Technische Universität München
TUM School of Engineering and Design

Alpine Hydrology under Different Climate Model Generations

Timo Schaffhauser

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. Niklas Boers

Prüfende der Dissertation:

1. Prof. Dr.-Ing. Markus Disse
2. Prof. Dr. Martin Volk
3. Prof. Dr. Nicola Fohrer

Die Dissertation wurde am 07.10.2024 bei der Technischen Universität München eingereicht und durch die TUM School of Engineering and Design am 14.01.2025 angenommen.

Affidavit

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.

Place and date:

Munich, September 15, 2024

Name and signature

Timo Schaffhauser

Abstract

The impacts of climate change on the hydrology in mountainous and glacierized regions is increasingly important to understand. An investigation for which suitable and reliable tools are indispensable. Therefore, continuous validation and improvement of these instruments is necessary for accurate hydrological representations. This dissertation investigates the impacts of climate change on hydrological processes in mountainous and glacierized catchments, utilizing and enhancing the Soil and Water Assessment Tool (SWAT).

First, multiple climate projections from two generations of General Circulation Models (GCMs) based on the The Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) are assessed in the Naryn Basin, the headwater of the famous Syr Darya located in Central Asia. Hence, climate projections stemming from the ISIMIP phases two and three are used to investigate climate impacts on the basin's hydrology. The two ISIMIP phases represent the two most recent stages of the Coupled Model Intercomparison Project that are widely used in the field (CMIP5 and CMIP6, respectively). The comparison and evaluation focuses on the low emission and high emission scenarios available in both generations, which correspond to RCP2.6 & SSP1-2.6, as well as RCP8.5 & SSP5-8.5. The comparative analysis covers multiple hydrological components, such as discharge, evapotranspiration (ETA), and soil moisture (SM). Besides, multiple aspects of uncertainties, inherent in climate impact studies, are analyzed. It is revealed that higher climate sensitivities, assigned to recent GCMs, can cause significantly differing hydrological projections due to elevated temperature projections. It is further demonstrated that peak discharge in the basin could occur around one month earlier. Increasing ETA alongside substantial decreases in soil moisture are expected to be additional hydrological changes. Under CMIP6, significant precipitation increases, not present in former projections, slightly buffer water-limited conditions, albeit soil moisture deficits are not sufficiently alleviated and water limitation could generally exacerbate.

Recognizing SWAT's limitations in representing glacier dynamics in high-mountainous regions, SWAT-GL is developed to extend SWAT and address current limitations. SWAT-GL introduces a novel mass balance module and glacier evolution routine based on the delta-h parameterization. The model development contributes to a better representation of high mountain hydrology and to improve the model's applicability in these environments. The extended model is particularly useful in data scarce regions as it involves a minimum of data requirements. Glacier changes can be simulated dynamically and includes a spatially-distributed representation of ice thickness changes. SWAT-GL and its model code is open-source, documented and easily accessible by the community to follow the FAIR (Findable, Accessible, Interoperable, and Reusable) principles. Initial tests in the Martelltal in Italy, where it successfully represents mass balance variations, serve as promising results.

To further validate SWAT-GL, a comprehensive evaluation study is carried out using four USGS (United States Geological Survey) Benchmark Glaciers. The evaluation relies fully on glaciological components and includes glacier mass balance, hypsometry and snow cover rather than discharge as in other studies.

The procedure allows to test the complex interactions of the new routine with the pre-existing snow routine, which is examined in detail by a sensitivity analysis using the Method of Morris. The work highlights the strong interactions between snow and glacier components, suggesting the need for simultaneous calibration of these processes. While the exclusion of discharge in the optimization process demonstrates SWAT-GL's capabilities in simulating glacier and snow processes, a further investigation shows how discharge is adequately reproduced calibrated when incorporated in the optimization procedure.
This dissertation advances hydrological modeling by providing hydrological projections under climate change and integrating glacier processes into SWAT, offering a sophisticated tool to represent the hydrology in complex, glacierized environments. These contributions are critical for better understanding water resource management and climate adaptation strategies in vulnerable mountainous regions.

Zusammenfassung

Die Auswirkungen des Klimawandels auf die Hydrologie in Gebirgs- und Gletscherregionen zu verstehen wird immer wichtiger. Eine Untersuchung, für die geeignete und zuverlässige Wekrzeuge unerlässlich sind. Daher ist eine kontinuierliche Validierung und Verbesserung dieser Werkzeuge für genaue hydrologische Darstellungen erforderlich. In dieser Dissertation werden die Auswirkungen des Klimawandels auf hydrologische Prozesse in gebirgigen und vergletscherten Einzugsgebieten untersucht, wobei das Soil and Water Assessment Tool (SWAT) verwendet und erweitert wird.
Zunächst werden mehrere Klimaprojektionen von zwei Generationen von globalen Klimamodellen (GCMs) auf der Grundlage des Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) im Naryn-Becken, dem Oberlauf des berühmten Syr Darya in Zentralasien, bewertet. Daher werden die Klimaprojektionen der zweiten und dritten ISIMIP-Phase verwendet, um die Klimaauswirkungen auf die Hydrologie des Beckens zu untersuchen. Die beiden ISIMIP-Phasen stellen die beiden jüngsten Generationen des Coupled Model Intercomparison Project dar, die in der Praxis weit verbreitet sind (CMIP5 bzw. CMIP6). Der Vergleich und die Bewertung konzentrieren sich auf die in beiden Generationen verfügbaren Szenarien mit niedrigen und hohen Emissionen, die dem RCP2.6 & SSP1-2.6 sowie dem RCP8.5 & SSP5-8.5 entsprechen. Die vergleichende Analyse umfasst mehrere hydrologische Komponenten, wie Abfluss, Evapotranspiration (ETA) und Bodenfeuchte (SM). Außerdem werden mehrere Aspekte der Unsicherheiten, die in Studien über Klimaauswirkungen zwangsläufig auftreten, analysiert. Es zeigt sich, dass die höheren Klimasensitivitäten, die den neueren GCMs zugewiesen werden, aufgrund der höheren Temperaturprognosen zu erheblich abweichenden hydrologischen Projektionen führen können. Ferner wird gezeigt, dass der Spitzenabfluss im Einzugsgebiet etwa einen Monat früher auftreten könnte. Als weitere hydrologische Veränderungen werden ein Anstieg der ETA und eine erhebliche Abnahme der Bodenfeuchte erwartet. Im Rahmen von CMIP6 federn deutliche Niederschlagszunahmen, was in den früheren Projektionen nicht der Fall war, die wasserbegrenzten Bedingungen leicht ab, wenngleich die Bodenfeuchtigkeitsdefizite nicht ausreichend gemildert werden und sich die Wasserbegrenzung im Allgemeinen verschärfen könnte.
Da SWAT bei der Darstellung der Gletscherdynamik in Hochgebirgsregionen an seine Grenzen stößt, wurde SWAT-GL entwickelt, um SWAT zu erweitern und derzeitige Einschränkungen zu überwinden. SWAT-GL führt ein neues Massenbilanzmodul und eine Gletscherentwicklungsroutine ein, die auf der delta-h-Parametrisierung basiert. Die Entwicklung des Modells trägt zu einer besseren Darstellung der Hochgebirgshydrologie sowie zu einer verbesserten Anwendbarkeit des Modells in diesen Gebieten bei. Das erweiterte Modell ist besonders in datenarmen Regionen nützlich, da es ein Mindestmaß an Daten benötigt. Gletscherveränderungen können dynamisch simuliert werden und beinhalten eine räumlich verteilte Darstellung von Eisdickenänderungen. SWAT-GL und sein Modellcode sind frei, dokumentiert und für die Gemeinschaft leicht zugänglich, um den FAIR-Prinzipien (Findable, Accessible, Interoperable, and Reusable) zu entsprechen. Erste Tests im Martelltal in Italien, wo es erfolgreich Massenbilanzvariationen abbildet, zeigen vielversprechende Ergebnisse.

Zur weiteren Validierung von SWAT-GL wird eine umfassende Evaluierungsstudie anhand vier USGS (United States Geological Survey) Benchmark-Gletschern durchgeführt. Die Bewertung stützt sich vollständig auf glaziologische Komponenten und umfasst die Massenbilanz des Gletschers, die Hypsometrie sowie die Schneebedeckung und basiert im Gegensatz zu anderen Studien nicht auf dem Abfluss. Das Verfahren ermöglicht es, die komplexen Wechselwirkungen zwischen der neuen Routine und der bereits bestehenden Schneeroutine zu testen, was durch eine Sensitivitätsanalyse basierend auf der Morris-Methode untersucht wird. Die Arbeit verdeutlicht die starken Wechselwirkungen zwischen Schnee- und Gletscherkomponenten, was die Notwendigkeit einer gleichzeitigen Kalibrierung dieser Prozesse nahelegt. Während die Nichtberücksichtigung des Abflusses im Optimierungsprozess die Fähigkeiten von SWAT-GL in der Simulation von Gletscher- und Schneeprozessen demonstriert, zeigt eine weitere Untersuchung, dass der Abfluss adäquat widergespiegelt werden kann, wenn er in das Optimierungsverfahren einbezogen wird.

Diese Dissertation bringt die hydrologische Modellierung voran, indem sie hydrologische Projektionen unter dem Einfluss des Klimawandels liefert und Gletscherprozesse in SWAT integriert, wodurch ein hochentwickeltes Werkzeug zur Darstellung der Hydrologie in komplexen, vergletscherten Umgebungen zur Verfügung gestellt wird. Diese Beiträge sind entscheidend für ein besseres Verständnis des Wasserressourcenmanagements und der Klimaanpassungsstrategien in gefährdeten Bergregionen.

Acknowledgements

I would like to extend my deepest gratitude to all those who have supported me throughout the course of this dissertation. However, first of all I want to thank the Chair of Hydrology and River Basin Management at the Technical University of Munich and all its employees for hosting, supporting and enduring me, and for providing a great atmoshpere and research place. I am more than grateful that I now have the opportunity to foremost thank my doctoral supervisor, Prof. Dr.-Ing. Markus Disse, for giving me this opportunity to do my doctorate, which I could not take for granted. He not only made it possible for me to participate in great projects and to be part of his worldwide network, but he was not afraid to give me a lot of responsibility in these projects right from the start, which tremendously strengthened my personal development, not only at the technical level of hydrology.
I am also deeply grateful to my second supervisor, Prof. Dr. Martin Volk, who always showed great interest in my work and my progress and had valuable input, both professionally and personally, in every exchange. In particular, his past work in a similar field made every exchange a valuable enrichment and I very much appreciate his visits to Munich, despite the enormous time pressure, where he always took a lot of time for bilateral conversations. This also applies for my third supervisor, Prof. Dr. Nicola Fohrer, who did not hesitate in agreeing to take this role and gave me the unique opportunity to give a guest lecture at her one of courses at her department in Kiel to show my research. An experience for which I am grateful. Moreover, I feel honored and want to thank Prof. Dr. Niklas Boers for taking the role as chairman during my PhD defense.
Two people who were also of great importance for this work are Prof. Dr. Gabriele Chiogna and Tuo Ye. Both of them always took a lot of time to exchange ideas, from whom I learned an enormous amount and without whom the quality of my academic work would not be what it is now. Not to forget about Florian Betz, from whom I learned a lot about Central Asia and who facilitated the work in the region a lot. I must also emphasize the invaluable help of our good soul of the chair, Christiane Zach-Cretaine, without whom I would never have been able to complete this work.
Coming to my invaluable team members that not only essentially shaped this work but even more importantly, some of whom I can now call my friends. first and foremost, I would like to thank my longtime roommate, whether in the office or at our countless joint conferences, which were always a lot of fun, evaporation specialist, running and sports friend Fabian Merk, who always had an open ear for extensive hydrology as well as off-topic discussions and was one of the most important constants during this time, both professionally and personally. Hopefully more beers will follow up this dissertation. I will also not miss the time I spent with Lucas Alcamo, another friend who came out of work. Our time together in field in Kyrgyzstan will remain unforgettable and I am deeply grateful for your patience with me in manual work, which, in our experience, runs much less smoothly than things on the computer. I hope that we will be able to undertake many more joint field trips and scientific works together. I would also like to particularly highlight Florentin Hofmeister, Daniel Bittner, Moha Alquadi, Pablo Sarmiento, Jingshui

Huang, Michael Neumayer and Sonja Teschemacher, all of whom have enriched me both professionally and personally and with whom I hope to be able to stay in touch. Nevertheless, the list to whom I want to say thank you for all the time we spent and experiences we shared together, Francesca Perosa, Karl Broich , Teresa Pérez Ciria, Francesca Ziliotto, Nicole Scherer, Thomas Pflugbeil, Monica Basilio Hazas, Sisay Simachew Mekonen, Beatrice Richieri, Leonardo Arias.

I would like to extend my heartfelt thanks to my colleagues in California. In particular, I am deeply grateful to Scott Boyce from USGS for making my introduction to Fortran much smoother and for being an exceptional and patient teacher throughout the process. I also wish to thank Wes Henson and Randy Hanson from USGS for their continued support. Additionally, I would like to acknowledge the valuable guidance of Prof. Dr. Yolanda Gil from USC, whose help with academic writing and the FAIR principles has had a lasting impact on me.

I would also like to highlight the greatly appreciated support from the Graduate School, as well as the numerous opportunities provided by TUM, particularly regarding academic exchanges and research stays, which I made use of and which played a significant role throughout my doctorate.

Among the many outstanding students I had the privilege of working with, I would like to specifically thank Julian Machnitzke, Joshua Holzer, Juliane Strohmeyer and Hans Lee for their great contributions. I am excited to follow their future endeavors and successes.

This brings me to my amazing family, whose support was at least equally significant. To my greater family, namely Sylvia, Jürgen, Markus, Andrea, Juri, Leonie, Anna Lena, Trudel, Allo (and many more), who have always stood by me and shown incredible understanding when I visited for celebrations but had to spend nights working on projects and my doctoral thesis, even when I wasn't fully present. I know this is something not to be taken for granted, and I thank you from the bottom of my heart. My siblings, who never stopped making jokes about how much longer I would be studying, deserve just as much thanks. One of the people to whom I owe my deepest gratitude is my girlfriend Amira, who has endured my bad moods, sleepless nights, a constantly running computer, and my frequent absences due to travels throughout my PhD journey. Your patience, understanding, and unwavering support during this long and challenging time mean more to me than words can express. Thanks to all of you for being by my side through it all.

My friends, who probably can't believe that the work is finally finished, deserve inexpressible thanks as well. I can't count how many times I was forgiven when work had to take priority over our plans, and I truly wouldn't be where I am today without your constant support. You were always there when I needed you, while also ensuring I could step away from hydrology when necessary to focus on other, often more important, things. Of course, I would especially like to mention my long-time friend and roommate Frieder, who has put up with me for almost a decade and who supported me enormously during my studies and always helped me never to stray from my path.

Articles & Co-Author Contributions

Research Articles

- **Schaffhauser, T.**, Lange, S., Tuo, Y., & Disse, M. (2023). Shifted discharge and drier soils: Hydrological projections for a Central Asian catchment. In Journal of Hydrology: Regional Studies (Vol. 46, p. 101338). Elsevier BV. https://doi.org/10.1016/j.ejrh.2023.101338
- **Schaffhauser, T.**, Garijo, D., Osorio, M., Bittner, D., Pierce, S., Vargas, H., Disse, M., & Gil, Y. (2023). A framework for the broad dissemination of hydrological models for non-expert users. In Environmental modeling & Software (Vol. 164, p. 105695). Elsevier BV. https://doi.org/10.1016/j.envsoft.2023.105695
- **Schaffhauser, T.**, Tuo, Y., Hofmeister, F., Chiogna, G., Huang, J., Merk, F., & Disse, M. (2024). SWAT-GL: A new glacier routine for the hydrological model SWAT. In JAWRA Journal of the American Water Resources Association. Wiley. https://doi.org/10.1111/1752-1688.13199
- **Schaffhauser, T.**, Hofmeister, F., Chiogna, G., Merk, F., Tuo, Y., Machnitzke, J., Alcamo, L., Huang, J., and Disse, M.: Merits and Limits of SWAT-GL: Application in Contrasting Glaciated Catchments, Hydrol. Earth Syst. Sci. Discuss. [preprint], https://doi.org/10.5194/hess-2024-89, in review, 2024.
- Pertiwi, A.P.; Roth, A.; **Schaffhauser, T.**; Bhola, P.K.; Reuß, F.; Stettner, S.; Kuenzer, C.; Disse, M. Monitoring the Spring Flood in Lena Delta with Hydrodynamic Modeling Based on SAR Satellite Products. Remote Sens. 2021, 13, 4695. https://doi.org/10.3390/rs13224695
- Khanal, P., Hoek Van Dijke, A. J., **Schaffhauser, T.**, Li, W., Paulus, S. J., Zhan, C., and Orth, R.: Relevance of near-surface soil moisture vs. terrestrial water storage for global vegetation functioning, Biogeosciences, 21, 1533–1547, https://doi.org/10.5194/bg-21-1533-2024, 2024.
- Dežerický, D., Šinka, K., Pipíšková, P., Dumbrovský, M., **Schaffhauser, T.**, & Muchová, Z. (2023). An approach to classify areas for appropriate rainfall-runoff risk management: A case study in Drevenica basin, Slovakia. Ecological Engineering, 196, 107083. https://doi.org/10.1016/J.ECOLENG.2023.107083
- Rai, P., Bangelesa, F., Abel, D., Ziegler, K., Huang, J., **Schaffhauser, T.**, Pollinger, F., Disse, M., and Paeth, H. (2024). Extreme precipitation and temperature indices under future climate change in central Asia based on CORDEX-CORE. Theor Appl Climatol (2024). https://doi.org/10.1007/s00704-024-04976-w
- Merk, F., **Schaffhauser, T.**, Anwar, F., Tuo, Y., Cohard, J.-M., and Disse, M.: The Significance of the Leaf-Area-Index on the Evapotranspiration Estimation in SWAT-T for Characteristic Land

Cover Types of Western Africa, Hydrol. Earth Syst. Sci. Discuss. [preprint], https://doi.org/10.5194/hess-2024-131, in review, 2024.

