., complex topography). In addition, the long-term performance of the simulated snow distribution of WaSiM can be evaluated in order to check if the preferential snow deposition leads to glaciation in areas that are not glaciated. Since the snow accumulation and cover duration on glaciers control glacier melt (i.e., timing and amount), the snow storage is a critical boundary condition for the glacier model initialization. This boundary information can either be obtained from spatial snow observation data or estimated by a sensitivity study to find an optimal initialization time. Not considered in this dissertation, but highly important is the consideration of glacier dynamic in these high-

elevation catchments, especially in Martelltal and Kaunertal. If hydrological models include a glacier routine, it is usually based on simple empirical approaches (e.g., volume-area scaling or delta h) (Tiel et al., 2020). However, it is unclear to what extent these approaches can reproduce the observed glacier dynamics over longer periods (> 30 years). Another question is whether the empirical approaches have to be re-initialized and re-calibrated due to inconsistency of the model parameters and transient boundary conditions after a certain simulation time, which can be constrained by the availability of observation data. Although the primary aim of glacio-hydrological modeling studies is to assess future climate change impacts, long-term changes have rarely been taken into account in model performance evaluations (Tiel et al., 2020). Depending on the complexity of the research question and data availability, the implementation of glacier routines that consider the ice dynamics may be necessary (Pesci & Förster, 2022; Wortmann, Bolch, Su, & Krysanova, 2019).

The availability of global satellite-based products enables a better model calibration and validation. However, main challenges consist in the overview of the available products, their quality, and their limitations. The processing of satellite data is very resource and time consuming, and requires expert knowledge. Therefore, it is advisable to use already available and validated products when possible. For example, the Theia data center provides already processed snow cover maps from high-resolution optical satellites (i.e., Landsat and Sentinel 2) for the European mountains (Gascoin et al., 2019). Leading tech companies in the Big Data field or space agencies not only provide large storage facilities for downloading remote sensing data, but provide also clusters for the direct processing of the data, such as Earth Engine by Google (Google, 2023) and Copernicus Open Access Hub by ESA (ESA, 2023b). Another way of using snow information from remote sensing is their assimilation in distributed snow models, which is gaining more and more popularity because it can be used for more reliable monitoring and predictions of snow storage on a very large scale (e.g., European Alps), as recently developed in the framework of the EO4Alps-Snow project (ESA, 2023a). In the future, integrated solutions with assimilation of both snow and meteorological observations in a coupled atmosphere/snow model systems are likely to provide the most robust estimates of snow conditions, in particular SWE, at scales of practical and scientific interest (Largeron et al., 2020). In order to close the gap between point and areal SWE observations, promising advances in the application of active radar from Sentinel 1 (band X and Ku) have been published by Tsang et al. (2022) recently. However, limitations exist in the complex volume and surface scattering of the radar signal in and on the snow pack, and the complicated feedbacks in the forest. Therefore, ground-based measurements are still the most reliable method to collect snow information, which in turn can be used to validate and improve remote sensing products.

Comprehensive homogeneity testing of hydrometeorological time series is well established in the climate research community, however, it is often not explicitly considered in hydrological modeling studies, in which meteorological records serve as forcing and discharge records as calibration and validation objective (McMillan, Coxon, Sikorska- Senoner, & Westerberg, 2022; McMillan et al., 2018). The focus of model input uncertainty studies is usually on uncertainty in precipitation data, which is the strongest predictor of runoff. Whereas uncertainty analyses of hydrological modeling studies are particularly concerned with the spatiotemporal resolution of precipitation records (Bárdossy, Kilsby, Birkinshaw, Wang, & Anwar, 2022; McMillan et al., 2022) rather than the actual measurement error in the observed data. Beven et al. (2022) recommend that quality assurance of all available data must be carefully performed. Ideally, data should be checked for consistency and errors before a model is run. Since there is a multitude of different statistical tests and methods (e.g., relative and absolute homogeneity tests) for quality control of time series, as already described in the introduction of this dissertation, it is rather difficult to establish a standardized quality control procedure in the field of hydrology. Since 2005, the WMO has regularly published multiple guides for the establishment of national standardized quality management services in the field of hydrometeorology, which have been adopted into the ISO standards (WMO, 2017). Additionally, WMO published a guide for the homogenization of meteorological time series in 2020 (WMO, 2020). At the same time, the scientific development of quality assurance and quality control concepts is progressing continuously. Modified bootstrap approaches (e.g., moving block bootstrap MBB), which generate artificial resamples of a time series, can be used for the quality control of measurement series under the consideration of autocorrelation and deviations from Gaussian distribution (Mudelsee, 2010, 2019). Faybishenko et al. (2022) developed a framework to perform the entire quality assurance and quality control process for a variety of hydrometeorological variables in the R programming environment. The developed statistical framework and methods are suitable for both real-time and post-data-collection quality control analysis. However, this tool does not consider the comparison of neighboring stations for a relative homogeneity check, which is essential for investigating the reason for the inhomogeneity (i.e., anthropogenic factors or climate change). Particularly important in assessing the cause of detected inhomogeneity is the metadata, which should include all changes to the experimental setup such as relocation or change of a sensor. Unfortunately, these are not always available. Therefore, the collection of high-quality observational data that meet the WMO criteria and the documentation of changes in the measurement setup remain the most important data basis for all studies from time series analysis to data driven and hydrological model applications.

High-resolution streamflow observations with a low relative error (< 7 %) are essential for hydrological studies in high-elevation Alpine catchments. Operators of gauging stations of

Alpine creeks faces a variety of challenges, such as continuous or sudden changes in the natural cross-section at the gauge site. Therefore, regular and event-based streamflow measurements are of great importance for assessing the validity of a rating curve and for reducing uncertainty in the extrapolation range of the rating curve. Due to the highly turbulent flow, tracer-based streamflow measurements are still the only method to measure the streamflow reliably. An alternative for reliable streamflow estimations are high-resolution flow field data that describe the presence and interaction of flow velocity and coherent structures. Contactless UAV-based velocity measurement provides a new source of velocity field data for measurements of extreme hydrological events at a safe distance, and could allow for measurements of inaccessible areas (Thumser, Haas, Tuhtan, Fuentes-Pérez, & Toming, 2017). However, the contactless UAV is limited by timing of the site visit and weather condition during an extreme event. In addition, UAV surveys can only take place during daylight hours. Based on these facts, automated and stand-alone measurement systems are superior for the high-resolution recording of hydrological extreme events in high-elevation catchments. Furthermore, it is possible to use the AutoSalt measurements in a real time flood warning and forecasting framework via remote data transmission. This can increase the reliability of early warnings of dangerous runoff events in Alpine headwater catchments with sizes up to 60 km², which is similar to drainage area of the upper Marttelltal. Since the AutoSalt is a rather complex system that requires expert knowledge in the installation, operation, and processing of the measurements, it is necessary to establish knowledge transfer between the system developers and researcher, on the one hand, and the potential users from the public and private sector, on the other.

The emulation of a physically-based model by machine learning enables a fast and efficient sensitivity analysis, parameter optimization, and the estimation of confidence intervals (i.e., uncertainty quantification) (Reichstein et al., 2019). Besides of the applied SVR in article four, long-short term memory (LSTM) networks have great potential for characterizing the input-output conversion relationship of complex nonlinear numerical simulations and can be used as surrogate model of the simulation model. The LSTM method provides reliable predictions especially for large data sets (Li, Lu, & Luo, 2021).

The focus of this dissertation was mainly on the surface processes of Alpine hydrology (from the atmosphere, via snow cover to runoff). However, runoff generation in high-elevation catchments, especially during the low flow period in winter, is not directly fed by liquid precipitation, snow or glacier melt. The storage of liquid precipitation, snow and glacier melt components in the subsurface (e.g., sediments, soils, bedrock) and the delayed runoff contribution is still a major unknown in the research field of Alpine hydrology (Arnoux, Halloran, Berdat, & Hunkeler, 2020; Hayashi, 2020). Due to the rapid change of the geosystem of the

high-elevation areas (Heckmann & Morche, 2019), groundwater dynamics are not stationary and underlay transient boundary conditions (Müller, Lane, & Schaefli, 2022). Due to the high relevance of thís research topic, it was also included in the twenty-three unsolved questions in hydrology (Blöschl et al., 2019) and formulated as followed: *How will cold region runoff and groundwater change in a warmer climate (e.g. with glacier melt and permafrost thaw)?* The role of groundwater storage and dynamics can be investigated and quantified by the collection and analysis of stable isotopes on the catchment scale (Chiogna et al., 2014; Engel et al., 2019; Penna et al., 2017; Zuecco et al., 2018). However, the allocation of storage to specific hydrological units remains unclear (Müller et al., 2022). For this, the installation of piezometers along the river channel and in flat areas allow the observation of the river-groundwater interaction and the seasonal dynamic of the groundwater storage (Kobierska, Jonas, Kirchner, & Bernasconi, 2015; Müller et al., 2022). Another method for characterizing seasonal groundwater storage in alpine catchments is the application of time-lapse gravimetry (Arnoux et al., 2020). Distributed and physically-based hydrogeological models can unravel the surface-groundwater interaction in high-elevation catchments and the importance of glacier and snow melt contribution to groundwater recharge (Müller, Roncoroni, Mancini, Lane, & Schaefli, 2023). The hydrological subproject of the DFG SEHAG 2 research group (Chiogna & Schaefli, 2021) uses a combined approach of field observation (isotope samling and piezometer) and physically-based hydrogeological model to investigate the impact of climate change on groundwater storage in high-elevation Alpine catchments. The main hypothesis to be tested is: Distributed hydrological models can unravel the increased importance of groundwater for high Alpine environments under climate change. The potential changes are investigated in an ensemble approach (i.e., driven by different climate models) up to the year 2050.

Consequently, the holistic approach developed in this dissertation has to be extended to include other key components of the Alpine hydrology such as glacier processes and groundwater storage. However, one should also critically evaluate the relevance and spatiotemporal occurrence of individual processes for answering a given research question before the extension. Depending on this evaluation, the model complexity (i.e., empirical, conceptual or physically-based) and spatiotemporal resolution can be chosen. In turn, the selection of boundary conditions depends on the selected model complexity. In addition, the choice of model complexity should take into account the uncertainty resulting from either the lack of model complexity (i.e., structural uncertainty) or the lack of information to define or constrain the parameters of a model (i.e., parameter uncertainty). Furthermore, the complexity of the research question also defines which interdisciplinary cooperation is necessary and useful to answer the research question. Eventually, the complexity of the research question also depends on the availability and quality of observational data.

Bibliography

Alaoui, A., Willimann, E., Jasper, K., Felder, G., Herger, F., Magnusson, J., & Weingartner, R. (2014). Modelling the effects of land use and climate changes on hydrology in the Ursern Valley, Switzerland: SWISS ALPS; LAND USE CHANGES; CLIMATE CHANGE; TOTAL DISCHARGE. *Hydrological Processes*, *28*(10), 3602–3614. doi: 10.1002/hyp.9895

Allen, C. M., & Taylor, E. A. (1923). The salt velocity method of water measurement. *Trans. Am. Soc. Mech. Eng.*, *45*, 285.

Allen, L., O'Connell, A., & Kiermer, V. (2019). How can we ensure visibility and diversity in research contributions? How the Contributor Role Taxonomy (CRediT) is helping the shift from authorship to contributorship. *Learned Publishing*, *32*(1), 71–74. doi: 10.1002/leap.1210

Alonso-González, E., Aalstad, K., Baba, M. W., Revuelto, J., López-Moreno, J. I., Fiddes, J., … Gascoin, S. (2022). The Multiple Snow Data Assimilation System (MuSA v1.0). *Geoscientific Model Development*, *15*(24), 9127–9155. doi: 10.5194/gmd-15-9127-2022

Altmann, M., Piermattei, L., Haas, F., Heckmann, T., Fleischer, F., Rom, J., … Becht, M. (2020). Long-Term Changes of Morphodynamics on Little Ice Age Lateral Moraines and the Resulting Sediment Transfer into Mountain Streams in the Upper Kauner Valley, Austria. *Water*, *12*(12), 3375. doi: 10.3390/w12123375

Alvera-Azcárate, A., Van der Zande, D., Barth, A., Cardoso dos Santos, J. F., Troupin, C., & Beckers, J.-M. (2021). Detection of shadows in high spatial resolution ocean satellite data using DINEOF. *Remote Sensing of Environment*, *253*, 112229. doi: 10.1016/j.rse.2020.112229

Anagnostou, M. N., Nikolopoulos, E. I., Kalogiros, J., Anagnostou, E. N., Marra, F., Mair, E., … Borga, M. (2018). Advancing Precipitation Estimation and Streamflow Simulations in Complex Terrain with X-Band Dual-Polarization Radar Observations. *Remote Sensing*, *10*(8), 1258. doi: 10.3390/rs10081258

Arnoux, M., Halloran, L. J. S., Berdat, E., & Hunkeler, D. (2020). Characterizing seasonal groundwater storage in alpine catchments using time-lapse gravimetry, water stable isotopes and water balance methods. *Hydrological Processes*, *34*(22), 4319–4333. doi: 10.1002/hyp.13884

Aronica, G., Bates, P. D., & Horritt, M. S. (2002). Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. *Hydrological Processes*, *16*(10), 2001–2016. doi: 10.1002/hyp.398

Auer, I., Böhm, R., Jurkovic, A., Lipa, W., Orlik, A., Potzmann, R., … Nieplova, E. (2007). HISTALP—historical instrumental climatological surface time series of the Greater Alpine Region. *International Journal of Climatology*, *27*(1), 17–46. doi: 10.1002/joc.1377

Auer, I., Böhm, R., Jurković, A., Orlik, A., Potzmann, R., Schöner, W., … Mercalli, L. (2005). A new instrumental precipitation dataset for the greater alpine region for the period 1800–2002. *International Journal of Climatology*, *25*(2), 139–166. doi: 10.1002/joc.1135

Baewert, H., & Morche, D. (2014). Coarse sediment dynamics in a proglacial fluvial system (Fagge River, Tyrol). *Geomorphology*, *218*, 88–97. doi: 10.1016/j.geomorph.2013.10.021

Bahr, D. B., Meier, M. F., & Peckham, S. D. (1997). The physical basis of glacier volume-area scaling. *Journal of Geophysical Research: Solid Earth*, *102*(B9), 20355–20362. doi: 10.1029/97JB01696

Bahr, D. B., Pfeffer, W. T., & Kaser, G. (2015). A review of volume-area scaling of glaciers. *Reviews of Geophysics*, *53*(1), 95–140. doi: 10.1002/2014RG000470

Barandun, M., Fiddes, J., Scherler, M., Mathys, T., Saks, T., Petrakov, D., & Hoelzle, M. (2020). The state and future of the cryosphere in Central Asia. *Water Security*, *11*, 100072. doi: 10.1016/j.wasec.2020.100072

Barandun, M., Pohl, E., Naegeli, K., McNabb, R., Huss, M., Berthier, E., … Hoelzle, M. (2021). Hot Spots of Glacier Mass Balance Variability in Central Asia. *Geophysical Research Letters*, *48*(11). doi: 10.1029/2020GL092084

Bárdossy, A., Kilsby, C., Birkinshaw, S., Wang, N., & Anwar, F. (2022). Is Precipitation Responsible for the Most Hydrological Model Uncertainty? *Frontiers in Water*, *4*, 836554. doi: 10.3389/frwa.2022.836554

Bárdossy, A., & Singh, S. K. (2008). Robust estimation of hydrological model parameters. *Hydrology and Earth System Sciences*, *12*(6), 1273–1283. doi: 10.5194/hess-12-1273-2008

Barrou Dumont, Z., Gascoin, S., Hagolle, O., Ablain, M., Jugier, R., Salgues, G., … Morin, S. (2021). Brief communication: Evaluation of the snow cover detection in the Copernicus High Resolution Snow & Ice Monitoring Service. *The Cryosphere*, *15*(10), 4975–4980. doi: 10.5194/tc-15-4975-2021

Battaglia, F., & Protopapas, M. K. (2012). An analysis of global warming in the Alpine region based on nonlinear nonstationary time series models. *Statistical Methods & Applications*, *21*(3), 315–334. doi: 10.1007/s10260-012-0200-9

Becht, M. (1995). Untersuchungen zur aktuellen Reliefentwicklung in alpinen Einzugsgebieten. *Geobuch-Verlag*, *Münchener Geographische Abhandlungen*.