Further Scientific Contributions - Oral & Poster Presentations

The lists refers to first author conference contributions and selected co-author contributions.

Oral

- **Schaffhauser, T.**, Lange, S., Tuo, Y., and Disse, M. (2022). Change in Climate Impact Assessment from CMIP5 to CMIP6 in a High-Mountaineous Catchment of Central Asia 2022, EGU General Assembly 2022, Vienna, Austria, 23–27 May 2022, EGU22-12551, https://doi.org/10.5194/egusphere-egu22-12551, 2022.
- **Schaffhauser, T.**, Merk, F., Hofmeister, F., Tuo, Y., Huang, J., Alcamo L., Disse, M. Chiogna, G. (2023). SWAT-G - A freely and openly available glacier routine for SWAT, SWAT User Conference, Aarhus, Denmark
- **Schaffhauser, T.**, Merk, F., Machnitzke, J., Hofmeister, F., Tuo, Y., Huang, J., Alcamo L., Disse, M. Chiogna, G. (2024). SWAT-GL: Benefits, Challenges and the Way Forward, SWAT User Conference, Strasbourg, France
- **Schaffhauser, T.**, Lange, S., Tuo, Y., and Disse, M., (2022). Change in Climate Impact Assessment from CMIP5 to CMIP6 in a High-Mountainous Catchment of Central Asia, Doktorandenseminar 2022, Schillbach, Germany
- **Schaffhauser, T.**, Lange, S., Tuo, Y., and Disse, M. (2023). SWAT-G - A New Glacier Routine for the Hydrological Model SWAT, Doktorandenseminar 2023, Krün, Germany
- Betz, F., **Schaffhauser, T.**, Lauermann, M., Chymyrov, A., Cyffka, B., Disse, M. (2022). Eine auenökologische Perspektive auf den Water-Food-Energy-Nexus des Naryn/Syr Darya (2022). Tag der Hydrologie 2022, Munich, Germany

Poster

- **Schaffhauser, T.**, Hofmeister, F., Merk, F., Chiogna, G., Huang, J., Machnitzke, J., Alcamo, L., Tuo, Y., Disse, M. (2023). Benchmarking SWAT-Gl with four USGS Benchmark Glaciers. AGU23, San Francisco, United States
- **Schaffhauser, T.**, Hofmeister, F., Chiogna, G., Disse, M. (2022). SWAT-G, a new glacier routine for the hydrological model SWAT. AGU22, Chicago, United States
- **Schaffhauser, T.**, Betz, F., Tuo, Y., Lange, S., Disse, M. (2022). Abflussverschiebungen zwischen trockeneren Böden: Hydrologische Projektionen für ein zentralasiatisches Einzugsgebiet (2022). Tag der Hydrologie 2022, Munich, Germany
- Merk, F., **Schaffhauser, T.**, Anwar, F., Cohard, J.-M., Disse, M. (2023). Simple ET methods are sufficient when detailed LAI modeling is applied: benchmark testing the SWAT-T model with eddy covariance data in Western Africa. AGU23, San Francisco, United States

- Schürz C., Bauwe A., Bieger K., Julich S., Kiesel S., Mehdi-Schulz B., Rathjens H., **Schaffhauser, T.**, Strauch M., Wagner P. D., Guse B. (2022). Analysis of the spatiotemporal variability of simulated water balance components and related SWAT+ model parameters in seven German catchments. Abstract from 2022 International SWAT Conference, San Francisco, United States, Prague, Czech Republic

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Climate change, one of the biggest threats of the 21st Century is already affecting parts of the world where 3.6 billion people live (IPCC, 2023b). It alters climate extremes, such as extreme precipitation events, droughts, heatwaves or cyclones that lead to irreversible losses around the globe. Those most at risk are those who have contributed the least but may be hit the most by the consequences. What are potential consequences? Threats to food and water security are among the greatest concerns. However, the list of impacts is diverse and long, ecosystem losses, human health issues, loss of significant portions of the cryosphere, economic impacts across sectors and regions, ocean alterations through freshwater input and acidification or changes in land cover through desertification and general land degradation serve as further examples (IPCC, 2023a). With a mean global surface temperature of 1.45°C above pre-industrial times, 2023 represents a sad summit in the young history of man-made climate change (WMO World Meteorological Organization, 2024). Another of many records in the last decades, where most of the ten warmest years since 1850 occurred. The value impressively shows how we are approaching 1.5°C, one of the thresholds decided at COP21 (United Nations Climate Change Conference) in Paris (United Nations Framework Convention on Climate Change, 2015). The chance that the year in which this dissertation was written will be another record-shattering year is one-in-three, while there is less doubt that it will be among the five warmest on record (NOAA National Centers for Environmental Information, 2024). Our human fingerprint is unequivocal and mitigation and adaptation measures inevitable, yet not sufficiently implemented or intended. Six of nine planetary boundaries, processes which are determined as crucial for a stable and resilient Earth system, were found to be exceeded (Richardson et al., 2023). Freshwater, as one of the nine boundaries, represents the global hydrological balance and is one of the disturbed and transgressed boundaries. Both components, blue and green water, have been heavily modified by humans.

Human exploitation of Earth's freshwater resources is multifaceted. Water is harnessed to fulfill diverse needs, ranging from irrigation purposes in agriculture, to providing potable water or meeting various industrial applications. Consequently, water-related conflicts of interest often arise due to divergent demands and requirements. The rate of growth in water usage is exceeding population growth, indicating a trend toward increasingly water-intensive human behaviors as part of evolving economic activities. 70% of global water withdrawals can be attributed to the agricultural sector (Ritchie and Roser, 2018), although the fraction varies significantly across different regions. Rising demands for water, coupled with changing water availability patterns pose stress and risk to the resource. Importantly, human activities exert not only quantitative but also qualitative constraints on global water resources. Various stressors, as deficiencies in sanitation, agricultural non-point source pollution, and industrial effluents contribute

substantially to these constraints, potentially leading to severe consequences for ecosystems.
Risk is also put on our most important and largest freshwater storage, the cryosphere, which holds water in its solid form, including glaciers, ice sheets or snow. The cryosphere does not only play an important role in Earth's climate, as it is involved in numerous climate feedbacks, but is also one of the systems most vulnerable to changes in climate. It is directly linked to the hydrosphere and delayed releases of water can become an important source to sustain river flow during dry periods. Glaciated basins account for around a fourth of the global land surface and also supply almost a fourth of the world's population with freshwater (Huss and Hock, 2018). A change of the cryosphere can cause significant impacts on multiple scales, for both, the water cycle and the climate. Modifications can take many forms, declines in snow cover, temporal shifts of melting periods, glacier retreat or permafrost thawing are already visible in many parts of the world, often with negative implications for the local population. An imbalance of the system can, in worst case, lead to a loss of its functioning for water and food security. The full potential a disturbance of the system might cause may not be realized for centuries due to its complex climate feedbacks. Glacier retreat is one of the examples that remarkably demonstrates the impacts of climate change. The overwhelming speed with which glaciers have disappeared around the globe since pre-industrial times is terrifying. Regions scarce in precipitation heavily rely on seasonal melt water and thus could face severe consequences when glacier shrinkage persists. Record-breaking 2023 also had devastating effects on our glaciers, with the Swiss Alps losing 4% of their ice mass in the second worst summer on record. Together with the previous year 2022, Swiss Alps lost 10% of their mass in just two years (Swiss Academy of Sciences, 2023). Greatest amounts of ice outside the poles are stored in the Himalaya and Central Asia. However, the situation does not differ from that of the rest of the world. Latest projections estimate global glacier mass losses of 26% by the end of the century (relative to 2015), even at an optimistic warming level of 1.5°C (Rounce et al., 2023). Earth's current warming trend, exceeding 1.5°C by almost a factor of two, is associated with an ice loss of even 32%.
The impacts of climate change are usually quantified and evaluated with sector-specific impact models. Hydrological models, naming only one class of impact models and traditionally used to study global and regional water cycles, already differ widely in scale, spatial discretization, process representation and complexity. Complexity of a model is often characterized by its mathematical formulations and the degree of physicality. In terms of spatial discretization, models can be categorized in lumped models, where parameters are averaged in space not accounting for spatial heterogeneity, or they can be spatially distributed, considering spatial heterogeneity across the catchment. Based on process descriptions, models can either be empirical, conceptual or physically-based. However, boundaries between the latter two are not always clear and classifications can vary. A prevailing paradigm is that the more physically based a model is. the better it might perform. *Better*, which can mean a lot in a complex hydrological system, is typically evaluated through one or multiple performance criteria for one or multiple variables, most commonly streamflow - an aggregation of catchment information. If a model is determined reliable, modeled hydrological variables are then extrapolated in space and time. To assess the impacts of climate change, the extrapolation of well-performing hydrological models often covers the end of the century to explore how persistent climate change might affect the Earth's water cycle. Even with significant uncertainty along the model chain, modeling the impacts of climate change is paramount and serves as a crucial foundation for decision-making and the development of adaptation and mitigation strategies. To minimize structural uncertainties in hydrological modeling and enhance their reliability, models are continuously refined to establish, at best, a more robust baseline for assessing climate change impacts. Yet, the skill of these models in replicating real world hydrology depends heavily on the quality and availability of the underlying data with used to drive and evaluate the model - a constraint that applies

even to the most physically accurate models. Data challenges are particularly pronounced in remote, high-mountain environments, where accessibility is limited and conditions can be hostile, especially in glaciated catchments. The intricate hydrological dynamics of these areas, shaped by factors like high altitudes, steep terrain, and snow and glacier feedbacks, further complicate model representation.
This thesis endeavors to diminish structural uncertainties in modeling by enhancing process representations within glaciated, high-mountain catchments, establishing a robust foundation for analyzing climate change impacts. Accurate projections of the hydrological cycle are essential for devising effective adaptation and mitigation strategies in the face of future climate change. However, advances in model development can introduce discrepancies in impact statements, which need to be considered. This also applied to advances in climate modeling, which can cause significant deviations in the projections of meteorological variables - a facet that is given special attention in this work.

1.1 Overarching Goal

The overall goal of the underlying dissertation is:

„Advancing the representation of high-mountain hydrology and demonstrating the implications of evolving climate models on climate impact assessment“

This is accomplished using a synthesis of different modeling approaches, model development, data analysis and processing techniques. The research covers multiple study areas characterized by various scales, topographies and hydro-climatological conditions. To attain the overall goal of the dissertation, the use of a hydrological model is foundational.
The relevance of this work is determined by the necessity for robust models as a reliable basis for thorough studies on the effects of climate change. Given the uncertainties inherent in the modeling chain of climate impact assessment, the importance of reliable impact statements might be bigger than ever. As hydrologists, our ability to influence climate modeling developments may be limited, but we have the opportunity and responsibility to make substantial contributions within our own domain to achieve above-mentioned goal. This entails demonstrating the implications of advancements in fields like climate science for real world applications on the basis of which decisions may be made.

1.2 Research Questions

Based on the overarching goal, multiple research questions are derived and introduced in detail in the following. The formulated research questions can be assigned to one of the two components that are encompassed in the overarching goal, namely a model development component and a component that emphasizes the implications of climate model generations.

1.2.1 Research Question 1

Climate models, which are widely used to drive impact models such as hydrological models, are continuously evolving and improved. The start of the Coupled Model Intercomparison Project (CMIP), which is intended to promote the further development of climate models, dates back to 1995, since then the CMIP3, CMIP5 and CMIP6 generations have been introduced. Based on recent advances and increasing popularity of climate impact studies, the question is posed:

How does the evolution of climate model generations affect the impact assessment of the regional hydrological cycle?

The question is addressed in the case of the Naryn Basin and evaluated based on the developments made between CMIP5 and CMIP6.

1.2.2 Research Question 2

A variety of sources of uncertainties arise along the modeling chain of climate impact studies. These range from climate models, bias adjustment methods, downscaling techniques and reference datasets to structural uncertainties within hydrological models, model parameters or climate scenarios. Another, far less considered, source is the generation of climate models, which acts as proxy for developments within a climate model family that indicates changes in the same climate model across different generations. Together with the linkage to Research Question 1, we want to answer the question:

To which extent do generations of climate models contribute to the total uncertainty in hydrological projections?

For this purpose, a sensitivity analysis based on an ANOVA (Analysis of Variance) approach is used to decompose the total uncertainty in its components. Analogous to Research Question 1, the Naryn serves as a case study to answer this question.

1.2.3 Research Question 3

The hydrological model SWAT (Soil Water Assessment Tool) is widely used and acknowledged in the hydrological community for a multitude of applications. A significant number of studies deal with the modeling of, partially glaciated, high mountain catchments. The inadequate representation of glacial processes exacerbates already significant uncertainties that arise from common problems in alpine catchments. Accordingly, we investigate the question:

Does the introduction of a glacier-expanded version, called SWAT-GL, enhance the applicability and credibility of SWAT in high mountain environments?

Thus, glacier processes are integrated into SWAT, which aim to realistically represent glacier mass balance and glacier evolution, the temporal changes and dynamics a glacier undergoes. The implementation introduces glaciers as new spatial and time-varying objects that leverage the Δh-parameterization (Huss et al., 2010) to enable spatially-distributed glacier changes.

1.2.4 Research Question 4

SWAT-GL provides technical innovations that facilitate and improve hydrological modeling of glaciated catchments. However, it has not yet been sufficiently tested and evaluated to ensure its unequivocal applicability in glaciated areas. Nor has it been demonstrated in detail whether it is superior to SWAT and how the newly introduced processes interact with existing processes, particularly those related to snow. Hence, no statements can be made about with respect to the performance of SWAT-GL. Accordingly, the following question arises:

How robust is SWAT-GL in contrasting, heavily glaciated basins, and what is the sensitivity of glacier parameters, including interactions with the existing snow routine?

The question is being investigated in four different glaciated catchments across the Unites States. A distinctive aspect of the study is its focus on evaluating SWAT-GL for its representation of cryospheric components, rather than the usual assessment solely based on lumped discharge signals. Aside, the question involves the identification of potential limitations that require action for the development of future versions. The catchments benefit from the availability of long glaciological, hydrological and meteorological information.

With a positive answer of the last research question, which builds on Research Question 3, it is assumed that the underlying dissertation can have a valuable contribution to both elements of the Overarching Goal. While the effect for a better representation of high-mountainous hydrology is intuitive, there is also an indirect but significant contribution in the field of climate impact assessment. This could indicate the potential for more robust climate impact statements in high mountain areas, which should increase the credibility of these studies and reduce uncertainty, even if not at the level of climate modeling.

1.3 Structure

This cumulative dissertation is based on two peer-reviewed articles published in the *Journal of Hydrology: Regional Studies* and *Journal of American Water Resources Association* as well as one article which is currently under review and available as pre-print in *Hydrology and Earth System Sciences* at the time of writing. The four Research Questions are mainly addressed in Chapter 4, Chapter 6 and Chapter 7, which cover the three articles in the order provided above, and are further supported by Chapter 5 and Chapter 8. Besides, Chapter 9 revisits all topics for a final discussion to critically highlight the most relevant aspects, some of which are elucidated beyond the analyzes in the previous chapters. Background information to topics of this work, such as climate impact assessment, catchment and high mountain hydrology or hydrological modeling is provided in Chapter 3. Due to variety of study areas and scales, all areas examined in this work are introduced and briefly summarized in Chapter 2.
The articles which exhibit the main body of the dissertation can be can be assigned to the Research Questions as follows:

- Chapter 4: Explores mainly RQ1, what general role climate model generations play in climate impact assessment and partly RQ2, whether these developments have a strong contribution to the overall uncertainty or not
- Chapter 5: Supplements Chapter 4 and focuses on RQ2, the role of climate model generations in climate change studies
- Chapter 6: Focuses on RQ3, and explores the feasibility of a glacier-extended SWAT version to increase its applicability and credibility in glaciated catchments
- Chapter 7: Addresses RQ4, to evaluate and quantify the capability and performance of a glacier-extended SWAT version in multiple and contrasting catchments
- Chapter 8: Further stresses RQ4 in a data-scarce environment and emphasizes the future way and role of the hydro-glaciological extensions developed attached to RQ3

The details and outline of the cumulative dissertation is further presented in Table 1.1.

Table 1.1: Outline and structure of the dissertation.