Beniston, M. (2012). Impacts of climatic change on water and associated economic activities in the Swiss Alps. *Journal of Hydrology*, *412–413*, 291–296. doi: 10.1016/j.jhydrol.2010.06.046

Beniston, M., Farinotti, D., Stoffel, M., Andreassen, L. M., Coppola, E., Eckert, N., … Vincent, C. (2018). The European mountain cryosphere: A review of its current state, trends, and future challenges. *The Cryosphere*, *12*(2), 759–794. doi: 10.5194/tc-12-759-2018

Berger, V. W., & Zhou, Y. (2014). Kolmogorov–Smirnov Test: Overview. In *Wiley StatsRef: Statistics Reference Online*. American Cancer Society. doi: 10.1002/9781118445112.stat06558

Bernhardt, J., Carleton, A. M., & LaMagna, C. (2018). A Comparison of Daily Temperature-Averaging Methods: Spatial Variability and Recent Change for the CONUS. *Journal of Climate*, *31*(3), 979–996. doi: 10.1175/JCLI-D-17-0089.1

Bernier, J. (1994). Statistical Detection of Changes in Geophysical Series. In L. Duckstein & E. Parent (Eds.), *Engineering Risk in Natural Resources Management* (pp. 159–176). Dordrecht: Springer Netherlands. doi: 10.1007/978-94-015-8271-1_9

Beven, K., Lane, S., Page, T., Kretzschmar, A., Hankin, B., Smith, P., & Chappell, N. (2022). On (in)validating environmental models. 2. Implementation of a Turing-like test to modelling hydrological processes. *Hydrological Processes*, *36*(10). doi: 10.1002/hyp.14703

Bittner, D., Richieri, B., & Chiogna, G. (2021). Unraveling the time-dependent relevance of input model uncertainties for a lumped hydrologic model of a pre-alpine karst system. *Hydrogeology Journal*. doi: 10.1007/s10040-021-02377-1

Blöschl, G. (1999). Scaling issues in snow hydrology. *Hydrological Processes*, *13*(14–15), 2149–2175. doi: https://doi.org/10.1002/(SICI)1099-1085(199910)13:14/15<2149::AID-HYP847>3.0.CO;2-8

Blöschl, G., Bierkens, M. F. P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., … Zhang, Y. (2019). Twenty-three unsolved problems in hydrology (UPH) – a community perspective. *Hydrological Sciences Journal*, *64*(10), 1141–1158. doi: 10.1080/02626667.2019.1620507

Borga, M., Degli Esposti, S., & Norbiato, D. (2006). Influence of errors in radar rainfall estimates on hydrological modeling prediction uncertainty: ERRORS IN RADAR RAINFALL ESTIMATES. *Water Resources Research*, *42*(8). doi: 10.1029/2005WR004559

Bormann, H., & Elfert, S. (2010). Application of WaSiM-ETH model to Northern German lowland catchments: Model performance in relation to catchment characteristics and sensitivity to land use change. *Advances in Geosciences*, *27*, 1–10. doi: 10.5194/adgeo-27-1-2010

Bouamri, H., Kinnard, C., Boudhar, A., Gascoin, S., Hanich, L., & Chehbouni, A. (2021). MODIS Does Not Capture the Spatial Heterogeneity of Snow Cover Induced by Solar Radiation. *Frontiers in Earth Science*, *9*, 640250. doi: 10.3389/feart.2021.640250

Brath, A., Montanari, A., & Toth, E. (2004). Analysis of the effects of different scenarios of historical data availability on the calibration of a spatially-distributed hydrological model. *Journal of Hydrology*, *291*(3), 232–253. doi: 10.1016/j.jhydrol.2003.12.044

Brockmann Consult GmbH. (2017). 2. The SNAP Cawa TCWV and CTP Processing System—CAWA Software User Manual. Retrieved June 2, 2021, from https://snap-cawa.readthedocs.io/en/latest/cawa_processing_system.html#the-idepix-pixel-classification-module

Brönnimann, S., Rajczak, J., Fischer, E. M., Raible, C. C., Rohrer, M., & Schär, C. (2018). Changing seasonality of moderate and extreme precipitation events in the Alps. *Natural Hazards and Earth System Sciences*, *18*(7), 2047–2056. doi: 10.5194/nhess-18-2047-2018

Brown, M. E., Racoviteanu, A. E., Tarboton, D. G., Gupta, A. S., Nigro, J., Policelli, F., … Tokar, S. (2014). An integrated modeling system for estimating glacier and snow melt driven streamflow from remote sensing and earth system data products in the Himalayas. *Journal of Hydrology*, *519*, 1859–1869. doi: 10.1016/j.jhydrol.2014.09.050

Brugnara, Y., & Maugeri, M. (2019). Daily precipitation variability in the southern Alps since the late 19th century. *International Journal of Climatology*, *39*(8), 3492–3504. doi: 10.1002/joc.6034

Brunetti, M., Maugeri, M., Nanni, T., Auer, I., Böhm, R., & Schöner, W. (2006). Precipitation variability and changes in the greater Alpine region over the 1800–2003 period. *Journal of Geophysical Research*, *111*(D11), D11107. doi: 10.1029/2005JD006674

Buchmann, M., Coll, J., Aschauer, J., Begert, M., Brönnimann, S., Chimani, B., … Marty, C. (2022). Homogeneity assessment of Swiss snow depth series: Comparison of break detection capabilities of (semi-)automatic homogenization methods. *The Cryosphere*, *16*(6), 2147–2161. doi: 10.5194/tc-16-2147-2022

Buckel, J., & Otto, J.-C. (2018). *The Austrian Glacier Inventory GI 4 (2015) in ArcGis (shapefile) format* (p. 4 MBytes) [Application/zip]. PANGAEA - Data Publisher for Earth & Environmental Science. doi: 10.1594/PANGAEA.887415

Bui, M. T., Lu, J., & Nie, L. (2020). A Review of Hydrological Models Applied in the Permafrost-Dominated Arctic Region. *Geosciences*, *10*(10), 401. doi: 10.3390/geosciences10100401

Cano-Paoli, K., Chiogna, G., & Bellin, A. (2019). Convenient use of electrical conductivity measurements to investigate hydrological processes in Alpine headwaters. *Science of The Total Environment*, *685*, 37–49. doi: 10.1016/j.scitotenv.2019.05.166

Carturan, L., De Blasi, F., Cazorzi, F., Zoccatelli, D., Bonato, P., Borga, M., & Dalla Fontana, G. (2019). Relevance and Scale Dependence of Hydrological Changes in Glacierized Catchments: Insights from Historical Data Series in the Eastern Italian Alps. *Water*, *11*(1), 89. doi: 10.3390/w11010089

Chang, Y., Mewes, B., & Hartmann, A. (2022). *Using LSTM to monitor continuous discharge indirectly with electrical conductivity observations* [Preprint]. Catchment hydrology/Modelling approaches. doi: 10.5194/hess-2022-77

Chen, J., Saunders, K., & Whan, K. (2021). Quality control and bias adjustment of crowdsourced wind speed observations. *Quarterly Journal of the Royal Meteorological Society*, *147*(740), 3647–3664. doi: 10.1002/qj.4146

Chiang, S., Chang, C.-H., & Chen, W.-B. (2022). Comparison of Rainfall-Runoff Simulation between Support Vector Regression and HEC-HMS for a Rural Watershed in Taiwan. *Water*, *14*(2), 191. doi: 10.3390/w14020191

Chiogna, G., Marcolini, G., Liu, W., Pérez Ciria, T., & Tuo, Y. (2018). Coupling hydrological modeling and support vector regression to model hydropeaking in alpine catchments. *Science of The Total Environment*, *633*, 220–229. doi: 10.1016/j.scitotenv.2018.03.162

Chiogna, G., Santoni, E., Camin, F., Tonon, A., Majone, B., Trenti, A., & Bellin, A. (2014). Stable isotope characterization of the Vermigliana catchment. *Journal of Hydrology*, *509*, 295–305. doi: 10.1016/j.jhydrol.2013.11.052

Chiogna, G., & Schaefli, B. (2021). *SEHAG Subproject 2—Impact of climate change on groundwater storage in high Alpine catchments: From observation to model predictions*. not published.

Cimpianu, C. (2018). Delineate snow using Normalized Difference Snow Index, Sentinel 2 and QGIS. Retrieved from GISCourse.com website: https://www.giscourse.com/delineate-snow-using-normalized-difference-snow-index-sentinel-2-and-qgis/

Clow, D. W., & Fleming, A. C. (2008). Tracer gauge: An automated dye dilution gauging system for ice-affected streams: AN AUTOMATED DYE DILUTION GAUGING SYSTEM. *Water Resources Research*, *44*(12), 11. doi: 10.1029/2008WR007090

Comiti, F., Mao, L., Penna, D., Dell'Agnese, A., Engel, M., Rathburn, S., & Cavalli, M. (2019). Glacier melt runoff controls bedload transport in Alpine catchments. *Earth and Planetary Science Letters*, *520*, 77–86. doi: 10.1016/j.epsl.2019.05.031

Cox, D. R., & Stuart, A. (1955). SOME QUICK SIGN TESTS FOR TREND IN LOCATION AND DISPERSION. *Biometrika*, *42*(1–2), 80–95. doi: 10.1093/biomet/42.1-2.80

Crespi, A., Matiu, M., Bertoldi, G., Petitta, M., & Zebisch, M. (2021). A high-resolution gridded dataset of daily temperature and precipitation records (1980–2018) for Trentino-South Tyrol (north-eastern Italian Alps). *Earth System Science Data*, *13*(6), 2801–2818. doi: 10.5194/essd-13-2801-2021

Crochemore, L., Isberg, K., Pimentel, R., Pineda, L., Hasan, A., & Arheimer, B. (2020). Lessons learnt from checking the quality of openly accessible river flow data worldwide. *Hydrological Sciences Journal*, *65*(5), 699–711. doi: 10.1080/02626667.2019.1659509

Cudennec, C., Lins, H., Uhlenbrook, S., & Arheimer, B. (2020). Editorial – Towards FAIR and SQUARE hydrological data. *Hydrological Sciences Journal*, *65*(5), 681–682. doi: 10.1080/02626667.2020.1739397

Debolskiy, M. V., Alexeev, V. A., Hock, R., Lammers, R. B., Shiklomanov, A., Schulla, J., … Prusevich, A. (2021). Water balance response of permafrost-affected watersheds to changes in air temperatures. *Environmental Research Letters*, *16*(8), 084054. doi: 10.1088/1748-9326/ac12f3

Dembélé, M., Oriani, F., Tumbulto, J., Mariéthoz, G., & Schaefli, B. (2019). Gap-filling of daily streamflow time series using Direct Sampling in various hydroclimatic settings. *Journal of Hydrology*, *569*, 573–586. doi: 10.1016/j.jhydrol.2018.11.076

Deschamps-Berger, C., Cluzet, B., Dumont, M., Lafaysse, M., Berthier, E., Fanise, P., & Gascoin, S. (2022). Improving the Spatial Distribution of Snow Cover Simulations by Assimilation of Satellite Stereoscopic Imagery. *Water Resources Research*, *58*(3). doi: 10.1029/2021WR030271

Di Baldassarre, G., & Montanari, A. (2009a). Uncertainty in river discharge observations: A quantitative analysis. *Hydrol. Earth Syst. Sci.*, 9.

Di Baldassarre, G., & Montanari, A. (2009b). Uncertainty in river discharge observations: A quantitative analysis. *Hydrology and Earth System Sciences*, *13*(6), 913–921. doi: 10.5194/hess-13-913-2009

Di Marco, N., Righetti, M., Avesani, D., Zaramella, M., Notarnicola, C., & Borga, M. (2020a). Comparison of MODIS and Model-Derived Snow-Covered Areas: Impact of Land Use and Solar Illumination Conditions. *Geosciences*, *10*(4), 134. doi: 10.3390/geosciences10040134

Di Marco, N., Righetti, M., Avesani, D., Zaramella, M., Notarnicola, C., & Borga, M. (2020b). Comparison of MODIS and Model-Derived Snow-Covered Areas: Impact of Land Use and Solar Illumination Conditions. *Geosciences*, *10*(4), 134. doi: 10.3390/geosciences10040134

Dibike, Y. B., Velickov, S., Solomatine, D., & Abbott, M. B. (2001). Model Induction with Support Vector Machines: Introduction and Applications. *Journal of Computing in Civil Engineering*, *15*(3), 208–216. doi: 10.1061/(ASCE)0887-3801(2001)15:3(208)

Dietz, A. J., Wohner, C., & Kuenzer, C. (2012). European Snow Cover Characteristics between 2000 and 2011 Derived from Improved MODIS Daily Snow Cover Products. *Remote Sensing*, *4*(8), 2432–2454. doi: 10.3390/rs4082432

Dozier, J. (1989). Spectral signature of alpine snow cover from the landsat thematic mapper. *Remote Sensing of Environment*, *28*, 9–22. doi: 10.1016/0034-4257(89)90101-6

Dozier, J., & Painter, T. H. (2004). MULTISPECTRAL AND HYPERSPECTRAL REMOTE SENSING OF ALPINE SNOW PROPERTIES. *Annual Review of Earth and Planetary Sciences*, *32*(1), 465–494. doi: 10.1146/annurev.earth.32.101802.120404

Duethmann, D., Peters, J., Blume, T., Vorogushyn, S., & Güntner, A. (2014). The value of satellite-derived snow cover images for calibrating a hydrological model in snow-dominated catchments in Central Asia. *Water Resources Research*, *50*(3), 2002–2021. doi: 10.1002/2013WR014382

Durand, Y., Guyomarc'h, G., Mérindol, L., & Corripio, J. G. (2005). Improvement of a numerical snow drift model and field validation. *Cold Regions Science and Technology*, *43*(1–2), 93–103. doi: 10.1016/j.coldregions.2005.05.008

Dyck, S. (1980). *Angewandte Hydrologie, Teil 1: Berechnung und Regelung des Durchflusses der Flüsse* (2., völlig überarb. Aufl). Berlin ; München ; Düsseldorf: Ernst.

Ebner, P. P., Koch, F., Premier, V., Marin, C., Hanzer, F., Carmagnola, C. M., … Lehning, M. (2021). Evaluating a prediction system for snow management. *The Cryosphere*, *15*(8), 3949–3973. doi: 10.5194/tc-15-3949-2021

El-Shaarawi, A. H., & Damsleth, E. (1988). PARAMETRIC AND NONPARAMETRIC TESTS FOR DEPENDENT DATA. *Journal of the American Water Resources Association*, *24*(3), 513–519. doi: 10.1111/j.1752-1688.1988.tb00901.x

Engel, M., Penna, D., Bertoldi, G., Vignoli, G., Tirler, W., & Comiti, F. (2019). Controls on spatial and temporal variability in streamflow and hydrochemistry in a glacierized catchment. *Hydrology and Earth System Sciences*, *23*(4), 2041–2063. doi: 10.5194/hess-23-2041-2019

ESA. (2023a). Eo4alps snow project. Retrieved January 4, 2023, from Earth Observation for operational hydrology 'eo4alps snow'—Alps regional initiative—Eo science for society website: https://eo4society.esa.int/projects/eo4alps-snow/

ESA. (2023b). Open Access Hub. Retrieved March 4, 2023, from Copernicus Open Access Hub website: https://scihub.copernicus.eu/

Essery, R., Li, L., & Pomeroy, J. (1999). A distributed model of blowing snow over complex terrain. *Hydrological Processes*, *13*(14–15), 2423–2438. doi: 10.1002/(SICI)1099-1085(199910)13:14/15<2423::AID-HYP853>3.0.CO;2-U

European Environment Agency. (2009). *Regional climate change and adaptation: The Alps facing the challenge of changing water resources*. LU: Publications Office. Retrieved from https://data.europa.eu/doi/10.2800/12552

European Space Agency. (2019). *Sentinel Application Platform (SNAP)*.

European Space Agency. (2021). Sentinel-2—Missions—Sentinel Online—Sentinel Online. Retrieved July 20, 2021, from https://sentinel.esa.int/web/sentinel/missions/sentinel-2

Farsi, N., Mahjouri, N., & Ghasemi, H. (2020). Breakpoint detection in non-stationary runoff time series under uncertainty. *Journal of Hydrology*, *590*, 125458. doi: 10.1016/j.jhydrol.2020.125458

Fassnacht, S. R., López-Moreno, J. I., Ma, C., Weber, A. N., Pfohl, A. K. D., Kampf, S. K., & Kappas, M. (2017). Spatio-temporal snowmelt variability across the headwaters of the Southern Rocky Mountains. *Frontiers of Earth Science*, *11*(3), 505–514. doi: 10.1007/s11707-017-0641-4

Fathom Scientific Ltd. (2020a). *AutoSalt: AQAc M5—Operating Manual*. Retrieved from https://www.fathomscientific.com/wp-content/uploads/2017/09/AQAc_Manual_V1.1.pdf

Fathom Scientific Ltd. (2020b). *THRECS - Operating Manual*. Retrieved from https://www.fathomscientific.com/wp-content/uploads/2020/12/T-HRECS_M7_Manual_v0.6_web.pdf

Faybishenko, B., Versteeg, R., Pastorello, G., Dwivedi, D., Varadharajan, C., & Agarwal, D. (2022). Challenging problems of quality assurance and quality control (QA/QC) of meteorological time series data. *Stochastic Environmental Research and Risk Assessment*, *36*(4), 1049–1062. doi: 10.1007/s00477-021-02106-w

Faybishenko, Boris, Versteeg, R., Pastorello, G., Dwivedi, D., Varadharajan, C., & Agarwal, D. (2021, August 5). *Challenging Problems of Quality Assurance and Quality Control (QA/QC) of Meteorological Time Series Data*. doi: 10.21203/rs.3.rs-510046/v1

Fey, C., Schattan, P., Helfricht, K., & Schöber, J. (2019). A compilation of multitemporal TLS snow depth distribution maps at the Weisssee snow research site (Kaunertal, Austria). *Water Resources Research*, 2019WR024788. doi: 10.1029/2019WR024788

Fiebrich, C. A., Morgan, C. R., McCombs, A. G., Hall, P. K., & McPherson, R. A. (2010). Quality Assurance Procedures for Mesoscale Meteorological Data. *Journal of Atmospheric and Oceanic Technology*, *27*(10), 1565–1582. doi: 10.1175/2010JTECHA1433.1

Fleischer, F., Haas, F., Piermattei, L., Pfeiffer, M., Heckmann, T., Altmann, M., … Becht, M. (2021). Multi-decadal (1953–2017) rock glacier kinematics analysed by high-resolution topographic data in the upper Kaunertal, Austria. *The Cryosphere*, *15*(12), 5345–5369. doi: 10.5194/tc-15-5345-2021

Foehn, A., García Hernández, J., Schaefli, B., & De Cesare, G. (2018). Spatial interpolation of precipitation from multiple rain gauge networks and weather radar data for operational applications in Alpine catchments. *Journal of Hydrology*, *563*, 1092–1110. doi: 10.1016/j.jhydrol.2018.05.027

Foppa, N., Stoffel, A., & Meister, R. (2005). *Snow depth mapping in the Alps: Merging of in situ and remotely-sensed data*. *4*(1). Retrieved from https://www.dora.lib4ri.ch/wsl/islandora/object/wsl:12578

Förster, K., Garvelmann, J., Meißl, G., & Strasser, U. (2018). Modelling forest snow processes with a new version of WaSiM. *Hydrological Sciences Journal*, *63*(10), 1540–1557. doi: 10.1080/02626667.2018.1518626

Förster, K., Oesterle, F., Hanzer, F., Schöber, J., Huttenlau, M., & Strasser, U. (2016). A snow and ice melt seasonal prediction modelling system for Alpine reservoirs. *Proceedings of the International Association of Hydrological Sciences*, *374*, 143–150. doi: 10.5194/piahs-374-143-2016

Frei, A., Tedesco, M., Lee, S., Foster, J., Hall, D. K., Kelly, R., & Robinson, D. A. (2012). A review of global satellite-derived snow products. *Advances in Space Research*, *50*(8), 1007–1029. doi: 10.1016/j.asr.2011.12.021

Freudiger, D., Kohn, I., Seibert, J., Stahl, K., & Weiler, M. (2017). Snow redistribution for the hydrological modeling of alpine catchments: Snow redistribution for hydrological modeling. *Wiley Interdisciplinary Reviews: Water*, *4*(5), e1232. doi: 10.1002/wat2.1232

Frey, S., & Holzmann, H. (2015). A conceptual, distributed snow redistribution model. *Hydrology and Earth System Sciences*, *19*(11), 4517–4530. doi: 10.5194/hess-19-4517-2015