Overarching Goal	Research Question	Chapters	Study Regions	References
Climate Impact Assessment	How does the evolution of climate model generations affect the impact assessment of the regional hydrological cycle?	Chapter 4	Naryn	Schaffhauser et al. (2023)
Climate Impact Assessment	To which extent do generations of climate models contribute to the total uncertainty in hydrological projections?	Chapter 4 Chapter 5	Naryn	Schaffhauser et al. (2023)
Model Development: SWAT-GL	Does the introduction of a glacier-expanded version, called SWAT-GL, enhance the applicability and credibility of SWAT in high mountain environments?	Chapter 6 Chapter 8	Martelltal	Schaffhauser et al. (2024b)
Model Development: SWAT-GL	How robust is SWAT-GL in contrasting, heavily glaciated basins, and what is the sensitivity of glacier parameters, including interactions with the existing snow routine?	Chapter 7 Chapter 8	Gulkana Wolverine South Cascade Lemon Creek Upper Naryn	Schaffhauser et al. (2024a)

Chapter 2

Study Areas

The work considers multiple study areas around the globe. The main region that examined is the *Naryn Basin* located in Kyrgyzstan and thus part of Central Asia. Together with the Kara Darya, the Naryn forms the Syr Darya which is one of the major tributaries of the well-known Aral Sea. The Naryn serves as the focal point of Chapter 4 and 5. Specific focus is put on the *Upper Naryn* in Chapter 8.
Besides, investigations cover the *Martelltal* in South Tyrol (Italy). The location is primarily chosen for demonstration purposes of the introduced technical novelties due to a good availability of hydro-meteorological and glaciological data. The Martelltal is the main location of Chapter 6.
For a comprehensive evaluation of the aforementioned technical developments, four contrasting and highly glaciated basins in the Northern United States are selected. The selection is based on the availability of long and homogenized datasets for multiple glaciological variables. The four catchments are part of the USGS Benchmark Glacier Project (McNeil et al., 2016; Baker et al., 2018), which also includes hydrological and meteorological datasets, all starting at least 50 years ago. This leads to the consideration as main area of interest with respect to the evaluation performed in Chapter 7. The glaciers of study are the *Gulkana, Wolverine, South Cascade and Lemon Creek* glacier, which are distributed over the Northern United States.
An overview of the study areas is given in Fig. 2.1.

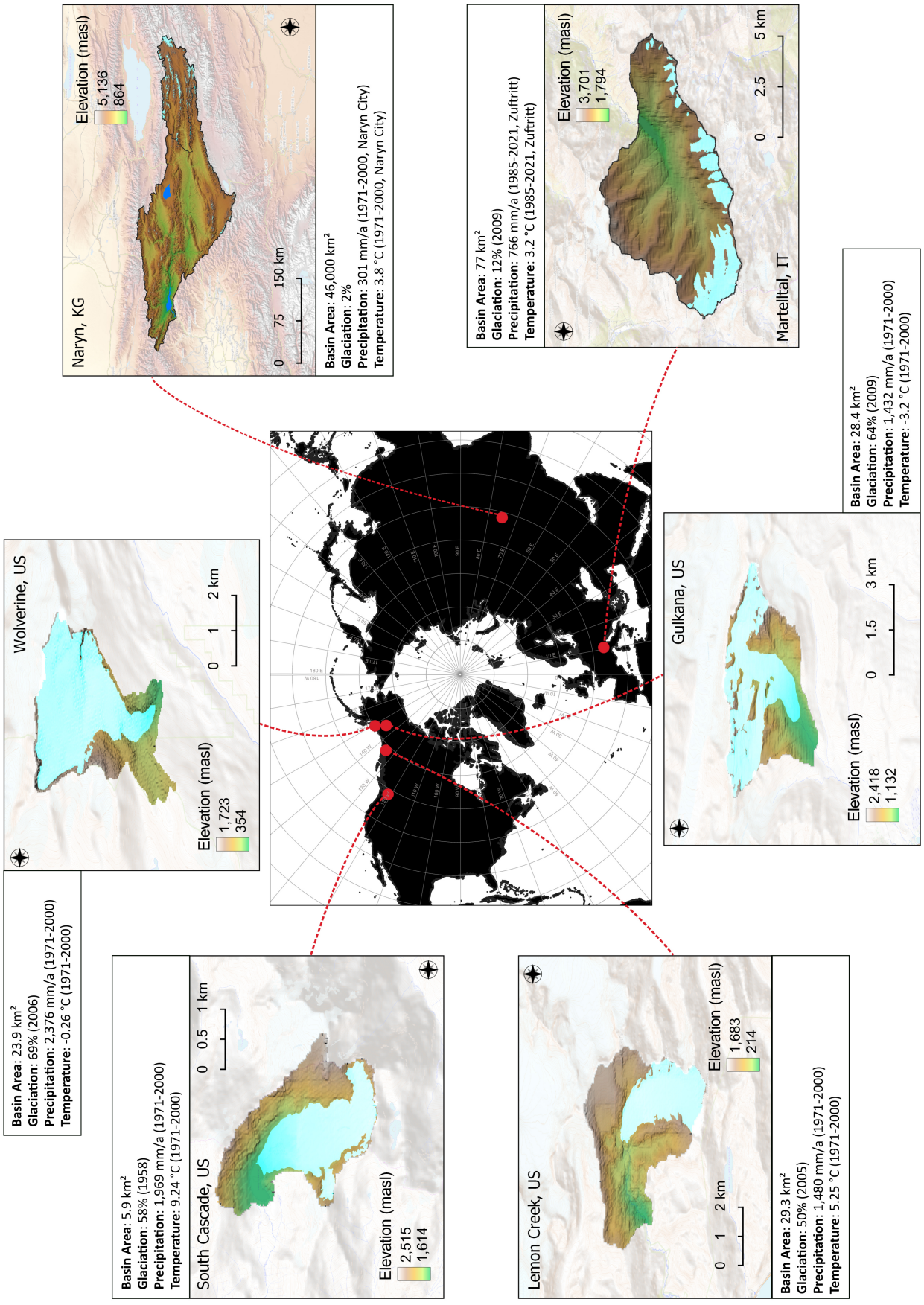

Figure 2.1: Overview of the study areas investigated in the underlying dissertation. Cyan areas represent glaciers.

Chapter 3

Background and State of the Art

In the following, an overview of the evolution of climate models and the resulting climate impact domain is given, followed by putting emphasis on the example of hydrology in the context of impact modeling.

3.1 A Time Travel in Climate Modeling

Climate models, nowadays widely-known under the term General Circulation Models (GCM), which represent Earth's atmospheric fluxes, at present commonly coupled with models that describe for example oceanic currents (another type of GCMs) are the major tools to represent and evaluate climate change (Hausfather et al., 2020; Tebaldi et al., 2021). Climate models, as the main tool to describe and improve the understanding of our climate, can look back on a long history that is characterized by paramount milestones not only for climate science but for many other domains.

Modern climate models have not much in common with those of their beginnings, which date back to the 19^{th} century already. Earlier days were dominated by conceptual approaches that attempted to translate atmospheric compositions into climatic phenomena. Some theoretical fundamentals, such as the transport of heat pole-wards or the concept of large-scale atmospheric cells for example, served as the fundamentals of modern climate models (Edwards, 2011). In parallel, geochemical advances in the understanding for example of the carbon cycle further supported the field of climate change. However, apart of all improvements of the physical knowledge on Earth's energy balance and its climate interactions, one (of many) cornerstone was laid by the famous Swedish physicist Svante Arrhenius, who estimated that a doubling of CO_2 could cause a warming of the Earth between 5 to 6°C (Arrhenius, 1896). His estimate was founded on the energy balance, including albedo and atmospheric effects such as absorption of infrared radiation, although treating Earth as a point mass and further simplifications were made (Edwards, 2011). A significant contribution in determining the role of greenhouse gases in climate change, even if the work has logically been refined and extended with time. Arrhenius' energy balance model calculated energy exchange one-dimensionally, along the meridians, yet, variants allowing longitudinal energy transport also were developed. Besides, energy balance models were developed that took into account vertical heat transfer through convection and made it possible to determine vertical temperature profiles (Edwards, 2011). Approaches that are all of great importance for modern climate modeling.

The advent of computers revolutionized the field of climate modeling and paved the way for numerical models in the 1950s and 1960s, however decades later than first numerical models based on finite-differences were suggested but rejected due to poor results (Richardson, 1922). Numerical weather pre-

diction, originally applied mainly on the local and regional scale, was the base for GCMs, which basically extended their concepts to the global scale. The concept of mass and energy transport on a vertical and horizontal grid was often a compromise between simulation time scale and spatio-temporal resolution. Norman Phillips, an American meteorologist, was one the first scientists being able to develop a GCM that was capable to capture Earth's atmospheric behavior in the mid 1950s (Phillips, 1956). A pioneering work that served as inspiration for later generations of climate modeling and set the base for building climate labs, research programs and departments, such as the UCLA Department of Meteorology or the Geophysical Fluid Dynamics Laboratory (GFDL) (Edwards, 2011).
Further progress has been made over time and GCMs began to become more complex by refining concepts and expanding the models with new components of the Earth such as oceans and feedback mechanisms. One example of the outstanding work, which was ongoing was the Community Climate Model developed at the US National Center for Atmospheric Research (NCAR). The aforementioned incorporation of ocean circulation, which coupled Atmospheric GCMs (AGCM) with Oceanic ones (OGCM) also referred to as AOGCMs (with first attempts going back into 1969 (Manabe and Bryan, 1969)), led especially in the 1980s to an increasing complexity in the system representations. The foundation for what is nowadays widely known as Earth System Model (ESM). ESMs emerged as more and more components that interact with the climate system were coupled with GCMs. The time when not only hydrological components were considered in Earth System Models but also the discipline of impact modeling evolved more and more (Katzav and Parker, 2015; Hausfather et al., 2020). Finally, the Intergovernmental Panel on Climate Change (IPCC) was established in 1988 with the aim of assessing the effects of climate change, as a response of growing evidence on human-induced climate change that could have severe consequences. In 1995 the first CMIP phase was launched. Since then, climate modeling as well as its interdisciplinarity has largely advanced and defined the base for nowadays climate impact assessment.

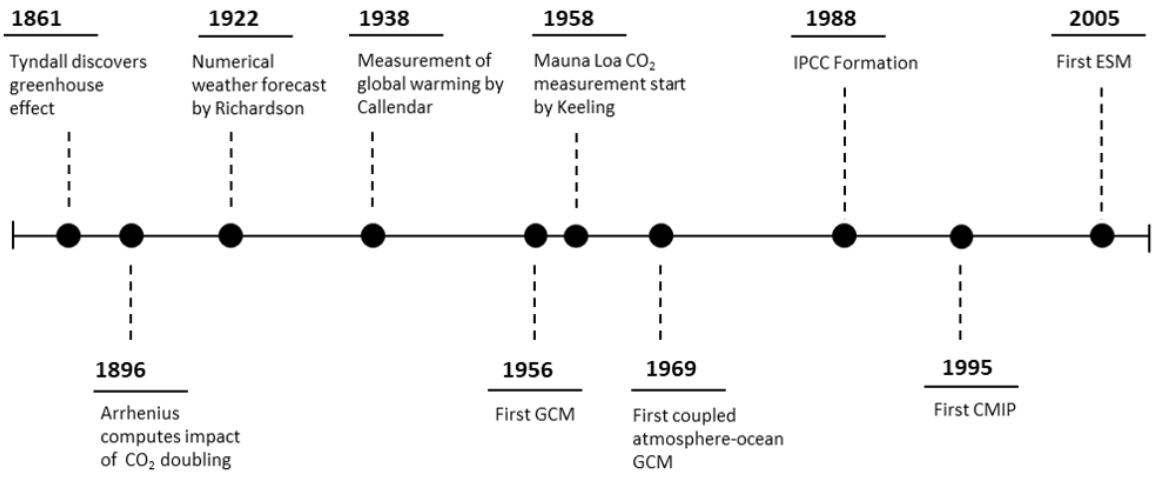

Figure 3.1: Timeline of some cornerstones in the history of climate modeling.

3.2 Modern Climate Impact Assessment

The mid-1990's set the way for the establishment of the Coupled Model Intercomparison Project (CMIP), an effort to compare climate models across the global community. The project aims to foster advances in climate modeling and to compare GCMs in a standardized way, making the evolution of climate modeling

a collaborative and community-driven approach. CMIP, being organized in phases and currently in its 6^{th} stage, shaped impact modeling as we know it today.

Almost 30 years after the first stage, the field of climate impact assessment has grown considerably, particularly in the hydrological community. The popularity of nowadays impact assessment, in the field of hydrology, can be emphasized with a literature search. The experiment consists of a literature search of three different topics based on Scopus: I) A review to demonstrate the general climate evolution of impact assessment in hydrology. II) Showing the popularity of the last stage of CMIP (stage 5) and its utilization in hydrology, and III) demonstrating the consideration of the current stage of CMIP (stage 6) and its utilization in hydrology. The results, illustrated in Fig. 3.2, highlight the increasing interest and popularity of climate change studies within the hydrological domain. In addition, it reveals the transition phase evident across different CMIP phases, suggesting continued use of earlier CMIP generations in hydrological studies. The data also reflects the ongoing increase in annual publications under CMIP6, surpassing the previous peak publication rate of CMIP5 already. Looking at present trends, it seems unlikely that the demand for studies addressing the impacts of climate change at local, regional, or global scales will decrease in the near future. However, while the sheer number of studies does not seem to be saturated, it is far less easy to distinguish which part is limited to replicas and which part represents significant progress.

ESMs have largely evolved in the last decades and nowadays integrate various components of different

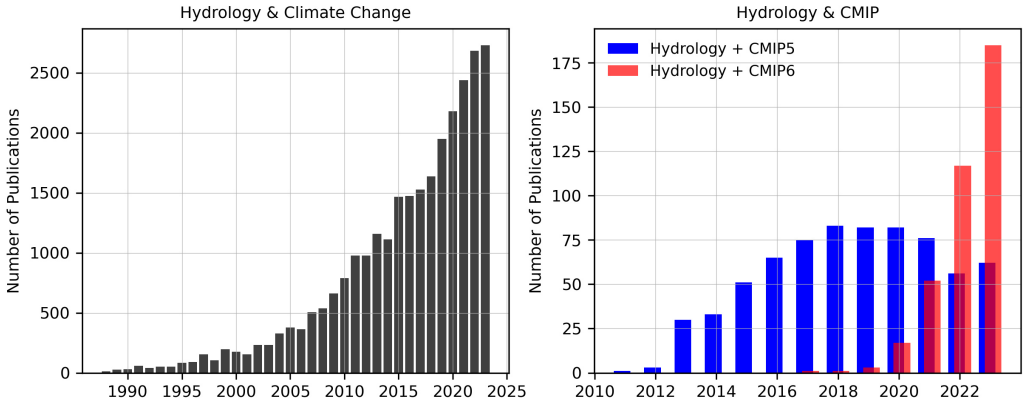

Figure 3.2: Literature results based on Scopus illustrating trends of climate change studies in the field of hydrology (left), and the utilization of CMIP5 and CMIP6 in the field of hydrology (right). The search of case 1 (left) is based on the keywords "Hydrological" OR "Hydrology" AND "Climate Change". The search of the case 2 and 3 (right) is based on the results of the search "Hydrological" OR "Hydrology" AND "CMIP5" and "CMIP6", respectively.

systems, such as the hydro- or the biosphere, recognizing the significance of an interdisciplinary assessment. In hydrology, impact assessment is often performed by forcing a hydrological model with climate projections of different trajectories, all representing a possible future climate based on different levels of CO_2. The differences in the climate pathways of various climate models can be substantial and impact modelers usually rely on an ensemble approach. However, the response in hydrological models to similar trajectories can also differ notably, which means that the choice of impact model is of general significance that can cover a big share of the inherent uncertainties of impact modeling. The choice is large, as are the differences, and the quality of the projected change of the hydrological system is usually assessed based on the model's reproducibility of the current state. Hydrological modeling therefore entails a responsibility, not only in assessing these impacts, but also in deriving measures to mitigate them.

3.3 Hydrological (Impact) Modeling

Impact models in the field of hydrology differ greatly in model structure, input demand, spatial representation and complexity. The impact assessment itself consists of multiple steps of varying structure and comprehensiveness. Nevertheless, model evaluation is the core of most impact assessment studies.
Model performance is mainly evaluated comparing past measurements of one or more variables, typically discharge, with the simulation results from the impact model. The complexity of the evaluation step can diverge widely. It ranges from simple assessments of the model, which just evaluate discharge at the outlet of the basin up to more complex and time demanding approaches that consider spatial differences, for instance using multiple gauges and variables of the hydrological cycle, such as evapotranspiration, snow-related processes or soil moisture. While none of the approaches is necessarily superior to another, the choice should be connected to the overall aim of the study. It was shown that impact statements stemming from simpler evaluation procedures can differ significantly from those based on comprehensive schemes (Krysanova et al., 2020). This is particularly true when climate impacts are evaluated for hydrological variables that were not considered in the evaluation procedure. A fact that should be at least clearly acknowledged by the modeler to enhance the clarity and reliability of the impact statement. Given the prevalent uncertainties of any impact assessment, this acknowledgment could serve to refine the robustness of the statements and draw attention to the potentially higher uncertainties.
Nevertheless, uncertainties, caused by a variety of sources and to which all approaches are subject to, remain. An appropriate communication of these is thus essential, not only to foster transparency, but also to build trust and enhance the credibility of the field. Throughout the modeling process, uncertainties arise from various sources including climate models, bias-adjustment, downscaling, and the impact model itself. Uncertainties attached to the impact model predominantly stem from two key factors: firstly, model structure and process representation, and secondly, the parameter choices derived from the evaluation procedure. A significant challenge for hydrological models lies in accurately capturing non-stationary catchment responses. A challenge that usually aggravates when the model is applied over periods significantly beyond its original evaluation period or under climatic or land surface conditions differing substantially from those encountered during the evaluation period. An inherent extrapolation problem that is not only constrained to hydrology.
An environment that might be particularly pronounced by non-stationarities are high-mountainous regions, where snow and glacier processes are crucial for the catchment hydrology. These areas often serve as major indicators of climate change due to the rapid and visible response of snow and glacier processes to climatic shifts. Shifted melt periods, towards earlier onsets, and accelerated melting rates are widespread across the globe. However, adequately representing these changes in models poses a significant challenge, especially when trends are pronounced or changes occur abruptly. In such cases, model quality and robustness of impact statements can be severely constrained. For instance, a model reliable in simulating snow processes may become less useful by the end of the century if snow dynamics cease to be significant in the catchment.
Challenges that the community should not only be aware of, but could be seen as opportunity to develop suitable solutions in order to eminently improve the field's reliability and to which this work aims to make a valuable contribution.