Galos, S. P., Klug, C., Maussion, F., Covi, F., Nicholson, L., Rieg, L., … Kaser, G. (2017). Reanalysis of a 10-year record (2004–2013) of seasonal mass balances at Langenferner/Vedretta Lunga, Ortler Alps, Italy. *The Cryosphere*, *11*(3), 1417–1439. doi: 10.5194/tc-11-1417-2017

Galos, S. P., & Klug, R. (2015). RECENT GLACIER CHANGES AND RELATED CONTRIBUTION POTENTIAL TO RIVER DISCHARGE IN THE VINSCHGAU / VAL VENOSTA, ITALIAN ALPS. *Geografia Fisica e Dinamica Quaternaria*, (38.2), 143–154. doi: 10.4461/GFDQ.2015.38.13

Gascoin, S., Barrou Dumont, Z., Deschamps-Berger, C., Marti, F., Salgues, G., López-Moreno, J. I., … Hagolle, O. (2020). Estimating Fractional Snow Cover in Open Terrain from Sentinel-2 Using the Normalized Difference Snow Index. *Remote Sensing*, *12*(18), 2904. doi: 10.3390/rs12182904

Gascoin, S., Grizonnet, M., Bouchet, M., Salgues, G., & Hagolle, O. (2019). Theia Snow collection: High-resolution operational snow cover maps from Sentinel-2 and Landsat-8 data. *Earth System Science Data*, *11*(2), 493–514. doi: 10.5194/essd-11-493-2019

Gauer, P. (1998). Blowing and drifting snow in Alpine terrain: Numerical simulation and related field measurements. *Annals of Glaciology*, *26*, 174–178. doi: 10.3189/1998AoG26-1-174-178

Geitner, C., & Becht, M. (1999). Sedimentologische und vegetationsgeschichtliche Untersuchungen an fluvialen Sedimenten in den Hochlagen des Horlachtales (Stubaier Alpen/Tirol). *Geobuch-Verlag*, *Münchener Geographische Abhandlungen*. Retrieved from https://www.uibk.ac.at/geographie/igg/berichte/1999/pdf/geitner.pdf

Girona-Mata, M., Miles, E. S., Ragettli, S., & Pellicciotti, F. (2019). High-Resolution Snowline Delineation From Landsat Imagery to Infer Snow Cover Controls in a Himalayan Catchment. *Water Resources Research*, *55*(8), 6754–6772. doi: 10.1029/2019WR024935

Gobiet, A., Kotlarski, S., Beniston, M., Heinrich, G., Rajczak, J., & Stoffel, M. (2014). 21st century climate change in the European Alps—A review. *Science of The Total Environment*, *493*, 1138–1151. doi: 10.1016/j.scitotenv.2013.07.050

González-Pinzón, R., Haggerty, R., & Dentz, M. (2013). Scaling and predicting solute transport processes in streams: Scaling Solute Transport in Streams. *Water Resources Research*, *49*(7), 4071–4088. doi: 10.1002/wrcr.20280

Google. (2023). Google Earth Engine. Retrieved March 4, 2023, from Meet Earth Engine website: https://earthengine.google.com

Gottardi, G., Maglionico, M., & Bolognesi, A. (2006). Discharge measurements in a small ungauged river: Comparison between conventional current-meter and tracer dilution methods. In E. Alves, A. Cardoso, J. Leal, & R. Ferreira (Eds.), *River Flow 2006* (pp. 1835–1842). Taylor & Francis. doi: 10.1201/9781439833865.ch200

Gruber, S. (2007). A mass-conserving fast algorithm to parameterize gravitational transport and deposition using digital elevation models: MASS-CONSERVING TRANSPORT AND DEPOSITION ON DEMS. *Water Resources Research*, *43*(6). doi: 10.1029/2006WR004868

Grünewald, T., Bühler, Y., & Lehning, M. (2014). Elevation dependency of mountain snow depth. *The Cryosphere*, *8*(6), 2381–2394. doi: 10.5194/tc-8-2381-2014

Grünewald, Thomas, & Lehning, M. (2011). Altitudinal dependency of snow amounts in two small alpine catchments: Can catchment-wide snow amounts be estimated via single snow or precipitation stations? *Annals of Glaciology*, *52*(58), 153–158. doi: 10.3189/172756411797252248

Grünewald, Thomas, & Wolfsperger, F. (2019). Water Losses During Technical Snow Production: Results From Field Experiments. *Frontiers in Earth Science*, *7*, 78. doi: 10.3389/feart.2019.00078

Gudmundsson, L., Do, H. X., Leonard, M., & Westra, S. (2018). The Global Streamflow Indices and Metadata Archive (GSIM) – Part 2: Quality control, time-series indices and homogeneity assessment. *Earth System Science Data*, *10*(2), 787–804. doi: 10.5194/essd-10-787-2018

Günther, D., Marke, T., Essery, R., & Strasser, U. (2019). Uncertainties in Snowpack Simulations—Assessing the Impact of Model Structure, Parameter Choice, and Forcing Data Error on Point-Scale Energy Balance Snow Model Performance. *Water Resources Research*, *55*(4), 2779–2800. doi: 10.1029/2018WR023403

Gurung, D. R., Maharjan, S. B., Shrestha, A. B., Shrestha, M. S., Bajracharya, S. R., & Murthy, M. S. R. (2017). Climate and topographic controls on snow cover dynamics in the Hindu Kush Himalaya. *International Journal of Climatology*, *37*(10), 3873–3882. doi: 10.1002/joc.4961

Haas, F., Heckmann, T., Hilger, L., & Becht, M. (2012). Quantification and Modelling of Debris Flows in the Proglacial Area of the Gepatschferner/Austria using Ground-based LIDAR. *Hydrological Sciences Journal*, (356), 293–302.

Hall, D. K., Foster, J. L., DiGirolamo, N. E., & Riggs, G. A. (2012). Snow cover, snowmelt timing and stream power in the Wind River Range, Wyoming. *Geomorphology*, *137*(1), 87–93. doi: 10.1016/j.geomorph.2010.11.011

Hamed, K. H. (2008). Trend detection in hydrologic data: The Mann–Kendall trend test under the scaling hypothesis. *Journal of Hydrology*, *349*(3–4), 350–363. doi: 10.1016/j.jhydrol.2007.11.009

Hamed, K. H., & Rao, A. R. (1998). A modified Mann-Kendall trend test for autocorrelated data. *Journal of Hydrology*, *204*(1–4), 182–196. doi: 10.1016/S0022-1694(97)00125-X

Hamon, W. R. (1961). Estimating Potential Evapotranspiration. *Journal of the Hydraulics Division*, *87*(3), 107–120. doi: 10.1061/JYCEAJ.0000599

Hänsel, S., Medeiros, D. M., Matschullat, J., Petta, R. A., & de Mendonça Silva, I. (2016). Assessing Homogeneity and Climate Variability of Temperature and Precipitation Series in the Capitals of North-Eastern Brazil. *Frontiers in Earth Science*, *4*. doi: 10.3389/feart.2016.00029

Hanzer, F., Helfricht, K., Marke, T., & Strasser, U. (2016). Multi-level spatiotemporal validation of snow/ice mass balance and runoff modeling in glacierized catchments. *The Cryosphere Discussions*, 1–37. doi: 10.5194/tc-2016-58

Hauet, A. C. (2020). Uncertainty of salt discharge measurement: The SUNY Framework. *Norwegian Water and Energy Directorate*, *29/2020*, 23.

Hauet, A. C., Morlot, T., & Daubagnan, L. (2018). Velocity profile and depth-averaged to surface velocity in natural streams: A review over alarge sample of rivers. *E3S Web of Conferences*, *40*, 06015. doi: 10.1051/e3sconf/20184006015

Hayashi, M. (2020). Alpine Hydrogeology: The Critical Role of Groundwater in Sourcing the Headwaters of the World. *Groundwater*, *58*(4), 498–510. doi: 10.1111/gwat.12965

Heckmann, T., Haas, F., Morche, D., Schmidt, K.-H., Rohn, J., Moser, M., … Becht, M. (2012). Investigating an Alpine proglacial sediment budget using field measurements, airborne and terrestrial LiDAR data. *IAHS Publ.*, 11.

Heckmann, T., & Morche, D. (Eds.). (2019). *Geomorphology of Proglacial Systems: Landform and Sediment Dynamics in Recently Deglaciated Alpine Landscapes*. Cham: Springer International Publishing. doi: 10.1007/978-3-319-94184-4

Hingray, B., Schaefli, B., Mezghani, A., & Hamdi, Y. (2010). Signature-based model calibration for hydrological prediction in mesoscale Alpine catchments. *Hydrological Sciences Journal*, *55*(6), 1002–1016. doi: 10.1080/02626667.2010.505572

Hirsch, R. M., Slack, J. R., & Smith, R. A. (1982). Techniques of trend analysis for monthly water quality data. *Water Resources Research*, *18*(1), 107–121. doi: 10.1029/WR018i001p00107

Hock, R. (1999). A distributed temperature-index ice- and snowmelt model including potential direct solar radiation. *Journal of Glaciology*, *45*(149), 101–111. doi: 10.3189/S0022143000003087

Hofmeister, F., Arias-Rodriguez, L. F., Premier, V., Marin, C., Notarnicola, C., Disse, M., & Chiogna, G. (2022). Intercomparison of Sentinel-2 and modelled snow cover maps in a high-elevation Alpine catchment. *Journal of Hydrology X*, *15*, 100123. doi: 10.1016/j.hydroa.2022.100123

Hofmeister, F., Graziano, F., Marcolini, G., Willems, W., Disse, M., & Chiogna, G. (2023). Quality assessment of hydrometeorological observational data and their influence on hydrological model results in Alpine catchments. *Hydrological Sciences Journal*, 02626667.2023.2172335. doi: 10.1080/02626667.2023.2172335

Hofmeister, F., Notarnicola, C., Chiogna, G., Disse, M., Arias-Rodriguez, L. F., Premier, V., & Marin, C. (2022). *Intercomparison of Sentinel-2 and modelled snow cover maps in a high-elevation Alpine catchment* [Data set]. Mendeley. doi: 10.17632/FRGXPBHZR8.1

Hofmeister, F., Rubens Venegas, B., Sentlinger, G., Tarantik, M., Blume, T., Disse, M., & Chiogna, G. (2023). Automated streamflow measurements in high-elevation Alpine catchments. *River Research and Applications*, rra.4203. doi: 10.1002/rra.4203

Hofmeister, F., Spadina, A., & Chiogna, G. (2022). Coupling Support Vector Machine and physically-based hydrological modeling for reducing the computational time in climate change studies. *Proceedings of the 39th IAHR World Congress*, 4827–4836. International Association for Hydro-Environment Engineering and Research (IAHR). doi: 10.3850/IAHR-39WC252171192022902

Horner, I., Le Coz, J., Renard, B., Branger, F., & Lagouy, M. (2022). Streamflow uncertainty due to the limited sensitivity of controls at hydrometric stations. *Hydrological Processes*, *36*(2). doi: 10.1002/hyp.14497

Hounkpè, J., Diekkrüger, B., Afouda, A. A., & Sintondji, L. O. C. (2019). Land use change increases flood hazard: A multi-modelling approach to assess change in flood

characteristics driven by socio-economic land use change scenarios. *Natural Hazards*, *98*(3), 1021–1050. doi: 10.1007/s11069-018-3557-8

Hsu, C.-W., Chang, C.-C., & Lin, C.-J. (2016). *A Practical Guide to Support Vector Classification*. 16.

Huang, S., Xia, J., Wang, Y., Wang, W., Zeng, S., She, D., & Wang, G. (2022). Coupling Machine Learning Into Hydrodynamic Models to Improve River Modeling With Complex Boundary Conditions. *Water Resources Research*, *58*(10). doi: 10.1029/2022WR032183

Huang, T.-M., Kecman, V., & Kopriva, I. (2006). *Kernel Based Algorithms for Mining Huge Data Sets* (Vol. 17). Berlin Heidelberg: Springer-Verlag.

Hurst, H. E. (1951). Long-Term Storage Capacity of Reservoirs. *Transactions of the American Society of Civil Engineers*, *116*(1), 770–799. doi: 10.1061/TACEAT.0006518

Huss, M., Bookhagen, B., Huggel, C., Jacobsen, D., Bradley, R. S., Clague, J. J., … Winder, M. (2017). Toward mountains without permanent snow and ice: MOUNTAINS WITHOUT PERMANENT SNOW AND ICE. *Earth's Future*, *5*(5), 418–435. doi: 10.1002/2016EF000514

Huss, M., Jouvet, G., Farinotti, D., & Bauder, A. (2010). Future high-mountain hydrology: A new parameterization of glacier retreat. *Hydrology and Earth System Sciences*, *14*(5), 815–829. doi: 10.5194/hess-14-815-2010

Immerzeel, W. W., Lutz, A. F., Andrade, M., Bahl, A., Biemans, H., Bolch, T., … Baillie, J. E. M. (2020). Importance and vulnerability of the world's water towers. *Nature*, *577*(7790), 364–369. doi: 10.1038/s41586-019-1822-y

Ipcc. (2022). *The Ocean and Cryosphere in a Changing Climate: Special Report of the Intergovernmental Panel on Climate Change* (1st ed.). Cambridge University Press. doi: 10.1017/9781009157964

JCGM. (2008). Evaluation of measurement data—Guide to the expression of uncertainty in measurement. *Int. Organ. Stand. Geneva ISBN*, *50*, 134.

Jennings, K. S., & Molotch, N. P. (2019). The sensitivity of modeled snow accumulation and melt to precipitation phase methods across a climatic gradient. *Hydrology and Earth System Sciences*, *23*(9), 3765–3786. doi: 10.5194/hess-23-3765-2019

Jonas, T., Marty, C., & Magnusson, J. (2009). Estimating the snow water equivalent from snow depth measurements in the Swiss Alps. *Journal of Hydrology*, *378*(1), 161–167. doi: 10.1016/j.jhydrol.2009.09.021

Jones, B. E. (1916). *A method of correcting river discharge for a changing stage* (No. No. 375-E). US Geological Survey. doi: 10.3133/wsp375E

Kääb, A., Winsvold, S., Altena, B., Nuth, C., Nagler, T., & Wuite, J. (2016). Glacier Remote Sensing Using Sentinel-2. Part I: Radiometric and Geometric Performance, and Application to Ice Velocity. *Remote Sensing*, *8*(7), 598. doi: 10.3390/rs8070598

Kendall, M. G. (1938). A NEW MEASURE OF RANK CORRELATION. *Biometrika*, *30*(1–2), 81–93. doi: 10.1093/biomet/30.1-2.81

Kendall, M. G. (1975). *Rank correlation methods* (4th ed.). London: Griffin.

Kiang, J. E., Gazoorian, C., McMillan, H., Coxon, G., Le Coz, J., Westerberg, I. K., … Mason, R. (2018). A Comparison of Methods for Streamflow Uncertainty Estimation. *Water Resources Research*, *54*(10), 7149–7176. doi: 10.1029/2018WR022708

Knoflach, B., Ramskogler, K., Talluto, M., Hofmeister, F., Haas, F., Heckmann, T., … Stötter, J. (2021). Modelling of Vegetation Dynamics from Satellite Time Series to Determine Proglacial Primary Succession in the Course of Global Warming—A Case Study in the Upper Martell Valley (Eastern Italian Alps). *Remote Sensing*, *13*(21), 4450. doi: 10.3390/rs13214450

Kobierska, F., Jonas, T., Kirchner, J. W., & Bernasconi, S. M. (2015). Linking baseflow separation and groundwater storage dynamics in an alpine basin (Dammagletscher, Switzerland). *Hydrology and Earth System Sciences*, *19*(8), 3681–3693. doi: 10.5194/hess-19-3681-2015

Kochendorfer, J., Earle, M., Rasmussen, R., Smith, C., Yang, D., Morin, S., … Meyers, T. (2022). How Well Are We Measuring Snow Post-SPICE? *Bulletin of the American Meteorological Society*, *103*(2), E370–E388. doi: 10.1175/BAMS-D-20-0228.1

Kochendorfer, J., Rasmussen, R., Wolff, M., Baker, B., Hall, M. E., Meyers, T., … Leeper, R. (2016). *The Quantification and Correction of Wind-Induced Precipitation Measurement Errors* [Preprint]. Hydrometeorology/Instruments and observation techniques. doi: 10.5194/hess-2016-415

Kocsis, T., Kovács-Székely, I., & Anda, A. (2020). Homogeneity tests and non-parametric analyses of tendencies in precipitation time series in Keszthely, Western Hungary. *Theoretical and Applied Climatology*, *139*(3–4), 849–859. doi: 10.1007/s00704-019-03014-4

Kraller, G., Warscher, M., Kunstmann, H., Vogl, S., Marke, T., & Strasser, U. (2012). Water balance estimation in high Alpine terrain by combining distributed modeling and a neural network approach (Berchtesgaden Alps, Germany). *Hydrology and Earth System Sciences*, *16*(7), 1969–1990. doi: 10.5194/hess-16-1969-2012

Krogh, S. A., Scaff, L., Kirchner, J. W., Gordon, B., Sterle, G., & Harpold, A. (2022). Diel streamflow cycles suggest more sensitive snowmelt-driven streamflow to climate change than land surface modeling does. *Hydrology and Earth System Sciences*, *26*(13), 3393–3417. doi: 10.5194/hess-26-3393-2022

Kuhn, M., & Olefs, M. (2020). Elevation-Dependent Climate Change in the European Alps. In M. Kuhn & M. Olefs, *Oxford Research Encyclopedia of Climate Science*. Oxford University Press. doi: 10.1093/acrefore/9780190228620.013.762

Kunstmann, H., Schneider, K., Forkel, R., & Knoche, R. (2004). Impact analysis of climate change for an Alpine catchment using high resolution dynamic downscaling of ECHAM4 time slices. *Hydrology and Earth System Sciences*, *8*(6), 1031–1045. doi: 10.5194/hess-8-1031-2004

Lafdani, E. K., Nia, A. M., & Ahmadi, A. (2013). Daily suspended sediment load prediction using artificial neural networks and support vector machines. *Journal of Hydrology*, *478*, 50–62. doi: 10.1016/j.jhydrol.2012.11.048

Largeron, C., Dumont, M., Morin, S., Boone, A., Lafaysse, M., Metref, S., … Margulis, S. A. (2020). Toward Snow Cover Estimation in Mountainous Areas Using Modern Data Assimilation Methods: A Review. *Frontiers in Earth Science*, *8*, 325. doi: 10.3389/feart.2020.00325

Lehning, M., Löwe, H., Ryser, M., & Raderschall, N. (2008). Inhomogeneous precipitation distribution and snow transport in steep terrain: SNOW DRIFT AND INHOMOGENEOUS PRECIPITATION. *Water Resources Research*, *44*(7). doi: 10.1029/2007WR006545

Lehning, M., Völksch, I., Gustafsson, D., Nguyen, T. A., Stähli, M., & Zappa, M. (2006). ALPINE3D: A detailed model of mountain surface processes and its application to snow hydrology. *Hydrological Processes*, *20*(10), 2111–2128. doi: 10.1002/hyp.6204

Leibundgut, C., Maloszewski, P., & Külls, C. (2011). *Tracers in Hydrology*. John Wiley & Sons.

Li, J., Lu, W., & Luo, J. (2021). Groundwater contamination sources identification based on the Long-Short Term Memory network. *Journal of Hydrology*, *601*, 126670. doi: 10.1016/j.jhydrol.2021.126670

Li, X., Wang, N., & Wu, Y. (2022). Automated Glacier Snow Line Altitude Calculation Method Using Landsat Series Images in the Google Earth Engine Platform. *Remote Sensing*, *14*(10), 2377. doi: 10.3390/rs14102377

Li, Z., Yan, Z., Zhu, Y., Freychet, N., & Tett, S. (2020). Homogenized Daily Relative Humidity Series in China during 1960–2017. *Advances in Atmospheric Sciences*, *37*(4), 318–327. doi: 10.1007/s00376-020-9180-0

Liljedahl, A. K., Boike, J., Daanen, R. P., Fedorov, A. N., Frost, G. V., Grosse, G., … Zona, D. (2016). Pan-Arctic ice-wedge degradation in warming permafrost and its influence on tundra hydrology. *Nature Geoscience*, *9*(4), 312–318. doi: 10.1038/ngeo2674

Lin, G.-F., Wang, T.-C., & Chen, L.-H. (2016). A Forecasting Approach Combining Self-Organizing Map with Support Vector Regression for Reservoir Inflow during Typhoon Periods. *Advances in Meteorology*, *2016*, 1–12. doi: 10.1155/2016/7575126

Lin, H.-T., Lin, C.-J., & Weng, R. C. (2007). A note on Platt's probabilistic outputs for support vector machines. *Machine Learning*, *68*(3), 267–276. doi: 10.1007/s10994-007-5018-6

Liston, G. E., Haehnel, R. B., Sturm, M., Hiemstra, C. A., Berezovskaya, S., & Tabler, R. D. (2007). Simulating complex snow distributions in windy environments using SnowTran-3D. *Journal of Glaciology*, *53*(181), 241–256. doi: 10.3189/172756507782202865

Liu, M., & Lu, J. (2014). Support vector machine—an alternative to artificial neuron network for water quality forecasting in an agricultural nonpoint source polluted river? *Environmental Science and Pollution Research*, *21*(18), 11036–11053. doi: 10.1007/s11356-014-3046-x

Liu, Z., Zhou, P., & Zhang, Y. (2015). A Probabilistic Wavelet–Support Vector Regression Model for Streamflow Forecasting with Rainfall and Climate Information Input*. *Journal of Hydrometeorology*, *16*(5), 2209–2229. doi: 10.1175/JHM-D-14-0210.1

López-Moreno, J. I., Revuelto, J., Gilaberte, M., Morán-Tejeda, E., Pons, M., Jover, E., … Pomeroy, J. W. (2014). The effect of slope aspect on the response of snowpack to climate warming in the Pyrenees. *Theoretical and Applied Climatology*, *117*(1–2), 207–219. doi: 10.1007/s00704-013-0991-0

López-Moreno, J. I., & Stähli, M. (2008). Statistical analysis of the snow cover variability in a subalpine watershed: Assessing the role of topography and forest interactions. *Journal of Hydrology*, *348*(3–4), 379–394. doi: 10.1016/j.jhydrol.2007.10.018

Lucio-Eceiza, E. E., González-Rouco, J. F., Navarro, J., Beltrami, H., & Conte, J. (2018). Quality Control of Surface Wind Observations in Northeastern North America. Part II: Measurement Errors. *Journal of Atmospheric and Oceanic Technology*, *35*(1), 183–205. doi: 10.1175/JTECH-D-16-0205.1

Lundquist, J., Hughes, M., Gutmann, E., & Kapnick, S. (2019). Our Skill in Modeling Mountain Rain and Snow is Bypassing the Skill of Our Observational Networks. *Bulletin of the American Meteorological Society*, *100*(12), 2473–2490. doi: 10.1175/BAMS-D-19-0001.1

Maity, R., Bhagwat, P. P., & Bhatnagar, A. (2010). Potential of support vector regression for prediction of monthly streamflow using endogenous property. *Hydrological Processes*, *24*(7), 917–923. doi: 10.1002/hyp.7535

Manara, V., Beltrano, M. C., Brunetti, M., Maugeri, M., Sanchez-Lorenzo, A., Simolo, C., & Sorrenti, S. (2015). Sunshine duration variability and trends in Italy from homogenized instrumental time series (1936-2013): SUNSHINE DURATION VARIABILITY IN ITALY. *Journal of Geophysical Research: Atmospheres*, *120*(9), 3622–3641. doi: 10.1002/2014JD022560

Manara, V., Brunetti, M., Celozzi, A., Maugeri, M., Sanchez-Lorenzo, A., & Wild, M. (2016). Detection of dimming/brightening in Italy from homogenized all-sky andclear-sky surface solar radiation records and underlying causes (1959–2013). *Atmospheric Chemistry and Physics*, *16*(17), 11145–11161. doi: 10.5194/acp-16-11145-2016

Mann, H. B. (1945). Nonparametric Tests Against Trend. *Econometrica*, *13*(3), 245. doi: 10.2307/1907187

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

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

Martínez-Carreras, N., Hissler, C., Gourdol, L., Klaus, J., Juilleret, J., Iffly, J. F., & Pfister, L. (2016). Storage controls on the generation of double peak hydrographs in a forested

headwater catchment. *Journal of Hydrology*, *543*, 255–269. doi: 10.1016/j.jhydrol.2016.10.004

Marty, C. (2008). Regime shift of snow days in Switzerland: REGIME SHIFT OF SNOW DAYS. *Geophysical Research Letters*, *35*(12), n/a-n/a. doi: 10.1029/2008GL033998

Matiu, M, Jacob, A., & Notarnicola, C. (2019). *Daily MODIS snow cover maps for the European Alps from 2002 onwards at 250m horizontal resolution along with a nearly cloud-free version* (Version v1.0.1) [Data set]. Zenodo. doi: 10.5281/ZENODO.3566703

Matiu, Michael, Crespi, A., Bertoldi, G., Carmagnola, C. M., Marty, C., Morin, S., … Weilguni, V. (2021). Observed snow depth trends in the European Alps: 1971 to 2019. *The Cryosphere*, *15*(3), 1343–1382. doi: 10.5194/tc-15-1343-2021

Matiu, Michael, Jacob, A., & Notarnicola, C. (2019). Daily MODIS Snow Cover Maps for the European Alps from 2002 onwards at 250 m Horizontal Resolution Along with a Nearly Cloud-Free Version. *Data*, *5*(1), 1. doi: 10.3390/data5010001

Maussion, F., Butenko, A., Champollion, N., Dusch, M., Eis, J., Fourteau, K., … Marzeion, B. (2019). The Open Global Glacier Model (OGGM) v1.1. *Geoscientific Model Development*, *12*(3), 909–931. doi: 10.5194/gmd-12-909-2019

McCuen, R. H. (2016). *Modeling Hydrologic Change: Statistical Methods* (0 ed.). CRC Press. doi: 10.1201/9781420032192

Mcmillan, H., Jackson, B., Clark, M., Kavetski, D., & Woods, R. (2010). Input Uncertainty in Hydrological Models: An Evaluation of Error Models for Rainfall. *Journal of Hydrology*, *400*, 83–94.

McMillan, H. K., Coxon, G., Sikorska-Senoner, A. E., & Westerberg, I. K. (2022). Impacts of observational uncertainty on analysis and modelling of hydrological processes: Preface. *Hydrological Processes*, *36*(2). doi: 10.1002/hyp.14481

McMillan, H. K., Westerberg, I. K., & Krueger, T. (2018). Hydrological data uncertainty and its implications. *WIREs Water*, *5*(6). doi: 10.1002/wat2.1319

Meißl, G., Geitner, C., Batliner, A., Klebinder, K., Kohl, B., & Markart, G. (2021). Brixenbach research catchment: Quantification of runoff process proportions in a small Alpine

catchment depending on soil moisture states and precipitation characteristics. *Hydrological Processes*, *35*(6). doi: 10.1002/hyp.14186

Menekay, D. (2019). Accuracy Assessment Experiments of Cloud Masking Programs (Sen2Cor, Fmask, Idepix) | by Deniz Menekay | Medium. Retrieved June 2, 2021, from https://medium.com/@denizmenekay/accuracy-assessment-experiments-of-cloud-masking-programs-sen2cor-fmask-idepix-8c1b44fae739

Mestre, O., Domonkos, P., Picard, F., Auer, I., Robin, S., Boehm, R., … Stepanek, P. (2013). HOMER: A homogenization software - methods and applications. *IDOJARAS*, *117*(1), 47–67.

Meyer, S., Blaschek, M., Duttmann, R., & Ludwig, R. (2016). Improved hydrological model parametrization for climate change impact assessment under data scarcity—The potential of field monitoring techniques and geostatistics. *Science of The Total Environment*, *543*, 906–923. doi: 10.1016/j.scitotenv.2015.07.116

Mohammadi, B. (2021). A review on the applications of machine learning for runoff modeling. *Sustainable Water Resources Management*, *7*(6), 98. doi: 10.1007/s40899-021-00584-y

Moore, R. D. (2004). Introduction to salt dilution gauging for streamflow measurement: Part 1. *Streamline Watershed Management Bulletin*, *7*(4), 20–23.

Moore, R. D. (2005). Introduction to Salt Dilution Gauging for Streamflow Measurement Part III: Slug Injection Using Salt in Solution. Retrieved April 22, 2022, from https://www.semanticscholar.org/paper/Introduction-to-Salt-Dilution-Gauging-for-Part-III-Moore/392e384b3ac76e4877d6412f1751ea27a8d2b951

Morgenschweis, G. (2018). *Hydrometrie: Theorie und Praxis der Durchflussmessung in offenen Gerinnen*. Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-662-55314-5

Mott, R., Schirmer, M., Bavay, M., Grünewald, T., & Lehning, M. (2010). Understanding snow-transport processes shaping the mountain snow-cover. *The Cryosphere*, *4*(4), 545–559. doi: 10.5194/tc-4-545-2010

Mott, R., Vionnet, V., & Grünewald, T. (2018). The Seasonal Snow Cover Dynamics: Review on Wind-Driven Coupling Processes. *Frontiers in Earth Science*, *6*, 197. doi: 10.3389/feart.2018.00197

Mudelsee, M. (2010). *Climate Time Series Analysis*. Dordrecht: Springer Netherlands. doi: 10.1007/978-90-481-9482-7

Mudelsee, M. (2019). Trend analysis of climate time series: A review of methods. *Earth-Science Reviews*, *190*, 310–322. doi: 10.1016/j.earscirev.2018.12.005

Müller, T., Lane, S. N., & Schaefli, B. (2022). Towards a hydrogeomorphological understanding of proglacial catchments: An assessment of groundwater storage and release in an Alpine catchment. *Hydrology and Earth System Sciences*, *26*(23), 6029–6054. doi: 10.5194/hess-26-6029-2022

Müller, T., Roncoroni, M., Mancini, D., Lane, S. N., & Schaefli, B. (2023). *Current and future role of meltwater-groundwater dynamics in a proglacial Alpine outwash plain* [Preprint]. Groundwater hydrology/Modelling approaches. doi: 10.5194/egusphere-2022-1503

Murphy, S. (2018). *Atmospheric Correction of Sentinel 2 Imagery in Google Earth Engine Using Py6S* [Jupyter Notebook]. Retrieved from https://github.com/samsammurphy/gee-atmcorr-S2 (Original work published 2017)

Mutzner, R., Tarolli, P., Sofia, G., Parlange, M. B., & Rinaldo, A. (2016). Field study on drainage densities and rescaled width functions in a high-altitude alpine catchment: Heterogeneous Drainage Densities in a High-altitude Alpine Catchment. *Hydrological Processes*, *30*(13), 2138–2152. doi: 10.1002/hyp.10783

Mutzner, R., Weijs, S. V., Tarolli, P., Calaf, M., Oldroyd, H. J., & Parlange, M. B. (2015). Controls on the diurnal streamflow cycles in two subbasins of an alpine headwater catchment. *Water Resources Research*, *51*(5), 3403–3418. doi: 10.1002/2014WR016581

Noether, G. E. (1956). Two Sequential Tests Against Trend. *Journal of the American Statistical Association*, *51*(275), 440–450. doi: 10.1080/01621459.1956.10501335

Notarnicola, C. (2020). Hotspots of snow cover changes in global mountain regions over 2000–2018. *Remote Sensing of Environment*, *243*, 111781. doi: 10.1016/j.rse.2020.111781

Notarnicola, C. (2022). Overall negative trends for snow cover extent and duration in global mountain regions over 1982–2020. *Scientific Reports*, *12*(1), 13731. doi: 10.1038/s41598-022-16743-w

Oke, T. R. (2002). *Boundary layer climates*. Routledge.

Olefs, M., Koch, R., Schöner, W., & 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. doi: 10.3390/atmos11121330

Orlowski, N., Breuer, L., Angeli, N., Boeckx, P., Brumbt, C., Cook, C. S., … McDonnell, J. J. (2018). Inter-laboratory comparison of cryogenic water extraction systems for stable isotope analysis of soil water. *Hydrology and Earth System Sciences*, *22*(7), 3619–3637. doi: 10.5194/hess-22-3619-2018

Østrem, G. (1964). A method of measuring water discharge in turbulent streams. *Geographical Bulletin*, (21), 21–43.

OTT. (2023). *Grundwasser-Datensammler / Datenlogger für Wasserstand, Temperatur und Leitfähigkeit OTT CTD*. Retrieved from https://www.ott.com/de-ch/produkte/wasserstand-49/ott-ctd-62/productAction/outputAsPdf/

Painter, K. J., Gentile, A., & Ferraris, S. (2023). A stochastic cellular automaton model to describe the evolution of the snow-covered area across a high-elevation mountain catchment. *Science of The Total Environment*, *857*, 159195. doi: 10.1016/j.scitotenv.2022.159195

Pandžić, K., Kobold, M., Oskoruš, D., Biondić, B., Biondić, R., Bonacci, O., … Curić, O. (2020). Standard normal homogeneity test as a tool to detect change points in climate-related river discharge variation: Case study of the Kupa River Basin. *Hydrological Sciences Journal*, *65*(2), 227–241. doi: 10.1080/02626667.2019.1686507

Paniconi, C., & Putti, M. (2015). Physically based modeling in catchment hydrology at 50: Survey and outlook. *Water Resources Research*, *51*(9), 7090–7129. doi: 10.1002/2015WR017780

Papathoma-Köhle, M., Zischg, A., Fuchs, S., Glade, T., & Keiler, M. (2015). Loss estimation for landslides in mountain areas – An integrated toolbox for vulnerability assessment and damage documentation. *Environmental Modelling & Software*, *63*, 156–169. doi: 10.1016/j.envsoft.2014.10.003

Parisouj, P., Mohebzadeh, H., & Lee, T. (2020). Employing Machine Learning Algorithms for Streamflow Prediction: A Case Study of Four River Basins with Different Climatic Zones in the United States. *Water Resources Management*, *34*(13), 4113–4131. doi: 10.1007/s11269-020-02659-5

Patakamuri, S. K., & O'Brien, N. (2021). *modifiedmk: Modified Versions of Mann Kendall and Spearman's Rho Trend Tests*. Retrieved from https://CRAN.R-project.org/package=modifiedmk

Pearson, C. P. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling: The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science: Vol 50, No 302. Retrieved September 20, 2021, from https://www.tandfonline.com/doi/abs/10.1080/14786440009463897

Pedersen, S. H., Tamstorf, M. P., Abermann, J., Westergaard-Nielsen, A., Lund, M., Skov, K., … Schmidt, N. M. (2016). Spatiotemporal Characteristics of Seasonal Snow Cover in Northeast Greenland from in Situ Observations. *Arctic, Antarctic, and Alpine Research*, *48*(4), 653–671. doi: 10.1657/AAAR0016-028

Penna, D., Engel, M., Mao, L., Dell'Agnese, A., Bertoldi, G., & Comiti, F. (2014). Tracer-based analysis of spatial and temporal variations of water sources in a glacierized catchment. *Hydrology and Earth System Sciences*, *18*(12), 5271–5288. doi: 10.5194/hess-18-5271-2014

Penna, D., Stenni, B., Šanda, M., Wrede, S., Bogaard, T. A., Gobbi, A., … Chárová, Z. (2010). On the reproducibility and repeatability of laser absorption spectroscopy measurements for δ^2H and $\delta^{18}O$ isotopic analysis. *Hydrology and Earth System Sciences*, *14*(8), 1551–1566. doi: 10.5194/hess-14-1551-2010

Penna, D., van Meerveld, H. J., Zuecco, G., Dalla Fontana, G., & Borga, M. (2016). Hydrological response of an Alpine catchment to rainfall and snowmelt events. *Journal of Hydrology*, *537*, 382–397. doi: 10.1016/j.jhydrol.2016.03.040

Penna, D., Zuecco, G., Crema, S., Trevisani, S., Cavalli, M., Pianezzola, L., … Borga, M. (2017). Response time and water origin in a steep nested catchment in the Italian Dolomites. *Hydrological Processes*, *31*(4), 768–782. doi: 10.1002/hyp.11050

Pepin, N. C., Arnone, E., Gobiet, A., Haslinger, K., Kotlarski, S., Notarnicola, C., … Adler, C. (2022). Climate Changes and Their Elevational Patterns in the Mountains of the World. *Reviews of Geophysics*, *60*(1). doi: 10.1029/2020RG000730

Perret, E., Lang, M., & Le Coz, J. (2022). A framework for detecting stage-discharge hysteresis due to flow unsteadiness: Application to France's national hydrometry network. *Journal of Hydrology*, *608*, 127567. doi: 10.1016/j.jhydrol.2022.127567

Pesci, M. H., & Förster, K. (2022). *Process-based water balance modelling with explicit ice flow dynamics: WaSiM-OGGM coupling scheme* [Other]. oral. doi: 10.5194/iahs2022-171

Pettitt, A. N. (1979). A Non-Parametric Approach to the Change-Point Problem. *Applied Statistics*, *28*(2), 126. doi: 10.2307/2346729