Chapter 4

Shifted Discharge and Drier Soils: Hydrological Projections for a Central Asian Catchment

The following chapter is based on the publication Schaffhauser et al. (2023)[1].

Abstract

We investigate the impacts of climate change in the basin based on two families of General Circulation Models (GCMs) using the hydrological model SWAT. The forcing datasets are the widely used ISIMIP2 (I2) and the newly derived ISIMIP3 (I3) data which refer to the 5th and 6th stage of the Coupled Model Intercomparison Project (CMIP). Due to notable differences in the forcing we evaluate their impacts on various hydrological components of the basin, such as discharge, evapotranspiration (ETA) and soil moisture (SM). Besides, a partial correlation (PC) analysis is used to assess the meteorological controls of the basin with special emphasize on the SM-ETA coupling. Agreement in the basin's projections is found, such as discharge shifts towards an earlier peak flow of one month, significant SM reductions and ETA increases. I3 temperature projections exceed their previous estimates and show an increase in precipitation, which differs from I2. However, the mitigating effects do not lead to an improvement in the region's susceptibility to soil moisture deficits. The PC study reveals enhanced water-limited conditions expressed as positive SM-ETA feedback under I2 and I3, albeit slightly weaker under I3.

4.1 Introduction

Central Asia (CA) is characterized by unevenly distributed renewable water resources and is prone to water-related conflicts (Bernauer and Siegfried, 2012; Berndtsson and Tussupova, 2020; Peña-Ramos et al., 2021). Water availability originates mostly from the high-mountainous part and relies to a significant part on nivo-glacial processes that supply water to the arid lowlands where it is used for agricultural production or to satisfy the energy demand (Didovets et al., 2021; Chen et al., 2017a; Hagg et al., 2007).

[1] **Schaffhauser, T.**, Lange, S., Tuo, Y., & Disse, M. (2023). Shifted discharge and drier soils: Hydrological projections for a Central Asian catchment. In Journal of Hydrology: Regional Studies (Vol. 46, p. 101338). Elsevier BV. https://doi.org/10.1016/j.ejrh.2023.101338

Central Asian's water towers are vulnerable to climate change, due to glacier recession and seasonal snow cover reductions which will lead to substantial impacts on the freshwater availability of the local population (Barandun et al., 2020, 2021; Huss and Hock, 2018; Saks et al., 2022). Further, climatic changes are already expected to exacerbate droughts, with partly adverse effects and risks on vegetation and ecosystems (Jiang et al., 2017; Zhou et al., 2015; Xing et al., 2022). CA is particularly vulnerable to land degradation, especially with its special form of desertification, which is already present and might be more pronounced in the future due to further climatic changes (Huang et al., 2020b; Zhang et al., 2018; Li et al., 2015). Reasons can be a multitude of climatic effects, such as precipitation decrease reducing moisture availability or temperature increases enhancing evapotranspiration as well as anthropogenic causes (Jiang et al., 2022; Peng et al., 2021). However, the relationships between these factors and the vegetation response are complex and challenging (jie Xu et al., 2016; Gessner et al., 2013).

Non-climatic factors such as mismanagement of water resources as it was the case for the prominent example of the Aral Sea Basin (Xenarios et al., 2018; Hamidov et al., 2016) and the general drivers of water stress in transboundary river basins, such as allocation, upstream water use or damming exacerbate water-related issues (Hill et al., 2017; Wilson et al., 2017; Döll et al., 2009).

Reported observed temperature rises range from 0.2-0.3°C per decade (1956 - 2007) in the Tian Shan (Kutuzov and Shahgedanova, 2009) up to 0.36°–0.42°C (1979 - 2011) per decade considering whole Central Asia (Hu et al., 2014). There is also evidence that the warming trends were intensified since the 1970s, while for precipitation observed changes are more heterogeneous without consistent patterns (Unger-Shayesteh et al., 2013). Climate projections in the context of the fifth phase of the Coupled Model Intercomparison Project (CMIP5) project significant temperature increases until 2099, e.g. Immerzeel et al. (2013) reports a temperature rise of around 2°C until 2050 under the RCP4.5 (Relative Concentration Pathway) for the Amu Darya and Syr Darya, which is consistent with the calculated increase of 1-2.2°C in High Asia by Lutz et al. (2014). Ozturk et al. (2017) estimates a temperature increase of up to 7°C under the high emission scenario RCP8.5 until 2099, which is in line with the projected changes in Summer of 6.5°C in Reyer et al. (2015). Precipitation trends are mostly minor and show a strong spatial diversity. Besides, precipitation changes are less significant and more uncertain for most projections (Lutz et al., 2014; Immerzeel et al., 2013; Luo et al., 2018). Only Ozturk et al. (2017) reports a more consistent precipitation decrease in their study.

More recent studies which are based on the newer Shared-Socioeconomic Pathways (SSP) and newer General Circulation Models (GCM) from CMIP6 show strong similarities in projected temperatures with slightly higher projections compared to CMIP5. Under SSP8.5 Lalande et al. (2021) calculated an upper bound of 9°C for the adjacent located High Mountain Asia (HMA). A different picture is found for precipitation changes over CA, where several studies report a clear increase in precipitation amounts, with ranges from 10.5%-14.4% under SSP2.6 and SSP8.5 until the end of the century (Jiang et al., 2020) or similar findings (Guo et al., 2021). However, CMIP6-based impact studies that investigate what these differences in the new estimates could mean are still rare but will likely increase in the near future.

The assessment of climate change impacts on different components, e.g. of the hydrological cycle is an essential element for the development of societal adaption and mitigation strategies but also to raise the general awareness of potential changes. However, climate impact studies are subject to a complex modelling chain, considering multiple scenarios, multiple GCMs (multi-model), downscaling and bias-adjustment methods and lastly an impact model such as an hydrological model (Huang et al., 2020a; Olsson et al., 2016). As a consequence impact studies are prone to large uncertainties initiated by these different sources (Wen et al., 2020; Bosshard et al., 2013). Nevertheless, the quantification and communication of uncertainties play an important role in impact modelling (Mishra et al., 2020; Vetter et al.,

2015), where one aim should be the minimization of uncertainty referred to the impact model to improve the reliability of the projected changes (Vetter et al., 2016; Krysanova et al., 2018). Thus, Krysanova et al. (2020) showed in their special issue how a robust impact model can reduce the magnitude of projected uncertainties and provided guidelines for a comprehensive model evaluation in addition to the standard calibration and validation procedure.

Due to the importance of CA's water towers and its dependence on agriculture the evaluation of its climate and hydrology is essential for adaption and mitigation. Climate impact assessment on the hydrological cycle in CA has been conducted in several studies, as it was done for eight CA catchments by Didovets et al. (2021) with varying results depending on the catchment characteristics and flow regime. Gan et al. (2015) reports a shift towards earlier peak discharge and runoff as well as precipitation declines in the Naryn Catchment. Slight reductions in runoff combined with similar significant seasonal shifts are found by Hagg et al. (2013) for the Upper Amu Darya. More substantial negative changes in runoff are estimated for the Chu River Basin by Changkun et al. (2015).

However, climate impact studies on the hydrological cycle based on CMIP6 are almost not available in the region. Thus, our aim is to provide an assessment of climate change impacts on different hydrological components within the Naryn Basin considering a subset of the recent CMIP6 GCMs and scenarios. Secondly, we do not solely focus on the climate projections based on CMIP6 but reveal how the impacts on the hydrological cycle changed from CMIP5 to CMIP6, which is particularly relevant in the context of adaptation and mitigation strategies. Our GCM subsets are derived from the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) versions 2b and 3b and will be designated as I2 and I3 in the following (see Section 4.2.1). Special emphasize is put on the differences in the projected coupling of soil moisture and evapotranspiration, initiated through disparate precipitation and temperature projections between the two generations. Hence, we perform a correlation analysis, where a partial correlation (PC) system is used to assess the meteorological effects on and changes within the soil moisture-evapotranspiration feedback.

4.2 Materials and Methods

4.2.1 Datasets used in this Study

We used the daily meteorological data of four GCMs and two scenarios from CMIP5 for the analysis of future changes and the corresponding successor models and scenarios from CMIP6 to drive our hydrological model and to evaluate how projected impacts differ between both CMIP generations. In total then 8 GCMs were utilized, 4 of each generation of climate models. In detail we used the GCM data from GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR and MIROC5 under CMIP5 and the corresponding successor GCMs GFDL-ESM4, UKESM1-0-LL (successor of HadGEM2-ES), IPSL-CM6A-LR and MIROC6 under CMIP6, respectively. The selected scenarios refer to a low and high emission scenario of both phases, namely RCP2.6, RCP8.5 and SSP1-2.6 as well as SSP5-8.5, respectively. In the following we denote them simply as RCP26, RCP85, SSP26, SSP85.

The data was obtained from the ISIMIP project, where the GCM data was separately downscaled and bias-adjusted for each CMIP phase. ISIMIP is (analogously to CMIP) structured in various phases, where ISIMIP2b (Frieler et al., 2017; Lange, 2018) corresponds to CMIP5, while ISIMIP3b (Lange, 2019) utilizes CMIP6 results. The ISIMIP2b procedure preserves trends in long-term mean values using adjustment factors or offsets for those statistics and transfer functions for the adjustment of day-to-day variability around the monthly mean value. In contrast, the ISIMIP3b procedure adjusts biases using

quantile mapping and preserves trends in all quantiles of the distribution of a climate variable. Statistical downscaling is done via spatial interpolation in ISIMIP2b and with multivariate quantile mapping in ISIMIP3b. Furthermore, the bias-adjustment was performed for different reference datasets, namely EWEMBI under ISIMIP2b (CMIP5 GCMs) (Lange, 2018) and W5E5 under ISIMIP3b (CMIP6 GCMs) (Cucchi et al., 2020; Lange et al., 2021) providing the base for the calibration and validation of our hydrological model. Both reference datasets were designed for the bias-adjustment of climate impact studies. Both datasets merge various data sources, EWEMBI hereby is based on EartH2Observe forcing, ERA-Interim, NASA/GEWEX Surface Radiation Budget data and WFDEI data. The predecessor W5E5 combines the methodology of version 1.0 WATCH Forcing Data, ERA5 reanalysis data, and precipitation data from version 2.3 of the Global Precipitation Climatology Project. The two reference datasets are the base for the bias-adjustment of the corresponding ISIMIP phase.

To evaluate simulated actual evapotranspiration (ETA) and soil moisture (SM) of SWAT, we considered

Table 4.1: Overview of used datasets. Note: The climate projections of ISIMIP2b and ISIMIP3b belong to CMIP5 and CMIP6, respectively.

Variable/Data Type	Abbreviations	Datasets	References
Temperature, precipitation	T, P	EWEMBI, W5E5	Lange (2018); Cucchi et al. (2020); Lange et al. (2021)
Evapotranspiration	ETA	GLEAM V3.5	Miralles et al. (2011); Martens et al. (2017)
Soil moisture	SM	GLDAS Noah V2.0	Rodell et al. (2004)
Digital elevation model	DEM	SRTM90	NASA JPL (2013b)
Land use	-	ESA CCI-LC	ESA (2017)
Soil	-	HWSM	FAO et al. (2012)
Climate projections/ models	I2, I3	ISIMIP2b:GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR, MIROC5 ISIMIP3b: GFDL-ESM4, UKESM1-0-LL, IPSL-CM6A-LR, MIROC6	Frieler et al. (2017); Lange (2018, 2019)

the widely used data from the Global Land Evaporation Amsterdam Model (GLEAM) (Miralles et al., 2011; Martens et al., 2017) and Global Land Data Assimilation System (GLDAS) (Rodell et al., 2004), respectively.

The Digital Elevation Model (DEM) with a resolution of 90 m was acquired from the Shuttle Radar Topography Mission (SRTM) (NASA JPL, 2013b). Land use data was obtained from the European Space Agency's Climate Change Inititiave Land Cover (ESA CCI-LC) project (ESA, 2017) with a resolution of 300 m and soil data was received from the Food and Agricultural Organization's (FAO) 1 km Harmonized World Soil Map (HWSM) (FAO et al., 2012). As the used ISIMIP data represents only a subset of available GCMs of each CMIP generation it has to be noted that our approach rather compares a sample of each CMIP generation based on the ISIMIP data than providing a complete picture, which was beyond the scope of that study.

4.2.2 Study Area - Naryn Basin

The Naryn Basin is located in Kyrgyzstan and is, along with the Kara Darya one of the two headwater catchments of the Syr Darya River in the Aral Sea Basin. The Naryn is formed by the two headwater rivers, Big and Small Naryn, which originate in the Tian Shan mountains. It flows westwards through Kyrgyzstan until it crosses the border to Uzbekistan, where it joins the Kara Darya in the Fergana Valley. The basin is characterized by a nivo-glacial flow regime, where the flood period begins in April

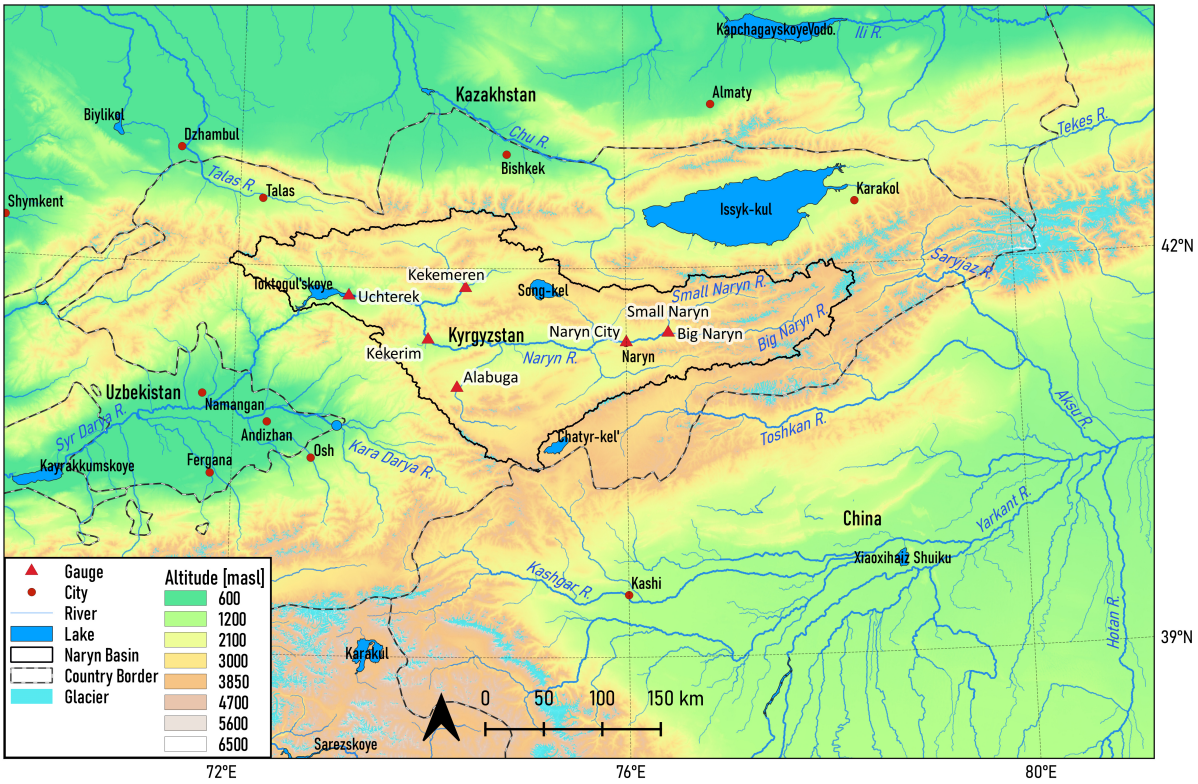

Figure 4.1: Overview of the Naryn Basin within Kyrgyzstan. The gauges refer to the subcatchments which were used for the multi-site calibration of this study. The gauges of the Small and Big Naryn are only a few meters apart (just before their confluence) and thus share one symbol.

and ends in September. The catchment size is around 58,000 km², however our study only considers the share of the catchment (~46,000 km²) until the Toktogul Reservoir, a multi-purpose reservoir used for hydroelectricity generation and irrigation. Around 2.2% of the catchment is glacierized, where the majority of the glaciers lie in the Eastern area.

The region has a dry continental and semi-arid climate with annual precipitation sums between 280 mm-450 mm (Aizen et al., 1995). Depending on the location and elevation within the basin, the annual maximum precipitation occurs in Spring or Summer. Temperatures are diverse throughout the basin and the mean Summer temperatures can reach 25 °C and during the Winter months the mean temperatures can fall down to -20 °C. An overview of the study area can be found in Figure 4.1.

4.2.3 Hydrological Model - SWAT

The Soil Water Assessment Tool (SWAT) is a process-based, semi-distributed and continuous model which was developed by the Blackland Research & Extension Center of the United States Department for Agriculture (USDA) (Arnold et al., 1998). SWAT was initially developed to evaluate the impact of land management practices in large watersheds, but is nowadays used for a multitude of applications all over the world. SWAT is widely used for sediment studies, water resources management, climate impact assessment or agricultural investigations (Arnold and Fohrer, 2005).

Spatially, the catchment is subdivided into subbasins, on the basis of which a further spatial discretization into Hyrdological Response Units (HRUs) takes place. The latter one serves as the smallest spatial unit within the model and represents a unique combination of land use, soil and slope on the subbasin level.

Computations are divided into a land and a water phases, where the calculations of the land phase are carried out on the HRU level. The land phase hereby covers the processes which are responsible for the movement of water, nutrients or sediment to the stream of a subbasin. SWAT first sums up the area dependent HRU loadings and fluxes of each subbasin before they are transferred to the corresponding stream. The water or routing phase refers to in-stream processes, such as the transport of water, sediment or nutrients through the channel network.