Philipona, R. (2013). Greenhouse warming and solar brightening in and around the Alps. *International Journal of Climatology*, *33*(6), 1530–1537. doi: 10.1002/joc.3531

Pomeroy, J. W., Gray, D. M., Brown, T., Hedstrom, N. R., Quinton, W. L., Granger, R. J., & Carey, S. K. (2007). The cold regions hydrological model: A platform for basing process representation and model structure on physical evidence. *Hydrological Processes*, *21*(19), 2650–2667. doi: 10.1002/hyp.6787

Poschlod, B., Willkofer, F., & Ludwig, R. (2020). Impact of Climate Change on the Hydrological Regimes in Bavaria. *Water*, *12*(6), 1599. doi: 10.3390/w12061599

Puspitarini, H. D., François, B., Zaramella, M., Brown, C., & Borga, M. (2020). The impact of glacier shrinkage on energy production from hydropower-solar complementarity in alpine river basins. *Science of The Total Environment*, *719*, 137488. doi: 10.1016/j.scitotenv.2020.137488

Qi, Y., Wang, H., Ma, X., Zhang, J., & Yang, R. (2021). Relationship between vegetation phenology and snow cover changes during 2001–2018 in the Qilian Mountains. *Ecological Indicators*, *133*, 108351. doi: 10.1016/j.ecolind.2021.108351

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

Racoviteanu, A. E., Rittger, K., & Armstrong, R. (2019). An Automated Approach for Estimating Snowline Altitudes in the Karakoram and Eastern Himalaya From Remote Sensing. *Frontiers in Earth Science*, *7*, 220. doi: 10.3389/feart.2019.00220

Ramskogler, K., Knoflach, B., Elsner, B., Erschbamer, B., Haas, F., Heckmann, T., … Tasser, E. (2023). Primary succession and its driving variables – a sphere-spanning approach applied in proglacial areas in the upper Martell Valley (Eastern Italian Alps). *Biogeosciences*, *20*(14), 2919–2939. doi: 10.5194/bg-20-2919-2023

Ranghetti, L., Boschetti, M., Nutini, F., & Busetto, L. (2020). "sen2r": An R toolbox for automatically downloading and preprocessing Sentinel-2 satellite data. *Computers & Geosciences*, *139*, 104473. doi: 10.1016/j.cageo.2020.104473

Rango, A., Wergin, W. P., & Erbe, E. F. (1996). Snow crystal imaging using scanning electron microscopy: II. Metamorphosed snow. *Hydrological Sciences Journal*, *41*(2), 235–250. doi: 10.1080/02626669609491495

Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A. P., … Gutmann, E. (2012). How Well Are We Measuring Snow: The NOAA/FAA/NCAR Winter Precipitation Test Bed. *Bulletin of the American Meteorological Society*, *93*(6), 811–829. doi: 10.1175/BAMS-D-11-00052.1

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. doi: 10.1038/s41586-019-0912-1

Resch, G., Koch, R., Marty, C., Chimani, B., Begert, M., Buchmann, M., … Schöner, W. (2022). A quantile-based approach to improve homogenization of snow depth time series. *International Journal of Climatology*, joc.7742. doi: 10.1002/joc.7742

Resources Information Standards Committee (RISC). (2018). *Manual of British Columbia hydrometric standards* (2.0). Victoria, B.C.: The Province of British Columbia. Retrieved from http://www.llbc.leg.bc.ca/public/pubdocs/bcdocs2019/694598/man_bc_hydrometric_stand_v2.pdf

Riano, D., Chuvieco, E., Salas, J., & Aguado, I. (2003). Assessment of different topographic corrections in landsat-TM data for mapping vegetation types (2003). *IEEE Transactions on Geoscience and Remote Sensing*, *41*(5), 1056–1061. doi: 10.1109/TGRS.2003.811693

Richardson, M., Moore, R. D. (Dan), & Zimmermann, A. (2017). Variability of tracer breakthrough curves in mountain streams: Implications for streamflow measurement by slug injection. *Canadian Water Resources Journal / Revue Canadienne Des Ressources Hydriques*, *42*(1), 21–37. doi: 10.1080/07011784.2016.1212676

Richardson, M., Sentlinger, G., Moore, R. D. (Dan), & Zimmermann, A. (2017). Quantifying the Relation Between Electrical Conductivity and Salt Concentration for Dilution Gauging Via Dry Salt Injection. *Confluence: Journal of Watershed Science and Management*, *1*(2), 16. doi: 10.22230/jwsm.2017v1n2a1

Riggs, G. A., & Hall, D. K. (2015). *MODIS Snow Products Collection 6 User Guide*. Retrieved from https://nsidc.org/sites/nsidc.org/files/files/MODIS-snow-user-guide-C6.pdf

Rogger, M., Chirico, G. B., Hausmann, H., Krainer, K., Brückl, E., Stadler, P., & Blöschl, G. (2017). Impact of mountain permafrost on flow path and runoff response in a high alpine

catchment. *Water Resources Research*, *53*(2), 1288–1308. doi: 10.1002/2016WR019341

Rom, J., Haas, F., Heckmann, T., Altmann, M., Fleischer, F., Ressl, C., … Becht, M. (2023). Spatio-temporal analysis of slope-type debris flow activity in Horlachtal, Austria, based on orthophotos and lidar data since 1947. *Natural Hazards and Earth System Sciences*, *23*(2), 601–622. doi: 10.5194/nhess-23-601-2023

Rom, J., Haas, F., Hofmeister, F., Fleischer, F., Altmann, M., Pfeiffer, M., … Becht, M. (2023). Analysing the Large-Scale Debris Flow Event in July 2022 in Horlachtal, Austria Using Remote Sensing and Measurement Data. *Geosciences*, *13*(4), 100. doi: 10.3390/geosciences13040100

Romanov, P., Gutman, G., & Csiszar, I. (2000). Automated Monitoring of Snow Cover over North America with Multispectral Satellite Data. *Journal of Applied Meteorology*, *39*(11), 1866–1880. doi: 10.1175/1520-0450(2000)039<1866:AMOSCO>2.0.CO;2

Rottler, E., Kormann, C., Francke, T., & Bronstert, A. (2019). Elevation-dependent warming in the Swiss Alps 1981–2017: Features, forcings and feedbacks. *International Journal of Climatology*, *39*(5), 2556–2568. doi: 10.1002/joc.5970

Rounce, D. R., Hock, R., Maussion, F., Hugonnet, R., Kochtitzky, W., Huss, M., … McNabb, R. W. (2023). Global glacier change in the 21st century: Every increase in temperature matters. *Science*, *379*(6627), 78–83. doi: 10.1126/science.abo1324

Sanchez-Lorenzo, A., Calbó, J., & Wild, M. (2013). Global and diffuse solar radiation in Spain: Building a homogeneous dataset and assessing their trends. *Global and Planetary Change*, *100*, 343–352. doi: 10.1016/j.gloplacha.2012.11.010

Satellite Imaging Corporation. (2021). Sentinel-2A SatelliteSensor | Satellite Imaging Corp. Retrieved April 29, 2021, from https://www.satimagingcorp.com/satellite-sensors/other-satellite-sensors/sentinel-2a/

Sauter, T., Möller, M., Finkelnburg, R., Grabiec, M., Scherer, D., & Schneider, C. (2013). Snowdrift modelling for the Vestfonna ice cap, north-eastern Svalbard. *The Cryosphere*, *7*(4), 1287–1301. doi: 10.5194/tc-7-1287-2013

Saydi, M., & Ding, J. (2020). Impacts of topographic factors on regional snow cover characteristics. *Water Science and Engineering*, *13*(3), 171–180. doi: 10.1016/j.wse.2020.09.002

Schattan, P., Schwaizer, G., Schöber, J., & Achleitner, S. (2020). The complementary value of cosmic-ray neutron sensing and snow covered area products for snow hydrological modelling. *Remote Sensing of Environment*, *239*, 111603. doi: 10.1016/j.rse.2019.111603

Scherrer, S. C., Fischer, E. M., Posselt, R., Liniger, M. A., Croci-Maspoli, M., & Knutti, R. (2016). Emerging trends in heavy precipitation and hot temperature extremes in Switzerland: Trends in Swiss Climate Extremes. *Journal of Geophysical Research: Atmospheres*, *121*(6), 2626–2637. doi: 10.1002/2015JD024634

Schmid, S. M., Fügenschuh, B., Kissling, E., & Schuster, R. (2004). Tectonic map and overall architecture of the Alpine orogen. *Eclogae Geologicae Helvetiae*, *97*(1), 93–117. doi: 10.1007/s00015-004-1113-x

Schneiderbauer, S., & Prokop, A. (2011). The atmospheric snow-transport model: SnowDrift3D. *Journal of Glaciology*, *57*(203), 526–542. doi: 10.3189/002214311796905677

Schöber, J., & Hofer, B. (2018). THE SEDIMENT BUDGET OF THE GLACIAL STREAMS IN THE CATCHMENT AREA OF THE GEPATSCH RESERVOIR IN THE ÖTZTAL ALPS IN THE PERIOD 1965-2015. *Twenty-sixth congress on large dams*. Presented at the Proceedings of ICOLD 2, Place of publication not identified. Place of publication not identified: CRC Press.

Schöber, J., Schneider, K., Helfricht, K., Schattan, P., Achleitner, S., Schöberl, F., & Kirnbauer, R. (2014). Snow cover characteristics in a glacierized catchment in the Tyrolean Alps—Improved spatially distributed modelling by usage of Lidar data. *Journal of Hydrology*, *519*, 3492–3510. doi: 10.1016/j.jhydrol.2013.12.054

Schöner, W., Koch, R., Matulla, C., Marty, C., & 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. doi: 10.1002/joc.5902

Schönwiese, C.-D. (2013). *Praktische Statistik für Meteorologen und Geowissenschaftler* (5 vollständig überarbeitete und erweiterte Aufl). Berlin, Stuttgart: Borntraeger.

Schulla, J. (2021). *Model Description WaSiM (Water balance Simulation Model)*. Retrieved from http://www.wasim.ch/downloads/doku/wasim/wasim_2021_en.pdf

Sentlinger, G., Fraser, J., & Baddock, E. (2019). Salt dilution flow measurement: Automation and uncertainty. *HydroSenSoft,International Symposium and Exhibition on Hydro-Environment Sensors and Software.*, *Vol. 8*, 8.

Serco Italia SPA. (2017). *Snow Cover Mapping with Sentinel-2 (SNAP)*. Research and User Support for Sentinel Core Products. Retrieved from https://rus-copernicus.eu/portal/wp-content/uploads/library/education/training/CRYO03_SnowCover_Sumava_Tutorial.pdf

Sevruk, B. (1991). International workshop on precipitation measurement I: Preface. *Hydrological Processes*, *5*(3), 229–232. doi: 10.1002/hyp.3360050302

Shen, L., Lu, L., Hu, T., Lin, R., Wang, J., & Xu, C. (2018). Homogeneity Test and Correction of Daily Temperature and Precipitation Data (1978–2015) in North China. *Advances in Meteorology*, *2018*, 1–17. doi: 10.1155/2018/4712538

Simoni, S., Padoan, S., Nadeau, D. F., Diebold, M., Porporato, A., Barrenetxea, G., … Parlange, M. B. (2011). Hydrologic response of an alpine watershed: Application of a meteorological wireless sensor network to understand streamflow generation. *Water Resources Research*, *47*(10), 16. doi: 10.1029/2011WR010730

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, *14*(3), 199–222. doi: 10.1023/B:STCO.0000035301.49549.88

Son, G., Kim, D., Kim, K., & Roh, Y. (2023). Performance of a Rectangular-Shaped Surface Velocity Radar for River Velocity Measurements. *KSCE Journal of Civil Engineering*, *27*(3), 1077–1092. doi: 10.1007/s12205-023-0272-0

Sonny. (2017). Digitale LiDAR-Geländemodelle von Italien, San Marino, Vatikanstadt, Malta | Digital LiDAR-Terrain Models of Italy, San Marino, Vatican City, Malta—Open Data Portal Austria. Retrieved October 12, 2021, from http://data.opendataportal.at/dataset/dtm-italy

Springer, J., Ludwig, R., & Kienzle, S. (2015). Impacts of Forest Fires and Climate Variability on the Hydrology of an Alpine Medium Sized Catchment in the Canadian Rocky Mountains. *Hydrology*, *2*(1), 23–47. doi: 10.3390/hydrology2010023

Stahl, K., Moore, R. D., Shea, J. M., Hutchinson, D., & Cannon, A. J. (2008). Coupled modelling of glacier and streamflow response to future climate scenarios: MODELLING OF GLACIER AND STREAMFLOW. *Water Resources Research*, *44*(2). doi: 10.1029/2007WR005956

Strasser, U., Förster, K., Formayer, H., Hofmeister, F., Marke, T., Meißl, G., … Schermer, M. (2019). Storylines of combined future land use and climate scenarios and their hydrological impacts in an Alpine catchment (Brixental/Austria). *Science of The Total Environment*, *657*, 746–763. doi: 10.1016/j.scitotenv.2018.12.077

Szentimrey, T. (1999). Multiple analysis of series for homogenization (MASH). *Proceedings of the Second Seminar for Homogenization of Surface Climatological Data*, 27–46. WMO Budapest, Hungary.

Tarrio, K., Tang, X., Masek, J. G., Claverie, M., Ju, J., Qiu, S., … Woodcock, C. E. (2020). Comparison of cloud detection algorithms for Sentinel-2 imagery. *Science of Remote Sensing*, *2*, 100010. doi: 10.1016/j.srs.2020.100010

Thornton, J. M., Brauchli, T., Mariethoz, G., & Brunner, P. (2021). Efficient multi-objective calibration and uncertainty analysis of distributed snow simulations in rugged alpine terrain. *Journal of Hydrology*, *598*, 126241. doi: 10.1016/j.jhydrol.2021.126241

Thumser, P., Haas, C., Tuhtan, J. A., Fuentes-Pérez, J. F., & Toming, G. (2017). RAPTOR-UAV: Real-time particle tracking in rivers using an unmanned aerial vehicle: RAPTOR-UAV: REAL-TIME PARTICLE TRACKING IN RIVERS USING AN UAV. *Earth Surface Processes and Landforms*, *42*(14), 2439–2446. doi: 10.1002/esp.4199

Tiel, M., Stahl, K., Freudiger, D., & Seibert, J. (2020). Glacio-hydrological model calibration and evaluation. *WIREs Water*, *7*(6). doi: 10.1002/wat2.1483

Totschnig, R., & Fuchs, S. (2013). Mountain torrents: Quantifying vulnerability and assessing uncertainties. *Engineering Geology*, *155*, 31–44. doi: 10.1016/j.enggeo.2012.12.019

Tsang, L., Durand, M., Derksen, C., Barros, A. P., Kang, D.-H., Lievens, H., … Xu, X. (2022). Review article: Global monitoring of snow water equivalent using high-frequency radar remote sensing. *The Cryosphere*, *16*(9), 3531–3573. doi: 10.5194/tc-16-3531-2022

Tudoroiu, M., Eccel, E., Gioli, B., Gianelle, D., Schume, H., Genesio, L., & Miglietta, F. (2016). Negative elevation-dependent warming trend in the Eastern Alps. *Environmental Research Letters*, *11*(4), 044021. doi: 10.1088/1748-9326/11/4/044021

Tuo, Y., Marcolini, G., Disse, M., & Chiogna, G. (2018). A multi-objective approach to improve SWAT model calibration in alpine catchments. *Journal of Hydrology*, *559*, 347–360. doi: 10.1016/j.jhydrol.2018.02.055

Tyralis, H. (2016). *HKprocess: Hurst-Kolmogorov Process*. Retrieved from https://CRAN.R-project.org/package=HKprocess

U.S. Environmental Protection Agency (EPA). (2014). Best Practices for Continuous Monitoring of Temperature and Flow in Wadeable Streams. *Global Change Research Program, National Center for Environmental Assessment*, p. 129.

Vanham, D., Fleischhacker, E., & Rauch, W. (2008). Technical Note: Seasonality in alpine water resources management – a regional assessment. *Hydrology and Earth System Sciences*, *12*(1), 91–100. doi: 10.5194/hess-12-91-2008

Vermote, E. F., Tanre, D., Deuze, J. L., Herman, M., & Morcette, J.-J. (1997). Second Simulation of the Satellite Signal in the Solar Spectrum, 6S: An overview. *IEEE Transactions on Geoscience and Remote Sensing*, *35*(3), 675–686. doi: 10.1109/36.581987

Vionnet, V., Marsh, C. B., Menounos, B., Gascoin, S., Wayand, N. E., Shea, J., … Pomeroy, J. W. (2021). Multi-scale snowdrift-permitting modelling of mountain snowpack. *The Cryosphere*, *15*(2), 743–769. doi: 10.5194/tc-15-743-2021

Vionnet, V., Martin, E., Masson, V., Guyomarc'h, G., Naaim-Bouvet, F., Prokop, A., … Lac, C. (2014). Simulation of wind-induced snow transport and sublimation in alpine terrain using a fully coupled snowpack/atmosphere model. *The Cryosphere*, *8*(2), 395–415. doi: 10.5194/tc-8-395-2014

Volchak, A. A., Sidak, S. V., & Parfomuk, S. I. (2021). STATISTICAL ASSESSMENT OF THE RUNOFF TIME SERIES HOMOGENEITY OF THE RIVERS IN BELARUS. *Vestnik of Brest State Technical University. Civil Engineering and Architecture*, 92–95. doi: 10.36773/1818-1112-2021-126-3-92-95

Wagner, T., Brodacz, A., Krainer, K., & Winkler, G. (2020). Active rock glaciers as shallow groundwater reservoirs, Austrian Alps. *Grundwasser*, *25*(3), 215–230. doi: 10.1007/s00767-020-00455-x

Wagner, T., Pauritsch, M., & Winkler, G. (2016). Impact of relict rock glaciers on spring and stream flow of alpine watersheds: Examples of the Niedere Tauern Range, Eastern Alps (Austria). *Austrian Journal of Earth Sciences*, *109*(1). doi: 10.17738/ajes.2016.0006

Wallace, D. L. (1959). Simplified Beta-Approximations to the Kruskal-Wallis H Test. *Journal of the American Statistical Association*, *54*(285), 225–230. doi: 10.1080/01621459.1959.10501508

Wan, Y., Gao, Q., Li, Y., Qin, X., Ganjurjav, Zhang, W., … Liu, S. (2014). Change of Snow Cover and Its Impact on Alpine Vegetation in the Source Regions of Large Rivers on the Qinghai-Tibetan Plateau, China. *Arctic, Antarctic, and Alpine Research*, *46*(3), 632–644. doi: 10.1657/1938-4246-46.3.632

Wang, W.-C., Chau, K.-W., Cheng, C.-T., & Qiu, L. (2009). A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology*, *374*(3–4), 294–306. doi: 10.1016/j.jhydrol.2009.06.019

Warscher, M. (2014). *Performance of Complex Snow Cover Descriptions in a Distributed Hydrological Model System and Simulation of Future Snow Cover and Discharge*

Characteristics: A Case Study for the High Alpine Terrain of the Berchtesgaden Alps. University Augsburg, Augsburg.