4.2.4 Calibration, Validation and Evaluation Test

This section describes the calibration and evaluation procedure of the hydrological model based on the reference datasets of both projection families. The approach follows the suggestions of Krysanova et al. (2018, 2020) and some modifications provided in Gelfan et al. (2020). The whole methodology is represented in Fig. 4.2. For each of the two reference datasets of the ISIMIP phases a SWAT model was set up individually. In detail we performed the following steps, (1) multi-site calibration of seven gauges considering multiple goodness of fit criteria (GOF) (Fig. 4.2a), (2) model robustness check under contrasting climate periods (Fig. 4.2b), (3) cross-evaluation of model robustness using evapotranspiration and soil moisture estimates (Fig. 4.2b).

First, the procedure was applied for the SWAT model forced with the EWEMBI reference dataset (representing I2), before we transferred the obtained parameter set to the W5E5 (representing I3) driven impact model and the same steps were performed again. Only if all steps for both hydrological models (forced with the two reference datasets of I2 and I3) could be carried out satisfactorily, the impact assessment is conducted. Due to inconsistencies in data availability and quality, daily discharge data was only available for four gauges and monthly data was obtained for the three remaining gauges. Based on the available data the calibration period was set to 1974 - 1981 and the validation period from 1982 - 1987 for the EWEMBI driven model. A limitation arose through different temporal coverage of EWEMBI and W5E5. Originally, both datasets are available from 1979 onward. However, for EWEMBI the extended GSWP3-EWEMBI (hereafter EWEMBI) product enabled a prolonged temporal coverage. A similar extension was not accessible for the W5E5 dataset at the time the calibration was done. Therefore the calibration period was set to the validation period of the model driven with EWEMBI. Due to these limitations a separate and suitable validation phase could not be used. As a preliminary study the two forcing datasets were compared and evaluated to quantify differences in the basin.

Step 1 of the procedure considered multiple rules: We calibrated from upstream to downstream and for each gauge multiple initial simulations were performed to close the water balance. The number varied throughout the subcatchments based on how fast the water balance could be closed. Multiple GOF criteria were selected during the calibration procedure. The water balance closure was assessed based on the Percent Bias (PBIAS), where we adapted and fixed the lapse rates for temperature and precipitation. Secondly, different parameter groups were calibrated (e.g. snow parameters, groundwater parameters etc.) and evaluated using the Nash Sutcliff Efficiency (NSE), Kling Gupta Efficiency (KGE) in combination with a seasonal RMSE to account for a trade off between daily/monthly and seasonal results. An overview of the selected parameters and their categories is presented in A.1.

When satisfactory results in the calibration and validation phase were achieved, a robustness check for contrasting climatic conditions was performed (step 2). The model performance is evaluated against wet and dry years, as well as for warm and cold years, by forming a sequential time series of non-continuous years. Each sequence then represents the respective climate. Thus, the mean annual precipitation and temperature for the calibration and validation period were calculated. Each year was then individually

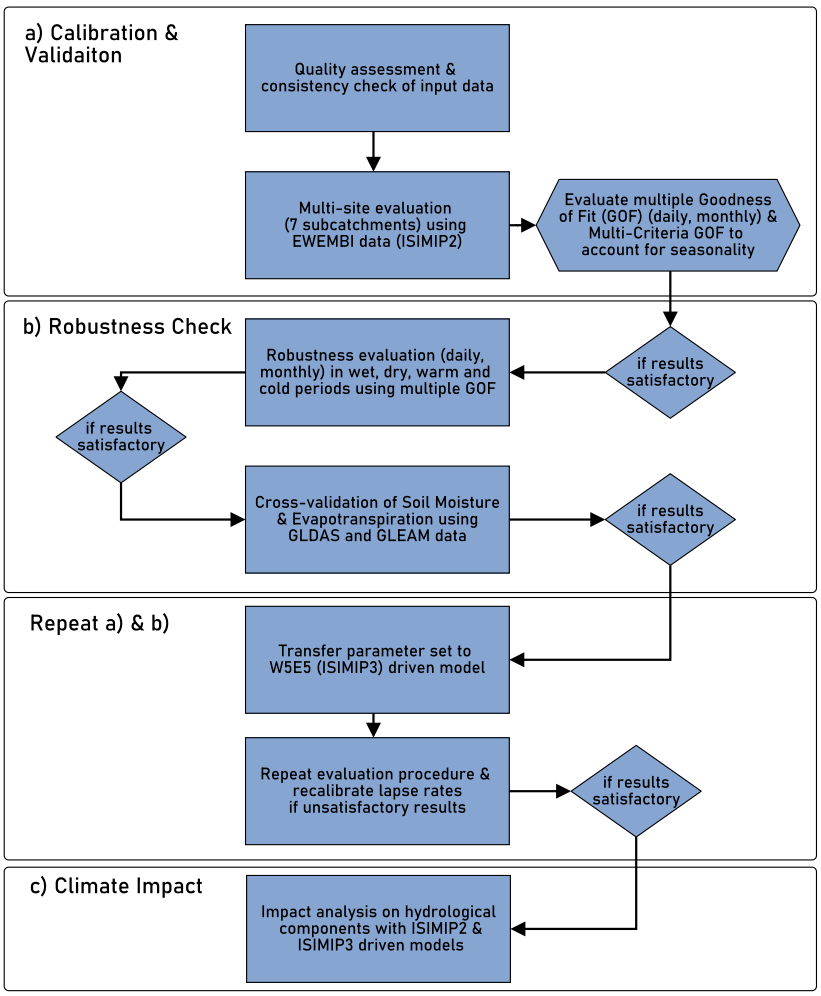

Figure 4.2: Flowchart showing the methodology of our study based on the comprehensive evaluation test proposed by Krysanova et al. (2020). The study is structured in three blocks, the calibration and validation, the robustness check followed by the final impact assessment. The multiple Goodness of Fit (GOF) correspond to Nash-Sutcliffe Efficiency (NSE), Kling-Gupta Efficiency (KGE), Root Mean Square Error (RMSE) and Percent Bias (PBIAS).

divided by the mean value and either classified as cold if the mean temperature of a year, Ti (where i is the individual year) is smaller than the mean annual temperature of the whole time series, $Tmean$, (or warm if $Ti > Tmean$), and wet if the annual precipitation of a year, Pi, $>$ the mean annual precipitation of the time series, $Pmean$ (or dry if $Pi < Pmean$). The four resulting discharge time series per SWAT model were then evaluated with the GOF criteria introduced under step 1.

Lastly (step 3), a cross-evaluation using ETA data from GLEAM as well as SM from GLDAS is performed to ensure an adequate internal model consistency and to represent various components of the hydrological system (Krysanova et al., 2020). The analysis was done on the basin and subbasin scale for monthly and annual estimates. The performance of ETA is illustrated by means of the the NSE, RMSE and the Pearson Correlation r. However, SM could not been validated in absolute values due to a mismatch in the soil layer definitions within the SWAT model and GLDAS, which provides SM values for different layers (e.g. 0-10 cm or 10-40 cm). Thus, we rather assessed seasonal and inter-annual SM anomalies, again by a (r) coefficient and a normalized version of the RMSE (NRMSE), computed by a normalization

Category	Parameter	File	Description
Water Balance	PLAPS	.sub	Precipitation Lapse Rate [mm/km]
	TLAPS	.sub	Temperature Lapse Rate [°C/km]
Snow	TIMP	.sno	Snowpack Temperature Lag Factor [-]
	SMFMX	.sno	Melt Factor for Snow on 21st June [mm H20/(°C*d)]
	SMFMN	.sno	Melt Factor for Snow on 21st December [mm H20/(°C*d)]
	SMTMP	.sno	Snowmelt Temperature [°C]
	SFTMP	.sno	Snowfall Temperature [°C]
Groundwater	ALPHA_BF	.gw	Baseflow Alpha Factor [days]
	GWQMN	.gw	Shallow Aquifer Threshold for Baseflow to Occur [mm]
	GW_DELAY	.gw	Groundwater Delay [days]
	RCHRG_DP	.gw	Fraction of Percolation for Deep Aquifer [-]
	REVAPMN	.gw	Shallow Aquifer Threshold for Revap to occur [mm]
ETP & Runoff	EPCO	.hru	Plant Uptake Compensation Factor [-]
	ESCO	.hru	Soil Evaporation Compensation Factor [-]
	SURLAG	.hru	Surface Runoff Lag Time [days]
	CN2	.mgt	Curve Number for Moisture Condition II [-]
Soil	SOL_AWC	.sol	Available Water Capacity for the Soil Layer [-]
	SOL_k	.sol	Available Water Capacity for the Soil Layer [mm/h]
	SOL_BD	.sol	Soil Bulk Density [mg/m^3]

Table 4.2: Overview of the model parameters which were calibrated in the hydrological models driven by the reference datasets of the two ISIMIP families. Parameter ranges were individually set and fixed for each calibrated subbasin/gauge.

with the interquartile range (difference between 75th percentile and 25th percentile). Similar difficulties with SM evaluations are presented for example in Poméon et al. (2018), Rajib et al. (2016) and Dembélé et al. (2020).

To allow for an appropriate comparison of the impacts based on the I2 and I3 driven models, we aimed for a consistent parameterization of both models. Thus, steps 1-3 were first performed for the SWAT model driven with EWEMBI and then, when satisfactory results were achieved, the same parameterization was used for the model forced with W5E5. Then all three steps were carried out again. A final parameterization was stated as acceptable when similar results are obtained for both models. However, due to minor differences in the forcing datasets (EWEMBI and W5E5) in their mean annual precipitation and temperature, minor adaptions in the parameterization were necessary to match the water balance. The adaptions were thus limited to the lapse rates and in seldom cases to the snow pack temperature lag factor.

4.2.5 Projected Changes of ISIMIP2b and ISIMIP3b

This section illustrates the comparison between both generations of projections. First, we show how the temperature and precipitation projections as well as their estimates in the historical phase differ. The comparison is done on the basin scale for multiple temporal representations (seasonal, annual). The future projections are derived from the four general circulation models of each projection phase (I2 and I3, respectively) and scenario. In order to ensure an adequate comparability, we selected GCMs that are both available in I2 as well as in the new phase I3, represented by a corresponding successor model (see 4.2.1). Accordingly, we focused on CMIP5's RCP scenarios that were translated and continued by a corresponding SSP scenario that characterizes a similar aggregated forcing. However, there are several differences between the RCP scenarios and their successors, which are explained and demonstrated in detail in Meinshausen et al. (2020) or Tebaldi et al. (2021). This assumption leads to the consideration

of the low emission (RCP2.6 and SSP2.6) and high emission (RCP8.5 and SSP8.5) scenarios. As SWAT requires $Tmax$ as input, we use the respective quantity as proxy for mean daily temperature changes. Precipitation changes ΔP refer to relative anomalies of the mean annual sum.

4.2.5.1 Differences in Meteorological Forcing

First of all we present the differences in the meteorological variables of all GCMs between I2 and I3. Therefore we calculate anomalies of two future 30-year periods, near (2020 - 2049) and far future (2070 - 2099), to the historical baseline period (1971 - 2000). Differences are illustrated individually for precipitation and temperature as multi-model mean (MMM) of the GCMs of each ISIMIP phase. Uncertainties are represented as GCM range. The radiative forcings of the selected scenarios of I2 and I3 are equal (e.g. SSP85 similar to RCP85). Deviations arise mainly from different climate sensitivities present in the GCMs of the different CMIP or ISIMIP phases.

4.2.5.2 Changes in Hydrological Components

As a next step, future hydrological changes in the basin are evaluated as a result of the various trajectories. Therefore, discharge, ETA and SM anomalies are assessed for the I2 and I3 forced hydrological models and compared afterwards. The model spread provides a representation of the uncertainty, and the ensemble mean represents the average future change per scenario and generation of projection. Analogously to the meteorological changes, hydrological changes are presented on a basin scale as seasonal and annual anomalies.

4.2.6 Meteorological Controls of the Soil Moisture and Evapotranspiration Feedback

Special emphasize is put on the relevance of meteorological drivers on variations in SM and ETA as well as the future development of the ETA-SM feedback under both generation of climate models. It was aimed to evaluate to what extent meteorological differences under I2 and I3 affect the coupling of SM and ETA. The interaction is implicitly considered in the model (non-linear decrease when SM is below field capacity). To identify the major controls and their effects on the SM-ETA feedback, a partial correlation system was developed, which gives further insights on the direction and strength of the SM-ETA mechanism. The evaluation focused on the MMM results of the GCMs (e.g. the MMM for each variable of a specific scenario and time period within the corresponding ISIMIP phase) to draw conclusions with regard to the mean signal. For the analysis monthly anomalies of the meteorological and hydrological variables were generated.
Since ETA and SM variations are larger during the vegetative season (which partly coincides with the rainy season, see Fig. A.2), the analysis was based on the period from April to September. The vegetation season is thus assumed to be stationary until 2099 and potential phenological shifts are not considered. For the analysis the vegetative season was split in two periods, April-June (AMJ) which roughly coincides with the rainy season in the basin and July to September (JAS), which is generally drier on the basin scale. It has to be noted that irrigation, which is the largest water consumer of Kyrgyzstan, was not explicitly considered in the simulations. The PCs allow to directly control for the effects of one or more independent variables to identify confounding effects (variable that explain part of the relationship which is present in the standard regression between a dependent and independent variable). We thus considered the combined effects of temperature, precipitation, SM and ETA to infer potential future shifts in the

SM-ETA interaction and their dominant controls. In our case combined effects refer to the consideration of two control variables, e.g. in the PC calculation between temperature and SM, the effects of the remaining variables ETA and precipitation are taken into account. Accordingly, the PC calculation is supported by the standard Pearson correlation values to highlight the confounding effect. A significant shift in the correlation values (Pearson and PC value) indicates that the relationship cannot be explained by one driver alone, but is partly based on the second variable. ETA variations are assessed by the explanatory strength of temperature and SM, while SM was evaluated using temperature. Since SM generally controls ETA in water-limited regimes and not vice versa, we present the SM-ETA coupling in the PC system of ETA rather than SM (Berg and Sheffield, 2019). Mathematically, the PC system can be formulated as follows:

$$r_{yx_1,x_2} = \frac{r_{yx_1} - r_{yx_2} r_{x_1x_2}}{\sqrt{(1 - r^2_{yx_2})(1 - r^2_{x_1x_2})}} \tag{4.1}$$

where, r_{yx_1,x_2} refers to the partial correlation coefficient between the dependent variable y (SM or ETA) and the independent variable x_1, while controlling for the effects of the independent variable x_2, the remaining r represent the Pearson correlation coefficients between the corresponding pairs of variables. The control variables can be precipitation, temperature, SM or ETA. E.g., if we want to evaluate potential effects of precipitation, the control variable x_2 would refer to precipitation and x_1 to temperature. Hence, $r_{x_1x_2}$ would refer to the Pearson correlation of precipitation and temperature.

4.3 Results

Hereafter, we present the results of the calibration procedure, the comparison of projected meteorological and hydrological changes under the new and old climate model generations, the relevance of the meteorological controls and the findings of the uncertainty decomposition.

4.3.1 Calibration, Validation and Evaluation Test

The following sections correspond to part a) and b) illustrated in Figure 4.2. Statistical results (w.r.t. different GOF criteria) for all seven gauges are presented in Tab. 4.3. Note, due to unsatisfactory data availability there is no validation phase for W5E5 (see 4.2.1).
Apart from the Big Naryn Basin, a significant deterioration between the monthly and daily calibration results, as well as between EWEMBI and W5E5 could not be detected. Best results were achieved at the basin outlet. On average the NSE values are around 0.75 for the calibration and validation period (KGE 0.82, PBIAS 9.4%) in the two models. Worst results are found at the two headwater streams Big and Small Naryn. The poorer results here are likely to be caused by the higher relevance of glacier processes, which were not represented by the model. A comprehensive discussion on that can be found in the discussion section at the end of this work. Besides, a notable difference in water volumes (expressed as PBIAS) was found at gauge Kekirim, which is caused by a systematic overestimation of late Summer and Autumn flow (July-October). Figure 4.3 shows the daily/monthly calibration and validation results for four selected gauges combined with their seasonal hydrographs. We selected one gauges which was calibrated monthly and three daily calibrated gauges to emphasize the comparable performance. For the daily calibrated gauges the selection comprised two well-performing gauges followed by the Big Naryn where poorer results were achieved. In general, both daily/monthly and seasonal hydrographs of both models show a good agreement with the observed flow with the exception of the Big Naryn. In

general, the model based on EWEMBI data as well the one driven with W5E5 are able to reproduce the seasonal behavior within the respective subcatchments. However, in case of the poorer performing headwater catchment it becomes apparant that the deterioration is likely caused by a mismatch during summer where a systematic underestimation occurs. The mean seasonal plots include observed flow for two different periods, each representing the maximum available data of EWEMBI (1974-1987) and W5E5 (1982-1987), respectively.

The robustness check, in which the model performance was individually assessed in wet/dry and

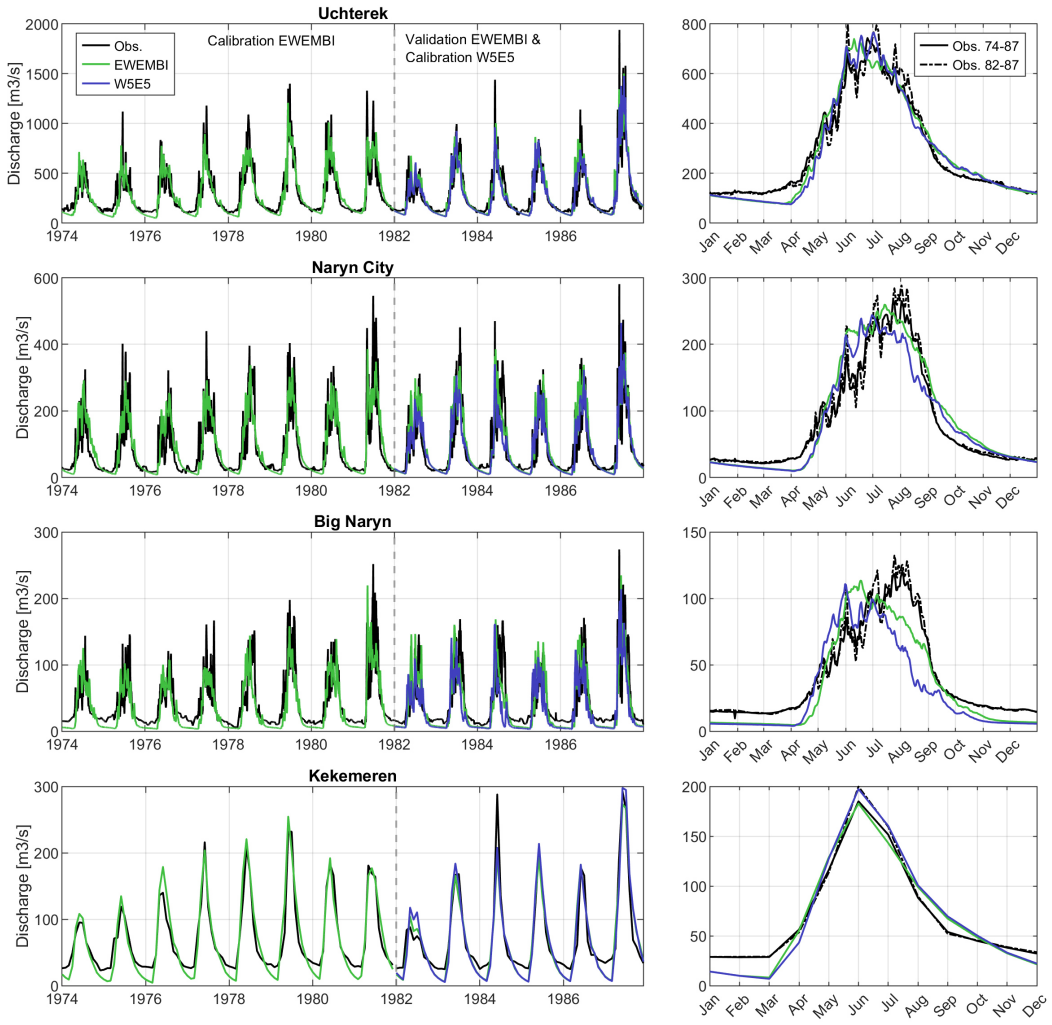

Figure 4.3: Calibration and validation results for four selected gauges. Gauge Uchterek, Naryn City and Big Naryn were calibrated on a daily scale. Kekemeren is based on a monthly calibration due to limited data availability. Note, for the period 1974-1981 only the extended EWEMBI dataset was available. Plots in the right column show the mean seasonal flow for the period 1974-1987 (calibration & validation phase of EWEMBI) and 1982-1987 (calibration phase W5E5): black lines - observed flow of the indicated periods, blue lines - W5E5 simulated flow 1982-1987, green lines - EWEMBI simulated flow 1974-1987

warm/cold periods (see section 4.2.4), led to comparable results as achieved in the original calibration and validation procedure. The results of the evaluation for all gauges can be found in Table 7.4. Overall, wet periods are better reproduced by both models for both GOF criteria than dry years (NSE_{wet} 0.8, NSE_{dry} 0.7, $PBIAS_{wet}$ $\pm 9\%$, $PBIAS_{dry}$ $\pm 11.8\%$). Also for warm and cold phases, model performances

are similar (NSE_{warm} 0.75, NSE_{cold} 0.71 and $PBIAS_{warm}$ ±9.1%, $PBIAS_{cold}$ ±10.46%). The numbers are average values of both models and all gauges. Similar to the original calibration, the same gauges exhibit the worst performance. Comparing the robustness of the EWEMBI and W5E5-driven models, no significant differences are observed. The model forced by EWEMBI data performs better in warm periods, slightly better in dry periods, similar in wet periods and slightly worse in cold years compared to its W5E5 counterpart. It can be concluded that the methodology is generally able to produce comparable results for both models. In overall, the models performed reasonable for most of the gauges in all climate phases, as well as for the standard calibration and validation procedure and thus successfully pass the first two steps of the evaluation test.
Gauges that produced unsatisfactory results are in line for both models, indicating the similarity in the parameterization and the meteorological forcing of both reference datasets. Poor results relate mainly to the Big Naryn and gauge Kekerim, where the poor performance mainly arises from water balance issues (indicated by PBIAS). A notable deterioration of the results is found in the Small Naryn Basin when moving from EWEMBI to W5E5. In contrast, an improvement is even obtained at gauge Kekerim.

Table 4.3: Calibration and validation results for the final parameter set used for the climate impact study. The last three gauges refer to the gauges where no daily discharge data was available and the calibration was performed on a monthly scale. Note that the calibration period for EWEMBI was 1974 - 1981 (validation period 1981 - 1987), while it was 1981 - 1987 for W5E5 (no validation phase) due to insufficient data.

	NSE [-]			KGE [-]			PBIAS [%]		
	Calibration		Validation	Calibration		Validation	Calibration		Validation
Gauge	EWEMBI	W5E5	EWEMBI	EWEMBI	W5E5	EWEMBI	EWEMBI	W5E5	EWEMBI
Uchterek (46,440 km²)	0.83	0.87	0.84	0.9	0.93	0.92	3.67	2.97	-2.35
Naryn City (10,291 km²)	0.74	0.74	0.72	0.84	0.84	0.82	-5.53	3.67	-10.79
Big Naryn (5,529 km²)	0.59	0.45	0.58	0.73	0.63	0.77	18.97	26.76	11.31
Small Naryn (3,877 km²)	0.78	0.59	0.63	0.84	0.62	0.81	5.38	21.29	-7.55
Alabuga (3,718 km²)	0.83	0.78	0.75	0.91	0.84	0.79	-3.82	5.6	-4.79
Kekerim (1,715 km²)	0.65	0.89	0.79	0.71	0.88	0.79	-26.27	-9.57	-12.92
Kekemeren (8,199 km²)	0.89	0.88	0.88	0.85	0.91	0.91	2.81	4.33	6.91

Table 4.4: Results of the robustness check for wet/dry and warm/cold years under consideration of the whole simulation period (calibration with validation phase) for the model driven with EWEMBI. For the W5E5-based model the periods were formed within the period 1982-1987, due to limited data.

	NSE [-]								PBIAS [%]							
	EWEMBI				W5E5				EWEMBI				W5E5			
Gauge	Wet	Dry	Warm	Cold	Wet	Dry	Warm	Cold	Wet	Dry	Warm	Cold	Wet	Dry	Warm	Cold
Uchterek	0.86	0.8	0.88	0.82	0.88	0.84	0.84	0.88	-5.5	-5.18	3.95	-2.11	-4.37	7.39	2.7	3.1
Naryn City	0.78	0.69	0.78	0.66	0.8	0.7	0.8	0.71	-5.7	-9.22	-2.28	-14.01	-4.18	7.95	-0.96	5.89
Big Naryn	0.66	0.5	0.62	0.55	0.62	0.3	0.53	0.42	15.75	15.55	19.69	11.27	23.04	28.95	26.49	26.90
Small Naryn	0.7	0.73	0.81	0.56	0.65	0.56	0.65	0.55	-6.8	3.48	8.17	-10.01	-6.11	32.29	21.37	21.25
Alabuga	0.86	0.71	0.87	0.68	0.85	0.74	0.69	0.77	-7.97	-1.64	-1.31	-7.5	-7.72	-0.59	13.67	5.55
Kekirim	0.77	0.67	0.75	0.72	0.82	0.71	0.89	0.9	-20.6	-19.65	-18.89	-21.34	-17.46	-19.5	3.03	-9.01
Kekemeren	0.9	0.87	0.91	0.87	0.9	0.88	0.53	0.89	0.82	7.23	4.32	4.93	0.18	6.63	-0.24	3.56

4.3.1.1 Cross-Validation of Actual Evapotranspiration

To improve the models reliability, they were cross-validated with respect to the GLEAM ETA product in order to increase the credibility of the projected changes. The cross-validation is performed to assure consistency of the hydrological processes within the catchment and to further improve model robustness. Due to the data availability it was performed from 1982 to 1987.

The simulated ETA of both models show a high agreement with the ETA estimates of GLEAM. At the catchment scale, a NSE of 0.97 is achieved based on the monthly evaporation estimates for the EWEMBI-forced model (RMSE = 4.1 mm, PBIAS = 3.2%). The performance of the W5E5-driven model is similar with a NSE of 0.96 (RMSE = 4.6 mm, PBIAS = 3.3%). Basin values for r are almost 1. On the subbasin scale the median NSE and r value are >0.8 and >0.9, respectively, for both models without any significant difference between them. NRMSE values are close to 0 for both models with both medians around 0.20. However, the small number of subbasins performing poor arises from both, a systematic over- and underestimation of SWAT in the Summer period, without any clear pattern. Due to the high agreement between the ETA estimates of both models and an overall very good performance, results are considered satisfactory.

Fig. 4.4 summarizes the monthly results on the basin and subbasin scale, where the latter is represented as boxplot for the individual performance metrics.

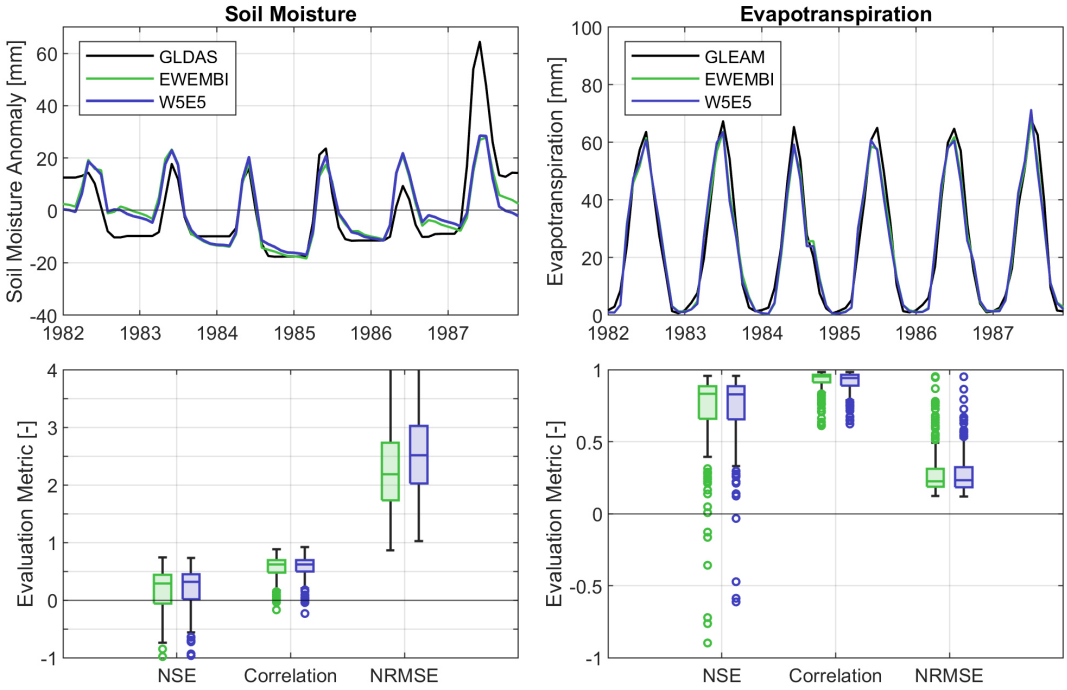

Figure 4.4: Evapotranspiration and soil moisture performance in the period 1982-1987. The top panels shows the evapotranspiration and soil moisture estimates from SWAT compared to the results obtained from GLEAM and GLDAS, respectively, at the basin scale. The lower row demonstrates the subbasin performance of the SWAT models evaluated with the Nash–Sutcliffe Efficiency (NSE), Pearson correlation (r) and Normalized Root Mean Square Error (NRMSE) metrics for soil moisture and evapotranspiration. Outliers (outside 1.5 times the interquartile range above or below the upper/lower quartile) are indicated as green or blue dots, respectively.

4.3.1.2 Cross-Validation of Soil Moisture

Similar to ETA, the SM estimates of the two SWAT models were evaluated using GLDAS data. The performance was assessed for soil moisture anomalies. We also used a NRMSE due to different soil layer assumptions (see section 4.2.4). Analogously, the evaluation period was from 1982 to 1987.
It was found that the resulting SM estimates of both models (based on the reference datasets EWEMBI and W5E5) reveal a satisfactory agreement on the basin scale. Besides, the general seasonal behavior in the models is reproduced in a satisfactory way with respect to the GLDAS SM, particularly at the basin scale. On the basin scale the NRMSE is 0.44 for the EWEMBI and W5E5-driven model, respectively. In addition, a r-value of 0.81 is obtained for both models. The basin NSE is 0.65 for the model forced with EWEMBI and 0.64 for the one based on W5E5.
On the subbasin scale the median (considering the performance of all individual subbasins) NRMSE is 2.20 ($r = 0.62$) for the EWEMBI-driven model and 2.24 ($r = 0.62$) for the W5E5 counterpart (see Fig. 4.4). However, on the subbasin scale a deterioration in all evaluation metrics can be observed, which is mainly due to the very poor performance of some subbasins. While the deficiencies are counterbalanced on the basin scale, some subbasins proved to produce much stronger seasonal cycle, which were not found in the GLDAS estimates. The effect is reflected in the weaker NSE results on the subbasin scale.
The results of both variables (ETA and SM) used in the cross-validation were moderate to very good in the catchment for both models. In summary both models performed equally during all steps of the evaluation procedure. A significant deterioration from EWEMBI to W5E5 was not found and most of

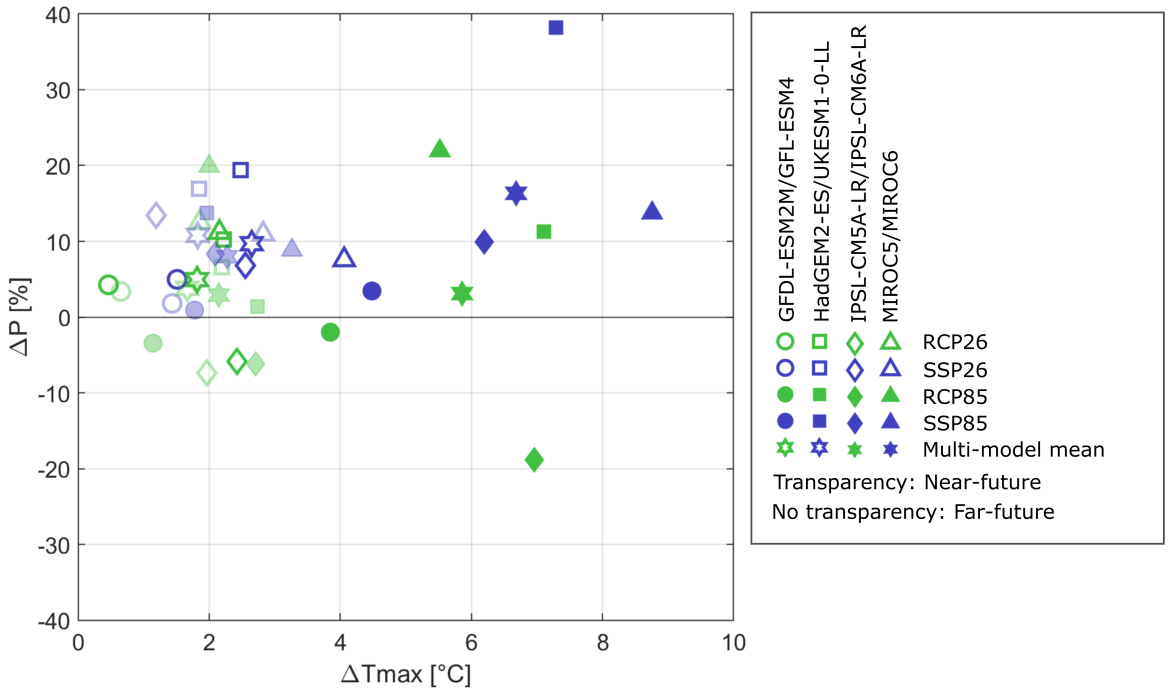

Figure 4.5: Mean projected meteorological changes of $Tmax$ & P for the near (2020-2049) and far (2070-2099) future periods, eight GCMs (four per ISIMIP phase), both scenarios and both ISIMIP phases compared to the baseline (1971-2000). $\Delta Tmax$ refers to the mean annual daily maximum temperature change. Analogously, ΔP corresponds to the mean annual precipitation change. Green - ISIMIP2, blue - ISIMIP3; filled/non-filled - high or low emission scenario; fill/outline transparency - near or far future; symbol - individual GCM (note, that the same symbol in green & blue indicates the respective predecessor/successor GCM. MMM refers to the multi-model mean in a specific period.

the discrepancies in the water balance were present at parts of the headwater subbasins. Both models were able to capture the contrasting climate periods with slight reductions in the quality in dry or cold phases. No notable differences were determined in the cross-validation of the water balance components SM and ETA. All steps were successfully performed and passed.

4.3.2 Projected Hydrometeorological Changes of ISIMIP2b and ISIMIP3b

Results are separately presented for the meteorological forcings, the hydrological projections, the importance of the meteorological controls as well as the sensitivity analysis.