Warscher, M., Strasser, U., Kraller, G., Marke, T., Franz, H., & Kunstmann, H. (2013). Performance of complex snow cover descriptions in a distributed hydrological model system: A case study for the high Alpine terrain of the Berchtesgaden Alps. *Water Resources Research*, *49*(5), 2619–2637. doi: 10.1002/wrcr.20219

Weijs, S. V., Mutzner, R., & Parlange, M. B. (2013). Could electrical conductivity replace water level in rating curves for alpine streams?: ELECTRICAL CONDUCTIVITY STREAMFLOW RATING CURVES. *Water Resources Research*, *49*(1), 343–351. doi: 10.1029/2012WR012181

Welber, M., Le Coz, J., Laronne, J. B., Zolezzi, G., Zamler, D., Dramais, G., … Salvaro, M. (2016). Field assessment of noncontact stream gauging using portable surface velocity radars (SVR): FIELD ASSESSMENT OF PORTABLE SURFACE VELOCITY RADARS. *Water Resources Research*, *52*(2), 1108–1126. doi: 10.1002/2015WR017906

Whitfield, P. H., & Pomeroy, J. W. (2017). Assessing the quality of the streamflow record for a long-term reference hydrometric station: Bow River at Banff. *Canadian Water Resources Journal / Revue Canadienne Des Ressources Hydriques*, *42*(4), 391–415. doi: 10.1080/07011784.2017.1399086

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. doi: 10.1038/sdata.2016.18

Willems, W., Dick, T., Stricker, K., & Kasper, G. (2013). *Ingenieurhydrologie, Angewandte Wasserwirtschaft und Geoinformatik. Hystat Benutzerhandbuch. Vesion 4.0.8.*

Wilson, R. T. (2012). Py6S: A Python interface to the 6S radiative transfer model. *Computers & Geosciences*, *51*, 166–171. doi: 10.1016/j.cageo.2012.08.002

WMO. (2017). *Guide to the Implementation of Quality Management Systems for National Meteorological and Hydrological Services and Other Relevant Service Providers*. World

Meteorological Organization. (Geneva, Switzerland). Retrieved from https://library.wmo.int/doc_num.php?explnum_id=4141

WMO. (2020). *Guidelines on homogenization*. World Meteorological Organization. (Geneva, Switzerland). Retrieved from https://library.wmo.int/doc_num.php?explnum_id=10352

Wood, P. J., & Dykes, A. P. (2002). The use of salt dilution gauging techniques: Ecological considerations and insights. *Water Research*, *36*(12), 3054–3062. doi: 10.1016/S0043-1354(01)00519-X

Wortmann, M., Bolch, T., Su, B., & Krysanova, V. (2019). An efficient representation of glacier dynamics in a semi-distributed hydrological model to bridge glacier and river catchment scales. *Journal of Hydrology*, *573*, 136–152. doi: 10.1016/j.jhydrol.2019.03.006

Yang, S., Wang, X. L., & Wild, M. (2018). Homogenization and Trend Analysis of the 1958–2016 In Situ Surface Solar Radiation Records in China. *Journal of Climate*, *31*(11), 4529–4541. doi: 10.1175/JCLI-D-17-0891.1

Young, C.-C., Liu, W.-C., & Wu, M.-C. (2017). A physically based and machine learning hybrid approach for accurate rainfall-runoff modeling during extreme typhoon events. *Applied Soft Computing*, *53*, 205–216. doi: 10.1016/j.asoc.2016.12.052

Yozgatligil, C., & Yazici, C. (2016). Comparison of homogeneity tests for temperature using a simulation study: COMPARISON OF HOMOGENEITY TESTS. *International Journal of Climatology*, *36*(1), 62–81. doi: 10.1002/joc.4329

Yuan, L., & Forshay, K. J. (2022). Evaluating Monthly Flow Prediction Based on SWAT and Support Vector Regression Coupled with Discrete Wavelet Transform. *Water*, *14*(17), 2649. doi: 10.3390/w14172649

Yue, S., & Wang, C. (2004). The Mann-Kendall Test Modified by Effective Sample Size to Detect Trend in Serially Correlated Hydrological Series. *Water Resources Management*, *18*(3), 201–218. doi: 10.1023/B:WARM.0000043140.61082.60

Zampieri, M., Scoccimarro, E., Gualdi, S., & Navarra, A. (2015). Observed shift towards earlier spring discharge in the main Alpine rivers. *Science of The Total Environment*, *503–504*, 222–232. doi: 10.1016/j.scitotenv.2014.06.036

Zanchettin, D., Traverso, P., & Tomasino, M. (2008). Po River discharges: A preliminary analysis of a 200-year time series. *Climatic Change*, *89*(3–4), 411–433. doi: 10.1007/s10584-008-9395-z

Zeng, Z., Chen, A., Ciais, P., Li, Y., Li, L. Z. X., Vautard, R., … Piao, S. (2015). Regional air pollution brightening reverses the greenhouse gases induced warming-elevation relationship. *Geophysical Research Letters*, *42*(11), 4563–4572. doi: 10.1002/2015GL064410

Zhang, J. L., Li, Y. P., Huang, G. H., Wang, C. X., & Cheng, G. H. (2016). Evaluation of Uncertainties in Input Data and Parameters of a Hydrological Model Using a Bayesian Framework: A Case Study of a Snowmelt–Precipitation-Driven Watershed. *Journal of Hydrometeorology*, *17*(8), 2333–2350. doi: 10.1175/JHM-D-15-0236.1

Zuecco, G., Carturan, L., De Blasi, F., Seppi, R., Zanoner, T., Penna, D., … Dalla Fontana, G. (2018). Understanding hydrological processes in glacierized catchments: Evidence and implications of highly-variable isotopic and electrical conductivity data. *Hydrological Processes*, hyp.13366. doi: 10.1002/hyp.13366

Article supplementary materials

A.1. Supplementary material to chapter 2

Hofmeister, Graziano, et al., 2023[5]

A.1.1: Metadata of all available hydrometeorological time series from Horachtal and Martelltal. A segment of a time series was included in the analysis, if it covered at least five years of observation. Number of analyzed segments are indicated in brackets.

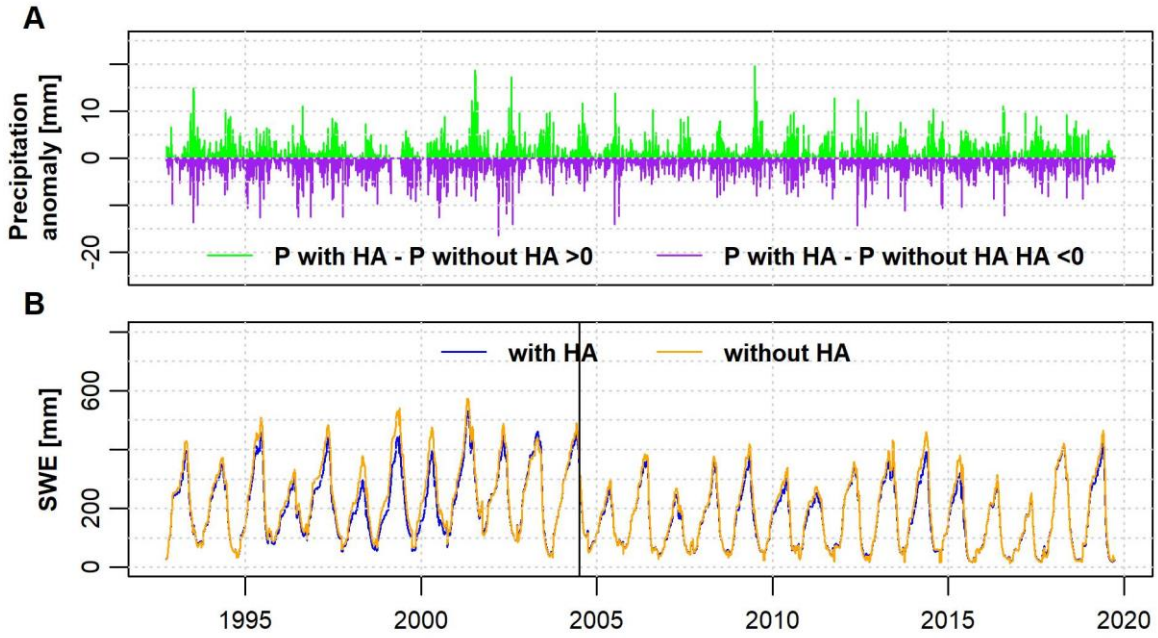
Station Name	Station code	Variable	Number of segments (analysed)	Latitude	Longitude	Elevation m a.s.l.	Obs. period	Provider
Durrapl att	1	P	2 (1)	46.5274	10.7373	1616	1989-2005	Bozen
Durrapl att	1	T	5 (0)	46.5274	10.7373	1616	1990-2005	Bozen
Hinterm artell	2	H	1 (1)	46.5169	10.7269	1720	2009-2020	Bozen
Hinterm artell	2	P	1 (1)	46.5169	10.7269	1720	2009-2020	Bozen
Hinterm artell	2	R	1 (1)	46.5169	10.7269	1720	2009-2020	Bozen
Hinterm artell	2	T	2 (2)	46.5169	10.7269	1720	2009-2020	Bozen
Hinterm artell	2	WS	1 (1)	46.5169	10.7269	1720	2009-2020	Bozen
Horlach alm	3	HS	4 (2)	47.1577	11.0125	1910	1988-2018	TIWAG
Horlach alm	3	P	1 (1)	47.1577	11.0125	1910	1989-2018	TIWAG
Horlach alm	3	T	1 (1)	47.1577	11.0125	1910	1987-2018	TIWAG
Horlach Fassun g	4	Q	1 (1)	47.1577	11.0125	1912	1985-2018	TIWAG
Kühtai	5	H	1 (1)	47.2071	11.0059	1918	1990-2019	TIWAG
Kühtai	5	P	2 (1)	47.2071	11.0059	1918	1990-2019	TIWAG

[5]Material from:
Hofmeister, F., Graziano, F., Marcolini, G., Willems, W., Disse, M., & Chiogna, G. (2023). Quality assessment of hydrometeorological observational data and their influence on hydrological model results in Alpine catchments. *Hydrological Sciences Journal*, 02626667.2023.2172335. doi: 10.1080/02626667.2023.2172335.

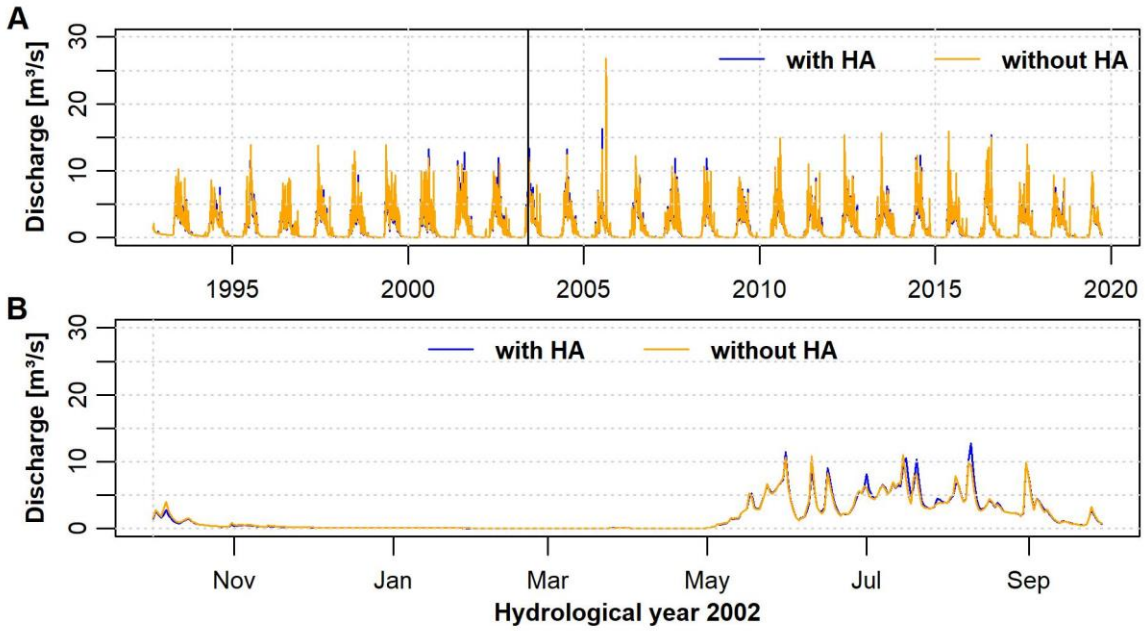
Kühtai	5	R	1 (1)	47.2071	11.0059	1918	1990-2019	TIWAG
Kühtai	5	T	1 (1)	47.2071	11.0059	1918	1990-2019	TIWAG
Kühtai	5	WS	2 (1)	47.2071	11.0059	1918	1990-2019	TIWAG
Lampenspitze Schneestation	6	H	1 (1)	47.153	11.121	2111	2008-2019	LWD
Lampenspitze Schneestation	6	R	1 (1)	47.153	11.121	2111	2008-2019	LWD
Lampenspitze Schneestation	6	T	1 (1)	47.153	11.121	2111	2011-2019	LWD
Lampenspitze Windstation	7	H	2 (1)	47.156	11.096	2870	2008-2019	LWD
Lampenspitze Windstation	7	T	1 (1)	47.156	11.096	2870	2011-2019	LWD
Lampenspitze Windstation	7	WS	2 (1)	47.156	11.096	2870	2008-2019	LWD
Langenferner-Felsköpfl	8	H	1 (0)	46.4725	10.6139	2967	2015-2019	ACINN
Langenferner-Felsköpfl	8	P	1 (0)	46.4725	10.6139	2967	2015-2019	ACINN
Langenferner-Felsköpfl	8	R	1 (0)	46.4725	10.6139	2967	2015-2019	ACINN
Langenferner-Felsköpfl	8	T	1 (0)	46.4725	10.6139	2967	2015-2019	ACINN
Langenferner-Felsköpfl	8	WS	1 (0)	46.4725	10.6139	2967	2015-2019	ACINN
Leiter	9	P	1 (1)	47.1228	10.9691	1564	1976-2019	TIWAG

Leiter	9	T	5 (2)	47.1228	10.9691	1564	1988-2019	TIWAG
Leiter	9	HS	1 (1)	47.1228	10.9691	1564	1976-2019	TIWAG
Malga Mare	10	P	1 (1)	46.4142	10.68	1950	1930-1984	Trentino
Niederthai	11	Q	1 (1)	47.1577	11.0125	1478	1985-2018	TIWAG
Rossbänke	12	H	1 (1)	46.4693	10.8194	2255	2022-2020	Bozen
Rossbänke	12	T	1 (1)	46.4693	10.8194	2255	2022-2020	Bozen
Rossbänke	12	WS	3 (1)	46.4693	10.8194	2255	2022-2020	Bozen
Schöntaufspitze	13	T	2 (1)	46.5029	10.6286	3328	2003-2020	Bozen
Schöntaufspitze	13	WS	4 (1)	46.5029	10.6286	3328	2003-2020	Bozen
Stausee Zufritt	14	P	1 (1)	46.509	10.7251	1851	1972-2019	Bozen
Stausee Zufritt	14	T	1 (1)	46.509	10.7251	1851	1973-2019	Bozen
Sulden	15	H	2 (1)	46.5159	10.5953	1907	2010-2020	Bozen
Sulden	15	P	5 (1)	46.5159	10.5953	1907	2003-2020	Bozen
Sulden	15	R	2 (0)	46.5159	10.5953	1907	2010-2020	Bozen
Sulden	15	T	3 (2)	46.5159	10.5953	1907	2003-2020	Bozen
Sulden Madritsch	16	H	4 (1)	46.4938	10.6144	2825	2000-2020	Bozen
Sulden Madritsch	16	HS	6 (1)	46.4938	10.6144	2825	2000-2020	Bozen
Sulden Madritsch	16	T	1 (1)	46.4938	10.6144	2825	2003-2020	Bozen
Sulden Madritsch	16	WS	1 (1)	46.4938	10.6144	2825	2008-2020	Bozen
Ulten Weißbrunn	17	H	1 (1)	46.4868	10.8318	1900	2010-2020	Bozen

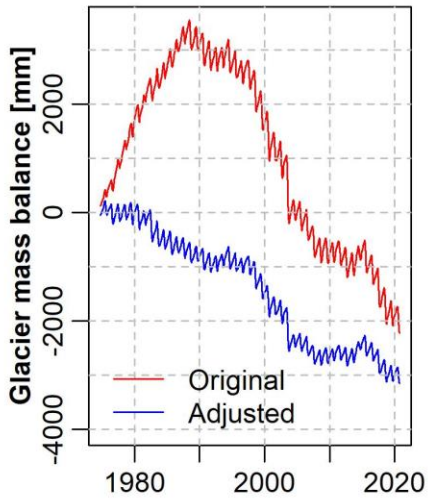
Ulten Weißbrunn	17	P	4 (1)	46.4868	10.8318	1900	2003-2020	Bozen
Ulten Weißbrunn	17	R	1 (0)	46.4868	10.8318	1900	2010-2020	Bozen
Ulten Weißbrunn	17	T	4 (3)	46.4868	10.8318	1900	1987-2020	Bozen
Umhausen	18	H	1 (1)	47.1392	10.9289	1025	1946-2019	ZAMG
Umhausen	18	P	1 (1)	47.1392	10.9289	1025	1946-2019	ZAMG
Umhausen	18	R	1 (1)	47.1392	10.9289	1025	2011-2019	ZAMG
Umhausen	18	T	2(2)	47.1392	10.9289	1025	1946-2019	ZAMG
Umhausen	18	WS	1 (1)	47.1392	10.9289	1025	1946-2019	ZAMG
Weißbrunnspitze	19	H	1 (1)	46.494	10.774	3253	2012-2020	Bozen
Weißbrunnspitze	19	T	1 (1)	46.494	10.774	3253	2012-2020	Bozen
Weißbrunnspitze	19	WS	3 (1)	46.494	10.774	3253	2012-2020	Bozen
Zufallhuette	20	Q	1 (1)	46.4763	10.6768	2311	2014-2020	Bozen



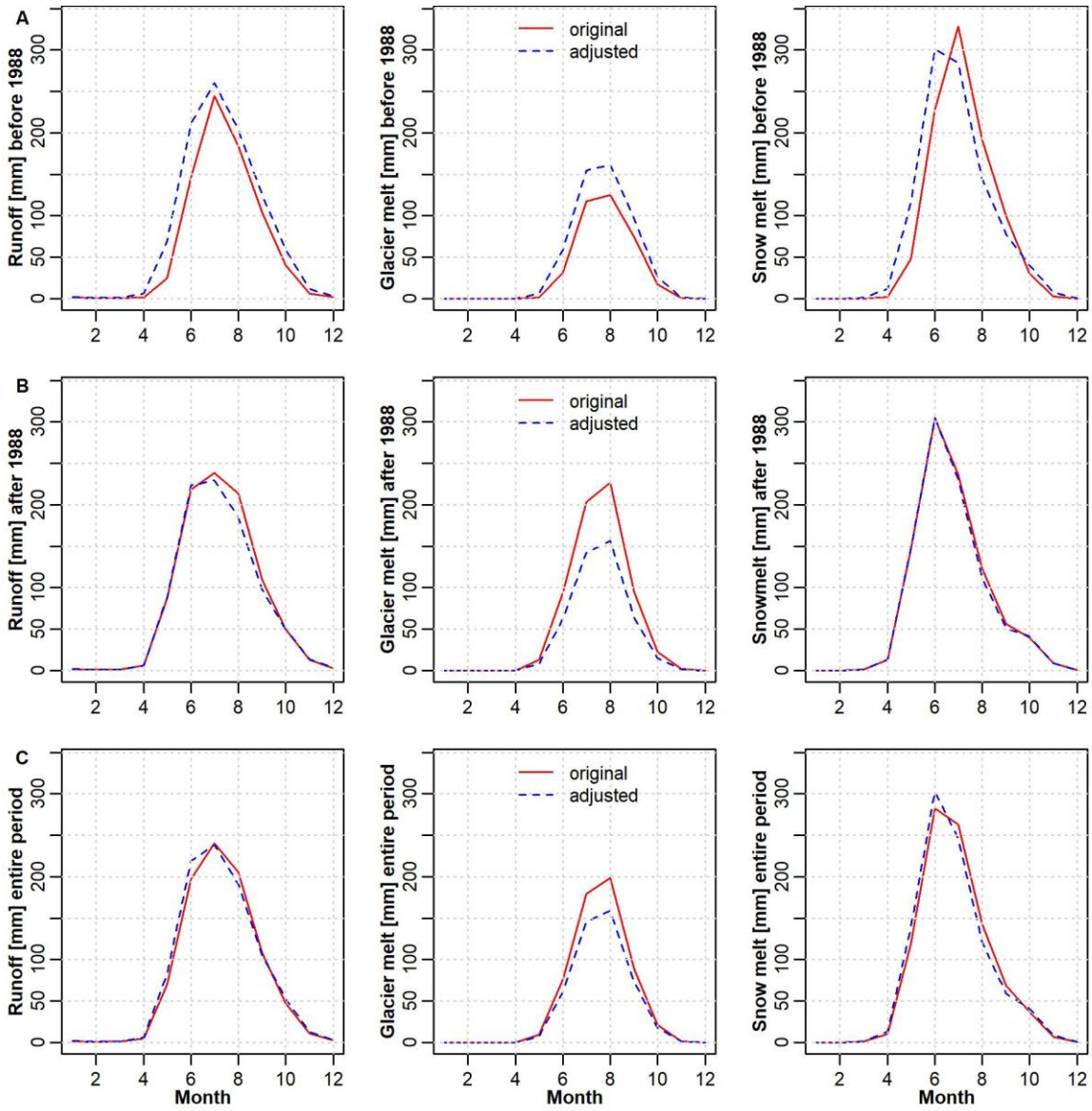
A.1.2: (A) Simulated daily positive (green) and negative (purple) precipitation anomaly through consideration of Horlachalm precipitation and (B) simulated snow water equivalent with (blue) and without (orange) considering Horlachalm precipitation. The vertical black line shows a significant change point detected by the Bernier and by the Pettitt tests.



A.1.3: (A) simulated daily discharge at the catchment outlet with (blue) and without (orange) considering Horlachalm precipitation and (B) detailed snap shot of the hydrological year 2002. The vertical black line shows a significant change point detected by the Bernier and by the Pettitt tests.



A.1.4: Simulated glacier mass balance using the original (red line) or the adjusted (blue line) temperature time series for the Zutritt station.



A.1.5: Simulated mean monthly runoff, glacier and snowmelt (A) before the change point in 1988, (B) after the change point and (C) over the entire period (1974-2020).

A.2. Supplementary material to chapter 3

Hofmeister, Rubens Venegas, et al., 2023[6]

The uncertainty analysis of the AutoSalt system is based on standard formula for error propagation proposed by the general framework of the Guide to the expression of Uncertainty in Measurement, GUM (JCGM, 2008). The relative uncertainty in discharge computed from an injection of dry salt can be expressed as (Sentlinger et al., 2019):

$$(\delta Q)^2 = \left(\frac{\partial Q}{\partial M}\right)^2 (\delta M)^2 + \left(\frac{\partial Q}{\partial \sum[EC(t) - EC_{BG}]}\right)^2 (\delta \sum[EC(t) - EC_{BG}])^2 + \left(\frac{\partial Q}{\partial dt}\right)^2 (\delta dt)^2 + \left(\frac{\partial Q}{\partial CF_T}\right)^2 (\delta CF_T)^2$$

A.2.1

where δQ represents the relative error in Q, which results from the uncertainty in the injected mass δM, followed by the uncertainty in the measurement of the difference between EC(t) and background conductivity EC_{BG}, the measurement time δdt and the uncertainty in the constant δCF_T. The uncertainty in time is essentially zero and can be ignored. The uncertainty in the mass of salt injected has two uncertainty components: 1) the brine concentration (salt content) and 2) the volume of the brine. The brine concentration, which usually has a concentration of 20% salt (NaCl) content, is measured with an optical refractometer at each site visit. The accuracy of brine volume depends on the accuracy of the flow meter. With a well-calibrated configuration, the uncertainty associated with the salt mass is < 4 %. The noise in the measurement signal that is represented by the standard error of the mean of EC_{BG} is computed with the following equation (Sentlinger et al., 2019):

$$\delta(EC(t) - EC_{BG}) = \frac{\sigma}{\sqrt{n}}$$

A.2.2

where σ is the standard deviation of all EC_{BG} measurements (pre- and post-tracer injection) and n is the total sample size. Ten samples before and after the wave should reduce the estimate of the mean EC_{BG} to 0.14 % of the sample σ. 30 should reduce the noise to 0.03 %. The larger n is, the lower the uncertainty in the derived discharge until the stream's natural drifting begins to increase σ. When the noise is smaller than the sensor resolution, half of the sensor resolution should be used as uncertainty. Another uncertainty source is associated with the constant CF_T (Sentlinger et al., 2019), which was explored in detail by Richardson, Sentlinger, *et al.* (2017). Since all T-HRECS probes are properly calibrated, the uncertainty

[6]Material from:
Hofmeister, F., Rubens Venegas, B., Sentlinger, G., Tarantik, M., Blume, T., Disse, M., & Chiogna, G. (2023). Automated streamflow measurements in high-elevation Alpine catchments. River Research and Applications, rra.4203. doi: 10.1002/rra.4203.

associated with CF_T is limited to 2.8 %. Fathom Scientific Ltd provides detailed documentation about T-HRECS calibration on its website (Fathom Scientific Ltd., 2020b). Uncertainty resulting from incomplete mixing of the tracer is not included in equation A.2.1, but considered in an indirect way by comparing the uncertainty of estimated Q from the left and right bank following equation A.2.3 (Sentlinger et al., 2019):

$$\frac{\delta \bar{Q}}{\bar{Q}} = max\left[QUnc_{LB}, QUnc_{RB}, \frac{abs(Q_{LB}-Q_{RB})}{\bar{Q}}\right] \quad \text{A.2.3}$$

where $\bar{Q}$ is the average of the left and right bank derived Q, Q_{LB}, and Q_{RB} respectively. $QUnc_{LB}$ and $QUnc_{RB}$ are the independent uncertainty estimates for each bank derived with equation A.2.1. If the difference in the derived discharges is greater than the sum of independent estimates of uncertainty for each bank then incomplete mixing is assumed and the measurement reach length should be adjusted (Sentlinger et al., 2019).

We estimated the uncertainty in the Ott and AQA water level measurements by considering the sensor accuracy and the mean hourly standard deviation of the water level measurement during steady flow periods, which indicates the variability (i.e., noise) in the stage records. The equation for calculating the relative accuracy of the stage measurements is:

$$Relative\ stage\ accuracy = \frac{\mu(\sigma\_hourly(stage))}{\mu(stage)} * (100 + sensor\ accurancy) \quad \text{A.2.4}$$

μ is a the arithmetic mean, σ is the hourly standard deviation of the 15-min stage time series, and the sensor accuracy is 0.05 % for the OTT and 0.1 % for the AutoSalt. We computed the relative stage accuracy for one respective measurement period and listed the results in Table A.2.5.

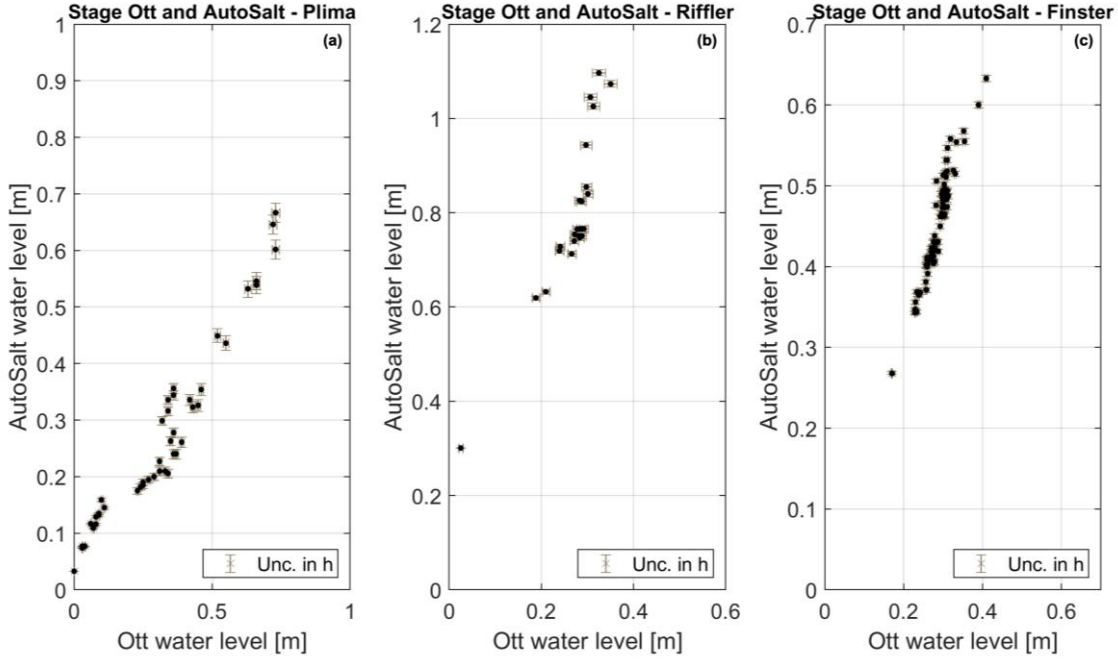
A.2.5: Relative stage accuracy at each experimental site computed with equation A.2.4.

Site	**Relative stage uncertainty Ott* [%]**	**Relative stage uncertainty AutoSalt** [%]**	**Period**
Lower Plima	2.32	2.2	2020
Riffler	2	1.26	2021
Finstertal	1	0.57	2020

*including the sensor accuracy of 0.05%

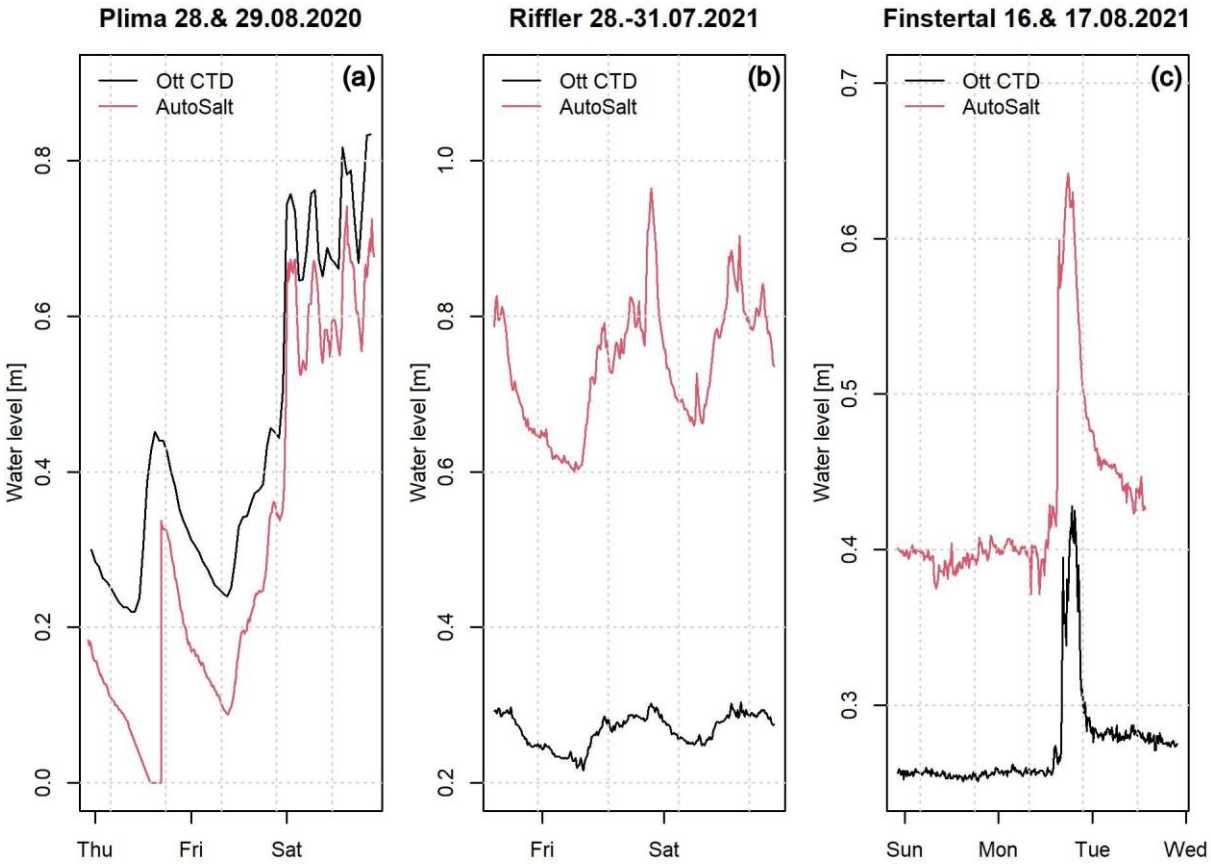
**including the sensor accuracy of 0.1%

Since the AutoSalt pressure sensor at the site of the injection and the downstream Ott CTD sensor both record the water level at the same 15-minute time interval, we compared the relationship between the two for all three experimental sites Figure A.2.6. The distance between the AutoSalt and the Ott CTD sensor is 181 m at the Plima, 51 m at Rifflerbach, and 41 m at Finstertal. Thus, the properties of the stream cross-section (e.g., geometry) of the respective gauging sites differ. In general, Figure A.2.6 shows a good agreement between the stage records at the different river sections. The rank correlation between the stage records is lowest for Plima and Rifflerbach (Kendall tau of 0.72 and 0.74), whereas the experimental site at Finstertal has the highest correlation with a Kendall tau of 0.81.



A.2.6: Stage records and sensor error that corresponds the sensor accuracy of ±0.05 % at the Ott and AutoSalt location in 15-minute time interval at each site (a) Plima, (b) Rifflerbach and (c) Finstertalbach.

Figure A.2.7 compares the water level records of the two sensors (i.e., Ott CTD and AutoSalt) during the highest event in the corresponding observation period. The noise in the stage signal is particularly evident at the Finstertalbach, which has the highest streambed gradient (i.e., 18 %) among the three creeks. The other two streams show little indication of noise in the two stage records.



A.2.7: Stage records of Ott (black line) and AutoSalt (red line) at each site (a) Plima, (b) Rifflerbach and (c) Finstertalbach during the highest discharge event in the respective observation period.

The lower part of the aluminium pipes, which protects and holds the sensors of the Ott CTDs at the same height, has a few holes to allow a constant water exchange. To prevent the pipe and the sensor from being filled up with sediments, we installed the pipe approx. 5 cm above the riverbed. This ensures that no sediment is deposited and affects the measurements (stage, temperature, and EC_T). We check the sensors at each site visit, which is about once a month during the summer period. This includes cleaning the pressure capillary and resetting the pressure level to zero if necessary. In addition, the sensor time is checked and adjusted if necessary. Once a year, we replace the lithium batteries. In case of increased wave action, we set the measuring interval of the pressure measurement to a few minutes (approx. 3 minutes) to reduce the noise in the stage signal.

The AutoSalt performs the measurements on the falling limb of the hydrograph. Although the stage is generally more stable on the falling limb, we cannot exclude that the high streamflow measurements are affected by the hysteresis effect, especially if there are multiple peaks (Martínez-Carreras et al., 2016). The hysteresis effect is caused by the fact that in the case of an approaching flood, the flow in the rising limb of a flood wave is greater than in the falling

limb at the same water level as a result of the greater water level gradient (Morgenschweis, 2018). According to the Jones (1916) formula, the hysteresis effect increases with a higher stage gradient and also with a lower bed slope. Therefore, it seems unlikely that hysteresis is strong in these steep Alpine streams. Nevertheless, the hysteresis effect can be checked based on the Jones (1916) formula as recently done by Perret, Lang, & Le Coz (2022).

A.3. Supplementary material to chapter 4

Hofmeister, Arias-Rodriguez, et al., 2022[7]

A.3.1: Spectral bands and resolutions of the Sentinel-2A MultiSpectral Instrument (MSI) (Satellite Imaging Corporation, 2021).