4.3.2.1 Differences in Meteorological Forcing

In the following we highlight the main differences of the two GCM generations with respect to the projected temperature and precipitation changes. Changes are represented as anomalies to the baseline for two future periods (see section 4.2.5.1) under both scenarios. Temperature changes are evaluated as mean annual daily maximum temperature anomalies (from now on $\Delta Tmax$) rather than mean temperatures. Mean daily temperature changes are represented as proxy derived from $Tmax$. Precipitation changes ΔP refer to relative anomalies of the mean annual sum. Fig. 4.5 provides an overview of the meteorological projections of all scenarios, ISIMIP phases, GCMs and future periods.
The newer generation of GCMs produce generally higher temperatures as their former counterparts (comparison of equal blue and green symbols). In contrast to the high emission scenario, the temperature rise of an individual GCM between near and far future is weaker under the low emission scenario (same symbol, but different outline intensity). Under the low emission scenario the GCMs range from +1.5°C to +4°C temperature increase until 2099 under I3, compared to +0.46°C to +2.4°C under the old generation. Most GCMs scatter around +2°C, without any significant difference between the two periods under RCP/SSP26. In contrast, under the high emission scenario the temperature rise ranges from +3.8°C to +7.1°C for RCP85, and 4.5°C to 8.8°C for SSP85. This underlines the higher climate sensitivity of the new generation of climate models in the study region. Comparing the warming rates in the far future on the GCM development level for the radiative forcing of 8.5W/m², differences range from -

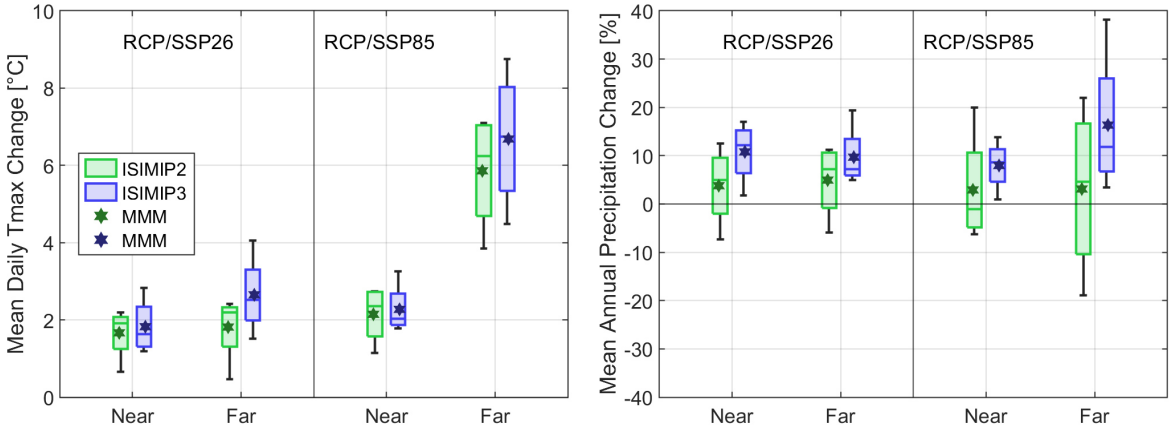

Figure 4.6: Evaluation of mean projected temperature and precipitation changes compared to 1971-2000. Left panel: $\Delta Tmax$ change of both GCM generations, represented as multi-model mean (MMM), indicated by the hexagon and GCM spread (boxplot). Right panel: ΔP change with same illustration as for $\Delta Tmax$. Near - 2020-2049, Far - 2070-2099.

0.77°C (predecessor GCM computes higher temperatures) to +3.2°C (successor GCM produces higher temperature). In average, the warming of the new climate models is around 0.82°C higher in the far future period under the high emission scenario (MMM of +5.86°C compared to MMM of 6.68°C). The MMM of the low emission scenario for I2 is +1.82°C and +2.65°C for I3 in the far future period. Those results are in line with a generally higher climate sensitivity of CMIP6 GCMs compared to CMIP5 GCMs (Meehl et al., 2020).

Precipitation anomalies show a higher disagreement between both climate generations. Especially the old generation of GCM is characterized by diverging and contrasting precipitation patterns. In comparison, under I3 the GCMs consistently exhibit positive precipitation anomalies, albeit with strong deviations in their magnitudes. Under SSP85 in the far future period the projections range from +3.5% up to 38.2%, compared to -18.9% to 22% under RCP85. In general, the differences in magnitude between the near and far future for each GCM are less pronounced under RCP26/ SSP26, and more distinctive for RCP85/ SSP85, respectively. The projected change represented by the MMM is small under both RCP scenarios with values of +3.1% and +5% in the far future for RCP85 and RCP26, respectively. For the respective scenarios of I3 we observe an increase in precipitation of +9.7% (SSP26) and +16.3% (SSP85). In summary, the positive precipitation anomalies reflected by the various MMMs are enhanced under I3. Besides, the inter-model range of I2 reveals a stronger spread (uncertainty) of precipitation anomalies for all future periods. The projected $\Delta Tmax$ anomalies present comparable inter-model ranges in all scenarios under I2 and I3 (Fig. 4.6). The discrepancy in the projected precipitation anomalies of the two generations arises mainly from specific regional contrasts, which are further addressed in the discussion section.

Moreover, deviations are found in the seasonal projections of precipitation. While both GCM families show an increase from October to May/June, they differ in late Summer and early Autumn. The MMM of the I3 ensemble does not significantly differ from the historical period in this period. However, in I2 the MMM shows strong declines of up to 30%. With respect to the seasonal patterns, all effects are more distinctive in the respective high-emission scenario and the far future period (See A.1, A.2).

4.3.2.2 Hydrological Impacts

The hydrological models were forced with the meteorological inputs of the different GCMs and scenarios until 2099. However, due to differences in the time of departure from the historical to the projection period between the two GCM generations (2006 and 2015), different time series lengths were available for the trajectories. The impact models for I2 and I3 are based on the individual parameter sets from the respective reference datasets (EWEMBI or W5E5). Here, we present the results of the climate impact study on various hydrological components on the basin scale. Discharge results are thus presented for gauge Uchterek.

Discharge

A summary of the results is provided in Fig. 4.7. In the far-future period the MMM of the annual mean flow is projected to change by +3.6% under SSP85 (+6.3% SSP26) and -0.7% under RCP85 (+5% RCP26). However, we observe significant seasonal shifts towards an earlier onset of the high flow period. Under the high emission scenario, peak flow is shifted by one month from June to May with a high agreement between the MMM of both, the I2 and I3-driven models for the far-future period. While Spring flow is projected to increase drastically under RCP26 (31.6%) and RCP85 (82.5%), Summer flow trajectories show significant decreases under the old generation of GCMs (-3.7% for RCP26 and -28.5% for RCP85).

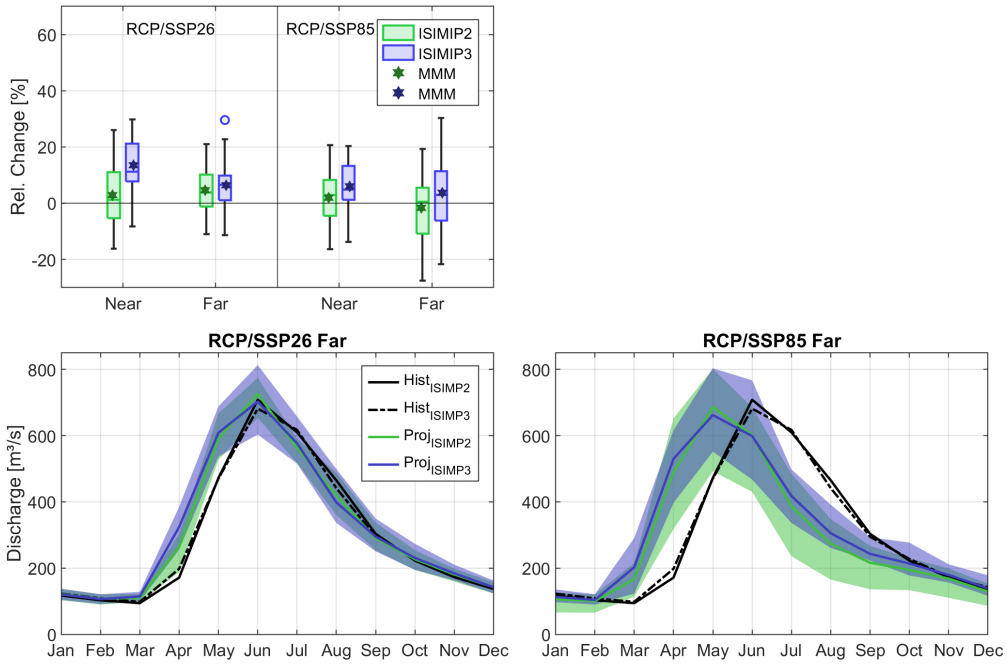

Figure 4.7: Mean annual and long-term seasonal discharge projections at the basin outlet (Gauge Uchterek) compared to the historical period (1971-2000). Upper left panel: Mean annual projections and model spread for two different future periods for both GCM families and both scenarios. Bottom panels: Seasonal discharge projections for the far future period under RCP/SSP26 and RCP/SSP85.Near - 2020-2049, Far - 2070-2099.

The I3-forced models exhibit similar patterns. Spring flow is projected to increase by 26.2% and 80.2%, respectively for SSP26 and SSP85 compared to a reduction of -2% (SSP26) and -24.4% (SSP85) in Summer.

The contrasting seasonal shifts in Spring and Summer have compensating effects with respect to annual mean flow changes. Despite an overall strong agreement in the seasonal MMM, both ISIMIP phases reveal differences regarding their uncertainties (see uncertainty bands in Fig. 4.7)). The agreement is detected in both scenarios in the far-future period. The uncertainties become larger with time and are more pronounced for the high emission scenarios. Besides, the lower boundary of the inter-model range is smaller under I2 in the high emission scenario (far-future) throughout the year.

Evapotranspiration

Stronger deviations between I2 and I3 are found in the ETA projections. Higher mean annual ETA anomalies are observed for I3 in both scenarios and both periods (expressed as MMM). We find a discrepancy of around 6% in the ensemble mean within the low emission scenario (Fig. 4.8) in the two periods. Under the high emission scenario the discrepancy between the I2 and I3-driven models increases with time. This leads to differences of 14% at the end of the century in the MMM (+9.32% I2 and +23.97% I3).

It becomes apparent that the reason is mainly due to negative Summer anomalies under RCP85 while the newer estimates are positive (-9% in MMM of I2 and +5% I3) for the far-future period. The effect is further amplified due to higher ETA estimates under SSP85 from March to June. The discrepancy of Summer ETA indicates differences in the available SM within this period (see also Fig. 4.9 lower right panel). The impact is likely to be induced by the deviations found in Summer precipitation of I2 and

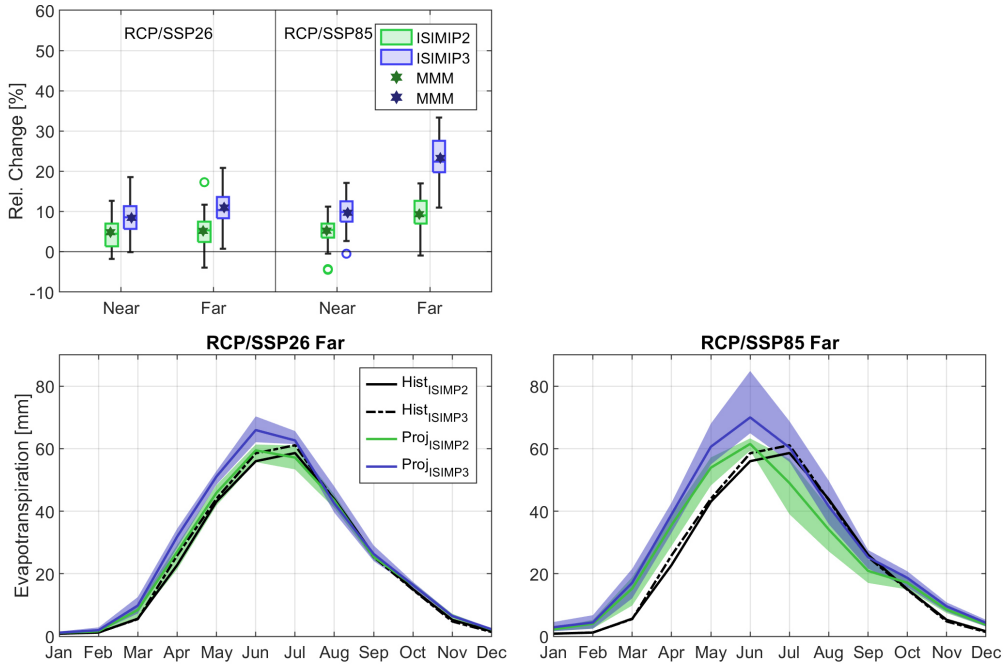

Figure 4.8: Mean annual and long-term seasonal actual evapotranspiration (ETA) projections at the basin scale. The panels information are identical to Fig. 4.7.

I3 (Fig. A.2). For example in the far future period the relative ETA changes from June to September under RCP85 are -9.8% compared to a reduction of -27.5% in precipitation (model median). In the same period under SSP85 precipitation increases by +1.8% supported by an ETA raise of +4.8%, which is almost in balance. Similar patterns also occur under the low emission scenario, although not as distinct. The results are illustrated in Fig. 4.8.

Soil Moisture

Fig. 4.9 illustrates the projected SM changes in the Naryn Basin. Both, I2 and I3 projections agree in their overall tendency of a significant decline in SM under all scenarios and periods. However, discrepancies are found in the uncertainties and magnitudes of the detected patterns. The projections of the older GCM family constitute larger uncertainties with less negative seasonal anomalies in the far-future period in the low emission scenario. In contrast, the negative seasonal anomalies in the high emission scenario are more pronounced than for I3. In particular, Summer and early Autumn are characterized by a significant decline in the MMM in all periods and scenarios under both climate model phases. For SSP85, available Summer SM is projected to decrease around -33% (far-future) which is in line with a decrease of -36% for RCP85. In the low emission scenario (again far-future) Summer SM projections reveal a decrease of -17% and -12% for SSP26 and RCP26, respectively. Similar reductions for the far-future are observed in Autumn, with MMM SM decreasing by -27% under SSP85, -38% under RCP85, -18% under SSP26 and -14% for RCP26. Positive SM anomalies are detected in April under both ISIMIP generations in the high emission scenario and far-future period (+11% MMM of RCP85 and +15.8% MMM of SSP85). Annual SM anomalies under RCP26 remain relatively constant from near to far-future (around -10% in MMM), while a marked decline under SSP26 is found (from -9.5% to -14.3%). In the high emission scenario, the MMM of both, the I2 and I3-driven models project a negative SM anomaly of -13% in the Near-future. For the far-future an even more distinct SM decline of -18% under SSP85 and -26% under

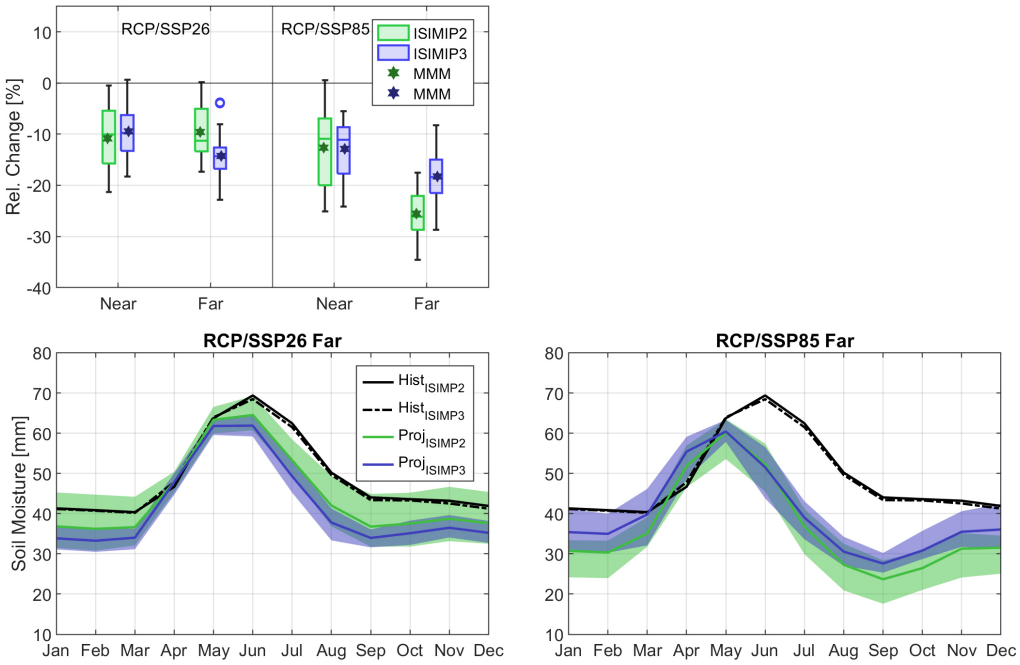

Figure 4.9: Mean annual and long-term seasonal soil moisture (SM) projections at the basin scale. The panels information are identical to Fig. 4.7.

RCP85 is detected, respectively. Besides, the lower boundary of the SM projections under RCP85 is -35% (around -29% for SSP85). The lower reductions under SSP85 are in line with the significant precipitation rise (see Fig. 4.6) compared to the I2-derived counterpart.