Sentine-2 Bands	Central Wavelength (µm)	Resolution (m)
Band 1 - Costal aerosol	0.443	60
Band 2 - Blue	0.490	10
Band 3 - Green	0.560	10
Band 4 - Red	0.665	10
Band 5 - Vegetation Red Edge	0.705	20
Band 6 - Vegetation Red Edge	0.740	20
Band 7 - Vegetation Red Edge	0.783	20
Band 8 - near-infrared	0.842	10
Band 8A - Vegetation Red Edge	0.865	20
Band 9 - Water vapour	0.945	60
Band 10 - short-wave infrared - Cirrus	1.375	60
Band 11 - short-wave infrared	1.610	20
Band 12 - short-wave infrared	2.190	20

The energy balance approach considers all energy fluxes into and out of the snow pack using equation A.3.2:

$$Q + H + E + A + G + Mae = 0 \quad \text{A.3.2}$$

where *Q* is the short-wave and long-wave radiation balance [W/m²], *H* is the sensible heat flux [W/m²], *E* is the latent heat flux [W/m²], *A* is the advective heat flux by precipitation [W/m²], *G* is the ground heat flux and set to constant 2 W/m², and M_{ae} is the energy available for melting and/or sublimation.

The fraction of snow on the total precipitation is calculated after equation A.3.3:

$$Psnow = \frac{T0R+Ttrans-T}{2\times Ttrans} \text{ for } (T0R\text{-}Ttrans) < T < (T0r + Ttrans) \quad \text{A.3.3}$$

where *Psnow* is the fraction of snow on the total precipitation (range 0 to 1), *T* is the air temperature in °C, *T0R* is the temperature at which 50 % of the precipitation falls as snow [°C] and *Ttrans* is half of the temperature-transition range from snow to rain [°C].

[7]Material from:
Hofmeister, F**.,** Arias-Rodriguez, L. F., Premier, V., Marin, C., Notarnicola, C., Disse, M., & Chiogna, G. (2022). Intercomparison of Sentinel-2 and modelled snow cover maps in a high-elevation Alpine catchment. *Journal of Hydrology X*, *15*, 100123. doi: 10.1016/j.hydroa.2022.100123.

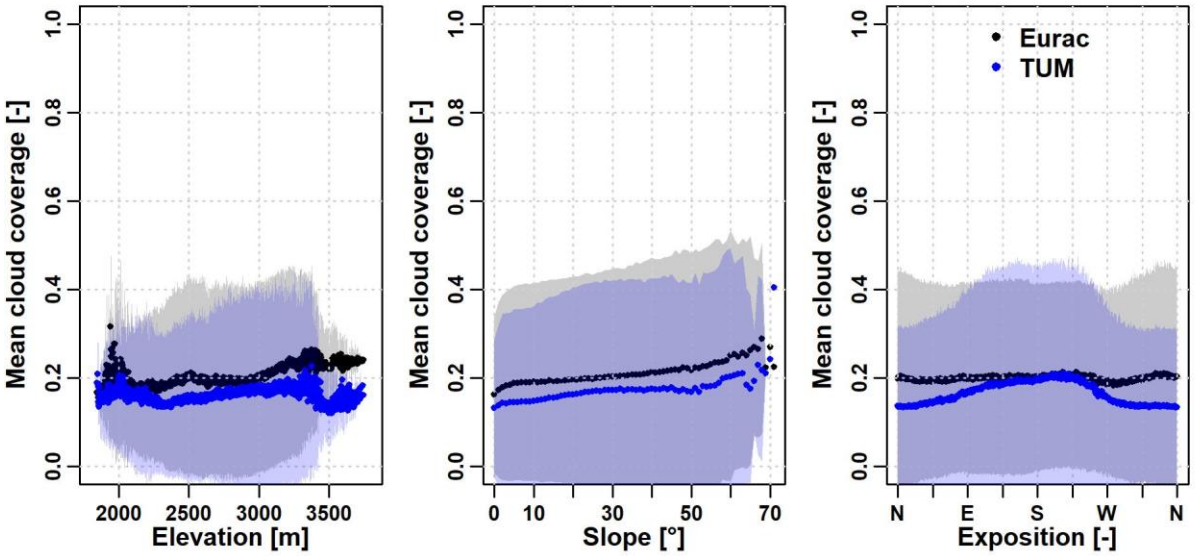
The only parameters of the energy balance to be calibrated are the correction factors for incoming and outgoing long-wave radiation in equations A.3.4 and A.3.5:

$$LWin = \sigma \times LWINcorr \times Tair^4 \times \left(0.61 + 0.05\sqrt{e}\right) \times (1 + cb \times CL^2) \quad \text{A.3.4}$$

where σ is the Stefan-Boltzmann constant (0.567×10^{-8} W/(m²K⁴)), *LWINcorr* is the correction factor for incoming long-wave radiation, *Tair* is the air temperature [K], *e* is the vapour pressure [Pa], *cb* is the factor for accounting for cloudiness (0.24 if precipitation; 0.08 for cloudiness < 0.2; 0.17 for cloudiness between 0.2 and 0.6, 0.2 for cloudiness > 0.6) and *CL* is the cloudiness which is calculated in the evaporation model depending on global radiation. The outgoing long-wave radiation is calculated by equation A.3.5:

$$LWout = -\sigma \times LWOUTcorr \times Tsurf^4 \quad \text{A.3.5}$$

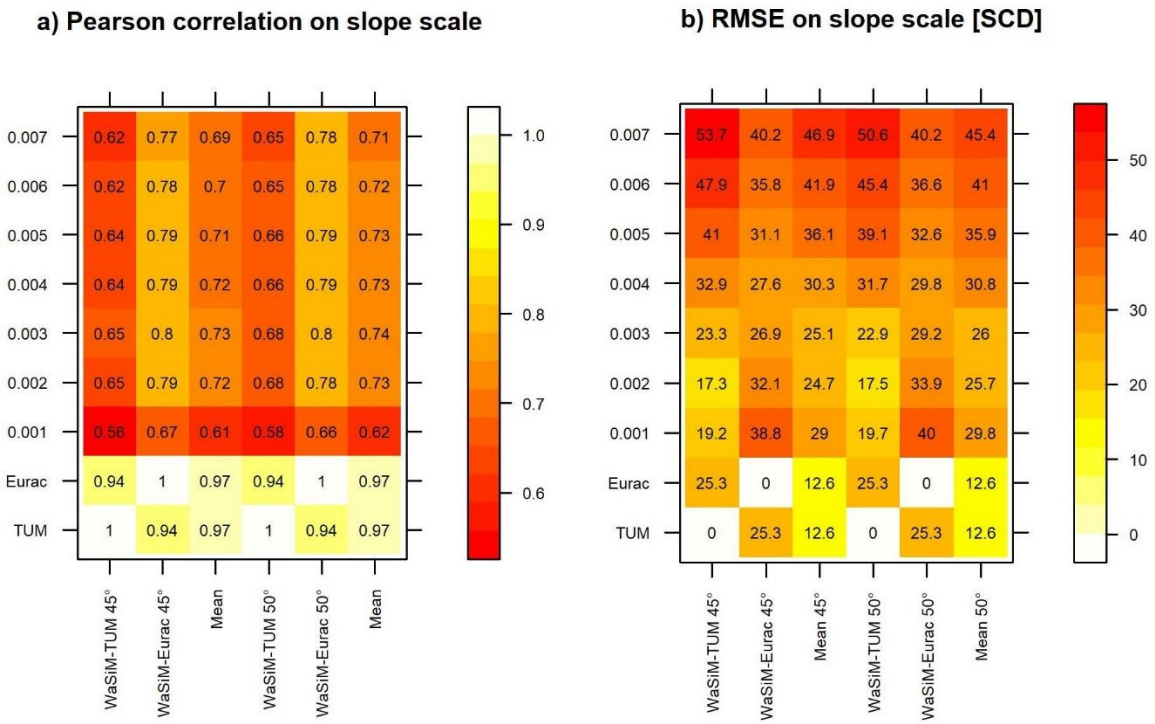
where σ is the Stefan-Boltzmann constant (0.567×10^{-8} W/(m²K⁴)), *LWOUTcorr* is the correction factor for outgoing long-wave radiation, *Tsurf* is the surface temperature [K] (273.16 K where *Tair* ≥ 0 and snow pack temperature where *Tair* < 0).



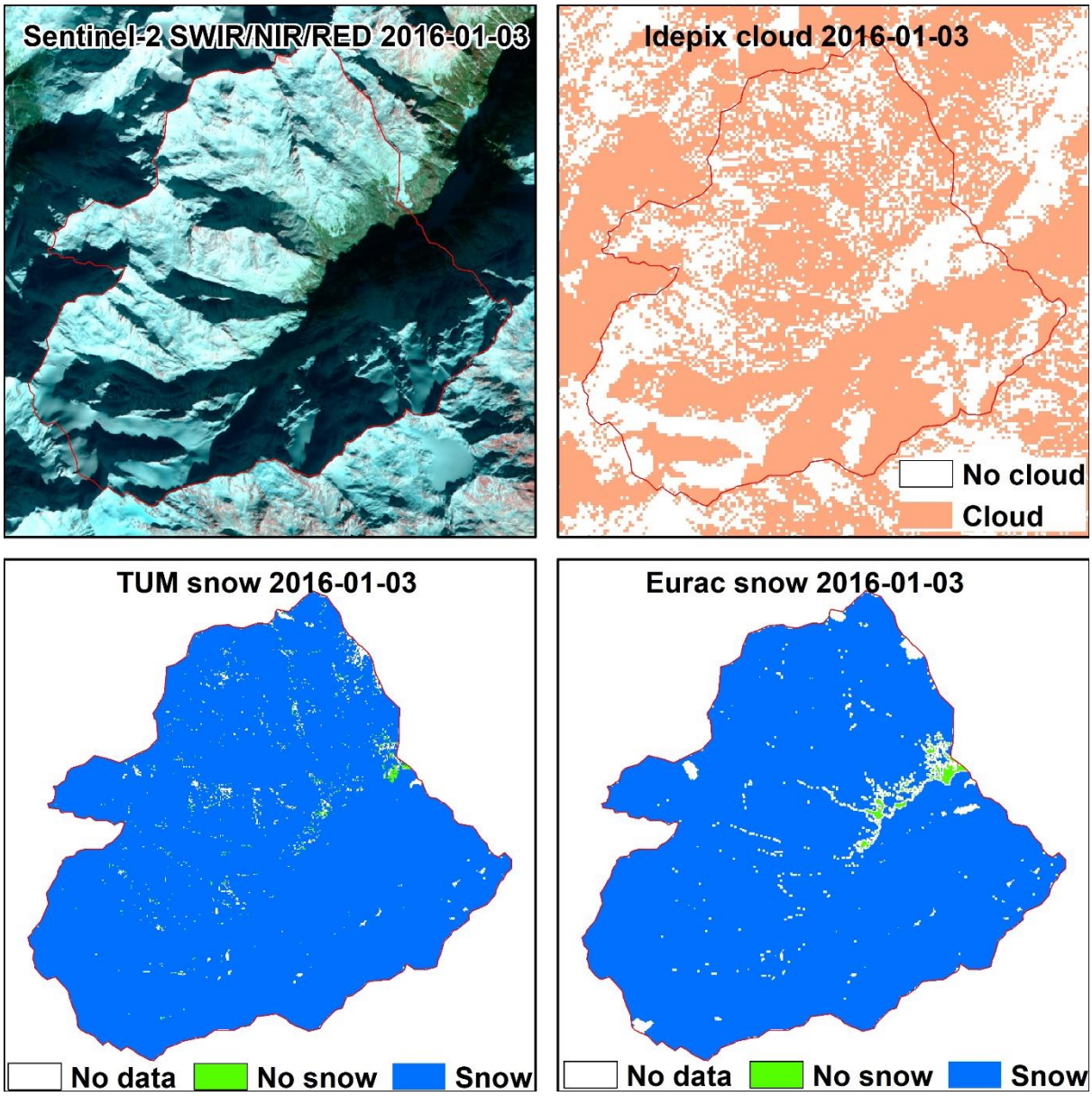
A.3.6: Cloud detection by TUM and Eurac against topographical features (elevation in one meter steps, slope and exposition in one degree steps) plus mean standard deviation as shaded areas.

A.3.7: Snow detection accuracy of TUM and Eurac at the three snow observation stations and for different snow depth thresholds.

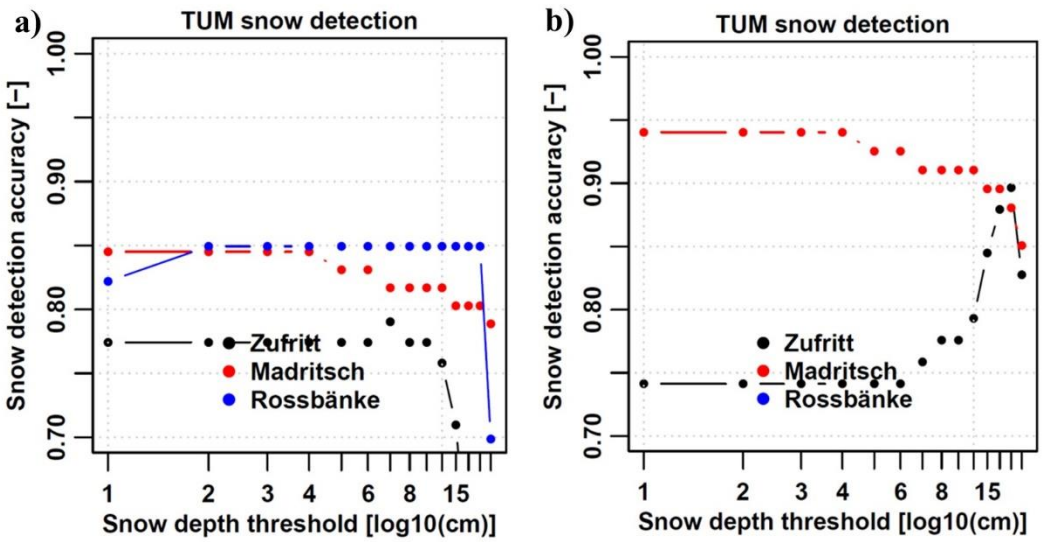
Thres-hold	TUM Zufritt	Eurac Zufritt	TUM Madritsch	Eurac Madritsch	TUM Ross-baenke	Eurac Ross-baenke	TUM mean	Eurac mean
1 cm	0.92	0.83	0.96	0.97	0.86	0.84	0.91	0.88
2 cm	0.92	0.83	0.95	0.97	0.88	0.87	0.92	0.89
3 cm	0.92	0.83	0.94	0.96	0.96	0.96	0.94	0.92
4 cm	0.92	0.83	0.94	0.96	0.98	0.99	0.95	0.92
5 cm	0.93	0.84	0.93	0.94	0.98	0.98	0.94	0.92
6 cm	0.93	0.84	0.93	0.94	0.98	0.98	0.94	0.92
7 cm	0.94	0.84	0.92	0.94	0.98	0.98	0.95	0.92
8 cm	0.93	0.85	0.92	0.94	0.98	0.98	0.94	0.92
9 cm	0.93	0.85	0.91	0.93	0.97	0.98	0.94	0.92
10 cm	0.92	0.85	0.91	0.93	0.97	0.98	0.94	0.92
15 cm	0.91	0.84	0.89	0.91	0.97	0.98	0.93	0.91
20 cm	0.89	0.84	0.88	0.9	0.95	0.95	0.91	0.9
30 cm	0.89	0.87	0.83	0.85	0.9	0.91	0.87	0.88
50 cm	0.87	0.86	0.75	0.76	0.81	0.83	0.81	0.82



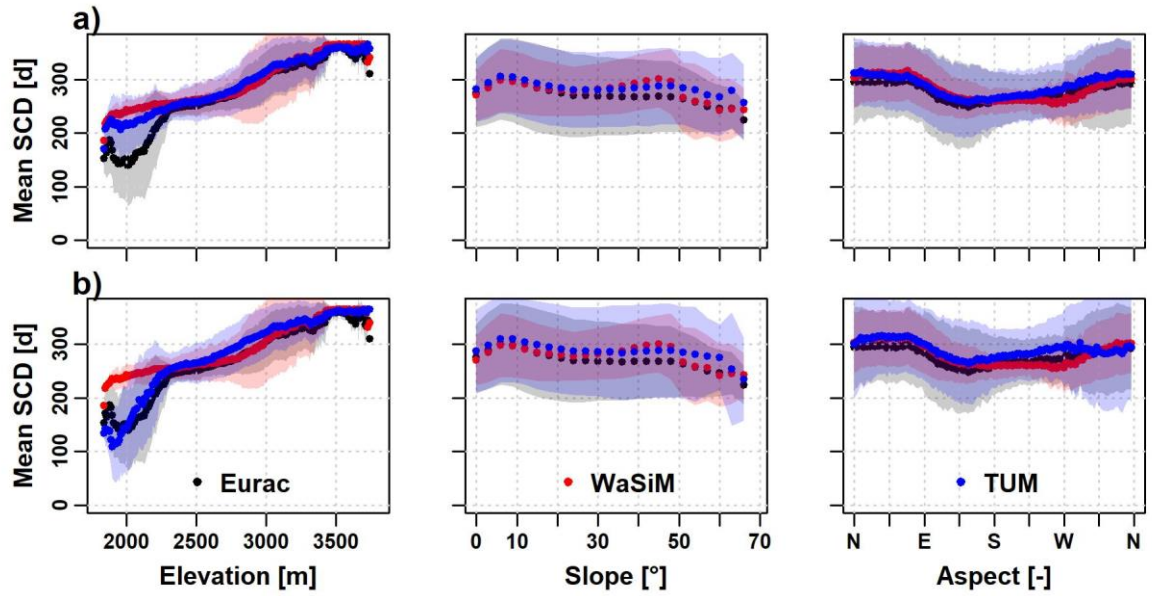
A.3.8: Heatmap (a) Pearson correlation and (b) Root Mean Square Error (RMSE) of optimized *ferosion* parameter on the slope scale for each observation product (TUM and Eurac) and as an average of both products as well as for different *ierosion* values (45° and 50°).



A.3.9: Sentinel-2 false colour image (SWIR/NIR/RED) of 2016-01-03 in 20 m resolution, Idepix cloud detection (60 m) and TUM and Eurac snow product (25 m) for the said date. No data pixels were considered as cloud or shadow by the algorithms.



A.3.10: Accuracy of TUM snow detection at in-situ snow observation stations without atmospheric correction (a) and with atmospheric correction (b) for the hydrological year 2019.



A.3.11: Mean snow cover duration over changes in elevation, slope and aspect without atmospherically corrected TUM product (a) and with atmospheric correction (b) for the hydrological year 2019.