4.3.3 Meteorological Controls of the Soil Moisture and Evapotranspiration Feedback

A PC system was established to evaluate the relevance of meteorological changes on the SM-ETA feedback and to determine differences in the feedback between I2 and I3. As the basin is generally water-limited, SM anomalies control ETA anomalies (positive relationship), while in energy-limited regions the coupling would be reversed and ETA anomalies would lead to SM variations (negative relationship) (Berg and Sheffield, 2019). For this reason, we calculated the PC between ETA and SM, rather than ETA and precipitation (which might be suitable as SM proxy). Monthly anomalies of the MMM for the two periods AMJ and JAS were used for the evaluation. The two periods together comprise the vegetative season. Due to hydrometeorological differences within the vegetative season, the season was split to identify potential contrasting patterns. The PC results are supported by standard Pearson Correlation values in order to highlight the magnitude of confounding effects.

The results of the PC analysis are illustrated in Fig. 4.10, where the upper four panels refer to the period from May to June and the lower four to June to September. In the AMJ period a strong positive relationship between temperature and ETA is found for the near-future within all scenarios under both ISIMIP phases. Here the effects of precipitation and SM are small (grey crosses versus colored bars), which indicates a sufficient water availability in this period in the near-future to cover the ETA demand. This is in agreement with the constant to slight increase in precipitation in these months (Fig. A.2). Vice versa, the SM control on ETA in the near-future of the AMJ period is rather small (<0.5), albeit

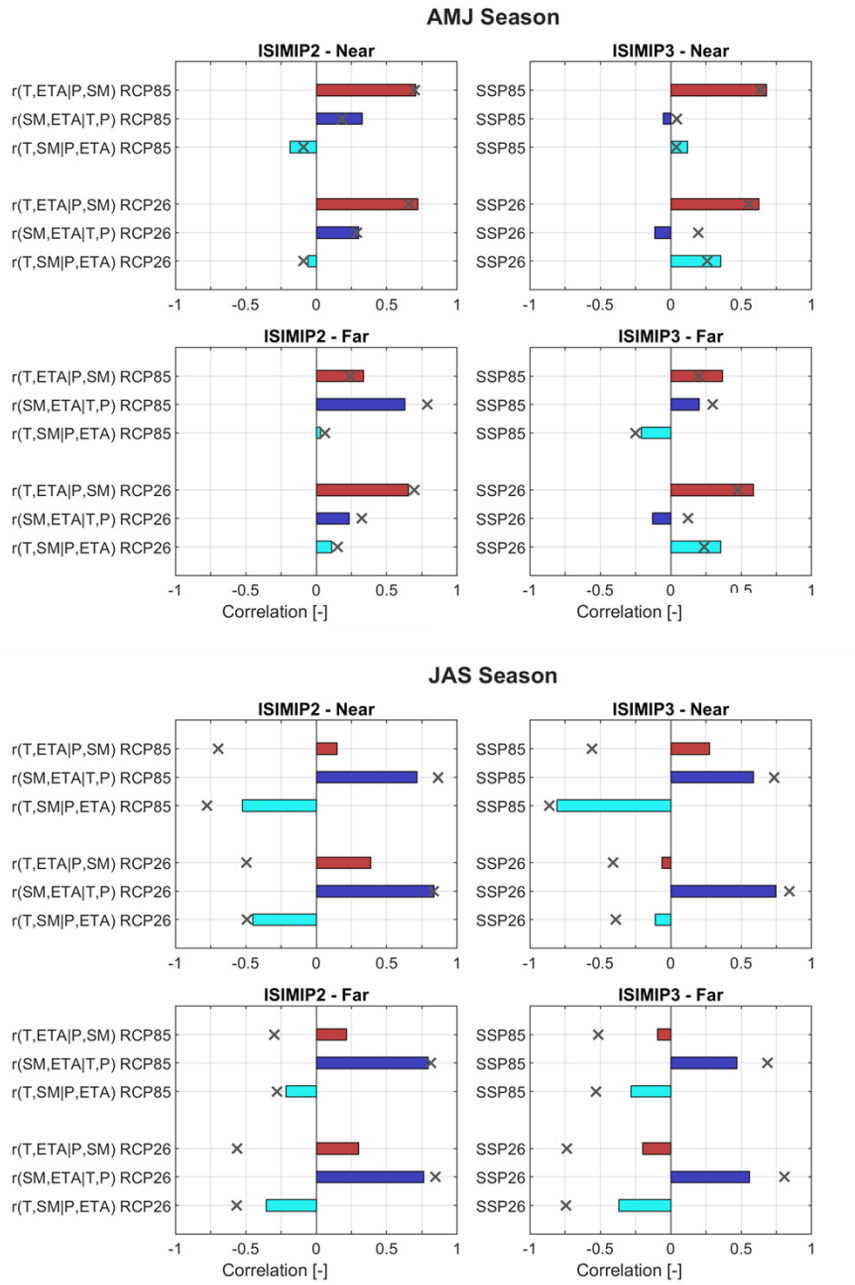

Figure 4.10: Relationship of monthly SM and ETA anomalies with temperature (T) or soil moisture (SM) anomalies for the two ISIMIP phases and two scenarios during the two periods April to June (AMJ) and July to September (JAS), which together represent the vegetation season.Shown in red, partial correlation of ETA and T while controlling for P and SM. Shown in blue, partial correlation of ETA and SM while controlling for T and P. Shown in cyan, partial correlation of SM and T while controlling for ETA and P. Grey crosses indicate the corresponding Pearson correlation coefficients. $r(T, SM|P, ETA)$ indicates the PC between T and SM controlling for P and ETA.

higher under I2 which is confirmed by higher precipitation projections causing a slightly higher water availability under I3 (Fig. A.2). Confounding effects in the ETA control are relatively weak and most pronounced in the SSP26 near-future relationship of SM and ETA. The ETA variance is therefore explained to a large degree of temperature and precipitation. We can already see a strong decline of the importance of temperature on ETA anomalies under the high emission scenario in the far future (I2 and

I3). This implies an increase in the positive SM-ETA coupling and a potential stressor for soil water. A cross-validation with Fig. 4.9 shows that this mostly refers to June where soil desiccation under RCP and SSP85 is substantial in the far future. In contrast to I3 the I2 projections of RCP85 the SM-ETA feedback increases significantly while under SSP85 the relationship of temperature and SM with ETA are close with small PC values (<0.5). Under the low emission scenario (far future) ETA variations are still controlled mainly by temperature, with PC magnitudes similar to the near future under I2 and I3.
The AMJ period coincides with the rainy season and the PC values of temperature and SM are predominantly weaker as for ETA. Nearly all PCs of temperature and SM are below 0.5 (all scenarios, periods and both ISIMIP phases) or even below 0.25. Under I2 no strong changes in the behavior of SM and temperature are found between near and far future (both scenarios). A similar development is shown under I3. However, the contra-intuitive positive association of temperature and SM is largely mitigated by the fact that the strengths of the relationships are insignificant and the signs heterogeneous. In the high emission scenario of I3 water-limitation becomes more apparent through a negative relationship (although weak) between SM and temperature in the far future. This is in accordance with the SM-ETA coupling in the AMJ period, where the T-ETA association is substantially decreased and SM becomes more important in the far future.
The generally drier JAS season is characterized by SM shortcomings and pronounced water-limited conditions. This is also reflected in the PC values of the JAS season. The SM-ETA feedback is strongly positively correlated in the near and far future of all scenarios under I2 and I3. The actual Pearson correlations between temperature and ETA show strong negative relationships, which would imply the unintuitive fact that rising temperatures are associated with a decline in ETA (again all scenarios of I2 and I3 and far and near future). However, most of the variance can be explained by the effects of precipitation and SM. Hence, the corresponding PCs of temperature and ETA are much weaker in absolute values than the Pearson correlations. For the I2 cases and the near future of the SSP85 the negative behaviour is even reversed (positive sign of PC). While under both I2 scenarios the dominant SM control of ETA anomalies remains constant from near to far future, the positive feedback is marginally reduced from near to far future in the I3 counterparts. This is especially in line with the precipitation projections in the JAS period of the high emission scenarios, which have a significant drop under I2 compared to I3 and is also visible in the SM anomalies of the high emission (Fig. A.2 and Fig. 4.9). The advanced water limitation under I2 compared to I3 is clearly visible in the corresponding PC values. Generally the higher the PC value of SM-ETA, the stronger the feedback and the more distinctive the water limitation. Accordingly, two entirely contrasting pictures can be revealed in the control of ETA variations.
The explanation of SM anomalies is again more ambiguous. For the high emission scenarios (I2 and I3) of the near future a substantial portion is explained by temperature anomalies. Both show a strong negative relationship indicating a strong SM decrease with rising temperatures. The negative relationship is also present in the far future (both scenarios) as well as the low emission scenarios of the near future, yet not as notable (<0.5). For the near future of the low emission scenarios the missing strength is attributed to the smaller positive temperature anomalies compared to the high emission scenarios. The reduced dominance of temperature in the far future is probably caused by an already advanced desiccation of soils (water limitation higher as in near future). The selection of ETA as one of the confounding variables for SM might be doubtful, since SM controls ETA and not vice versa. At least it becomes apparent that the confounding effects under I3 here are relatively high. Fig. A.4 provides an additional example of sensitivities of multiple variables to annual temperature anomalies.

4.4 Discussion

In this study, we investigated how climate impact assessment agrees and diverges under a subset of I2 and I3 GCMs. To produce a reliable and robust hydrological model for the impact assessment we followed the recommendations proposed by Krysanova et al. (2018, 2020).

4.4.1 Evaluation Test

The evaluation procedure produced meaningful and comparable results for both, the SWAT model forced with EWEMBI as well as with W5E5. All steps of the evaluation procedure were successfully passed and the models performed satisfactorily to well in most cases under contrasting climate periods, on multiple temporal scales, for various GOF criteria as well as for ETA and SM in addition to discharge. In detail, however, there were differences at the subbasin level, where individual gauges showed a poorer performance than for example at the basin outlet. Thus, two similar parameter sets were identified for both reference datasets serving as base for the impact assessment. GOF results were on average around 0.75 for NSE, 0.82 for KGE and 9.4% for PBIAS in the basin (considering all gauges). Moreover, warm and wet periods were better reflected (NSE_{wet} 0.8, NSE_{warm} 0.75) than cold and dry periods (NSE_{cold} 0.71, NSE_{dry} 0.70). Similar difficulties and weaker model performances are for example also reported in Vaze et al. (2010) for a catchment in Australia.
However, due to the crucial role of nivo-glacial processes in CA catchments, more emphasize should be placed on these processes in future studies. E.g. considering snow cover as in Hofmeister et al. (2022), glacier mass balance as Wortmann et al. (2018) did or snow water equivalent data (see Tuo et al. (2018)) in the calibration process, if available, could produce an even more consistent and reliable picture of the system. Deficiencies were mostly observed in the high-elevated headwater part (see Tab. 4.3 and Tab. 7.4), where NSE values were in average only around 0.6. These are mainly due to model weaknesses with regard to the representation of glaciers. We therefore agree with Saks et al. (2022) that future studies should consider these aspects and evaluate expected peak flow more closely.
It must also be mentioned that the approach followed here logically does not necessarily guarantee that the robustness of the model adequately reflects all projected changes. This is particularly the case for the contrasting climatological periods, since rates of change are likely to be more pronounced in the future on the one hand, and on the other hand the relatively short historical period (restricted due to data availability) may not adequately reflect the influence of persistent changes over a longer period of time. these deviations and assumptions that might be only partially representative are described in detail in Stephens et al. (2020).

4.4.2 Meteorological Changes

Projected temperature raises agree throughout the year in all periods under the new and old generation of GCMs. In contrast, precipitation disparities have shown to be one of the major differences between I2 and I3. While precipitation changes observed under I2 are mostly in line with other related studies (Lutz et al., 2014; Luo et al., 2018; Huang et al., 2014), the higher increase under I3 has not yet been comprehensively and extensively addressed in the literature. The same applies to the associated effects on the hydrological cycle. Discrepancies in the precipitation projections between I2 and I3 arise mainly from regional contrasts in the basin. Under SSP26 and SSP85 a continuous rise throughout the whole basin is projected, with a marked characteristic in the Kekerim subbasin. I2 contains two counteracting patterns, the Western part of the basin receives more precipitation, which is amplified in the high emis-

sion scenario, while on the other hand the Eastern (and headwater) part is characterized by a decline in precipitation. The decline is also even more significant in the high emission scenario and for the far future period.

We provide a first estimate of hydrological impacts triggered by precipitation changes under both ISIMIP generations. The deviations in projected precipitation anomalies (consistent positive precipitation anomalies with MMM of +16.3% under SSP85 compared to +3.1% under RCP85 in the far-future) come along with a higher climate sensitivity in the new ensemble of climate models (warming rate in average +0.82°C higher until 2100 in high emission scenario). However, a lot of the current research focuses on the higher climate sensitivity of CMIP6 models including the potential or necessity to constrain warming trends of models that tend to be on the upper edge (Tokarska et al., 2020; Sherwood et al., 2020). As only a subset of GCMs was available and included, future studies might extend the efforts with a larger sample.

Further, although projects like ISIMIP add a huge value for climate impact studies the generally coarse resolution of the GCMs might not be suitable for all studies and regions. Especially for small basins the resolution could be insufficient. Hydrological scale patterns could be missed, this problem might also affect the higher elevated parts of the study area, where the performance of the models (for all variables) was deteriorated. Meteorological processes of individual mountain ranges might be inaccurately represented. However, data limitations make it difficult to evaluate this thesis in detail. In any case, we assume that Regional Climate Models are favorable and would add value to future studies if possible to include.

We also would like to at least mention that (depending on the methodology and data availability) an assessment of the GCM performance might be reasonable in advance of the actual study to for example exclude poor-performing climate models or to take into account a weighting approach to reduce their importance within the results. If for some reasons GCMs are not applicable in a study, we would also like to emphasize to consider perturbations of the observed climatology to draw conclusions of the future hydrology based on these synthetic time series. Examples can for example be found in Rasouli et al. (2019b,a)

4.4.3 Hydrological Changes

The projected MMM of all hydrological variables generally agree between I2 and I3 on the annual and seasonal scale. MMM discharge projections were similar across both ISIMIP phases. They showed a significant shift in the flow regime. Peak discharge is projected to occur one month earlier (June to May) under the high emission scenarios by the end of the century. This is consistent with what was reported by other authors (Gan et al., 2015; Didovets et al., 2021). The shift is mainly driven by substantial increases in Spring snowmelt followed by a strong decline in Summer (A.3). The summer deficit agrees with the water shortage found for SM. Summer deficits of discharge and SM could cause and promote water stress in the region. This is particularly evident since summer water availability directly affects irrigation demand during the vegetative season and water storage for energy production. Especially for higher projected temperatures and thus an increase in potential evapotranspiration estimates, future water demand could significantly rise and constitute an additional stressor.

Moderate deviations were found in the ETA trajectories, with stronger increases projected under I3 (difference of 6% in MMM of I2 and I3 in low emission scenario and 14% in high emission scenario at the end of the century). The higher estimates are likely to be triggered by a combination of higher projected temperatures (higher energy-availability, Fig. 4.6) as well as weaker SM reductions (higher water-availability, Fig. 4.9). ETA increases referred mainly to the first half of the year, the period where the most significant

precipitation increases were observed. The period was also characterized by a strong increase in snowmelt (Fig. A.3), an additional source of water. The described impacts are more pronounced under the high emission scenario and the far-future. The ETA trajectories are consistent with recently reported trends for CA and globally. And the impacts could significantly contribute to future desertification, one of the major risks in CA (Huang et al., 2020a; Ma et al., 2021).

Notable SM deficits in Summer (-36% in MMM of RCP85 and -33% under SSP85 in the far-future) and Autumn (-38% in MMM of RCP85 and -27% under SSP85 in the far-future) were projected for both GCM generations, which could act as potential stressor for the agricultural productivity. Summer and Autumn SM deficits could thus induce and promote future agricultural droughts. Droughts that promote land degradation are already present in CA (Zhang et al., 2018). According to our results, indications of negative changes in the vegetation dynamics, caused by soil moisture deficits, which became particularly pronounced after the 90's (Li et al., 2015; Deng and Chen, 2017), could thus also represent a substantial future risk in the basin. The crucial role of SM for drying in CA is also illustrated in Hu et al. (2018), who used a correlation analysis to demonstrate the importance of SM. Although most of the studies available focus on whole CA, solely on the historical period or use related variables such as NDVI (or variants of it) and total water storage (Jiang et al., 2017; Xing et al., 2022; Zhou et al., 2015; Peng et al., 2021), our study supports to assess these future risks at the local level. The negative anomalies were slightly smaller under I3 due to an attenuation caused by a raise in precipitation (annual decrease in MMM of -18% for SSP85 and -26% for RCP85 in the far-future). Annual SM shortcomings tend to be stronger for I3 under the low emission scenario and smaller under the high emission scenario compared to I2 by end of the century. Summarized, the findings indicate the vulnerability of the region to SM shortages.

Our PC results provide evidence that shifts in the role of the meteorological drivers can be expected from near to far future as well as between I2 and I3 in both periods (AMJ and JAS). This leads to partly significant deviations and changes in the SM-ETA coupling (from near to far under both GCM generations as well as between them and in both seasons). We find clear differences in the strength of the SM-ETA relationship in the two seasons AMJ and JAS, where the latter is considerably more water-limited. While under I2 the pressure on soil moisture, reflected in the strong control of ETA, remains relatively constant from near to far future (both scenarios), we see a attenuation of the strongly positive relationship under I3 (both scenarios) from near to far. This development represents the different precipitation trajectories, where the I3 JAS rainfall does not show the drop which is especially visible in the RCP85 and thus likely leads to a mitigated negative SM evolution under I3, albeit still severe. The strong water limitations in JAS are expected to exacerbate the pressure on water resources to meet the demand. The resulting implications for example on land degradation or the occurence of droughts are similar to those mentioned already above. However, the PC results underline the assumption of a reinforced SM-ETA coupling udner both generations of climate models, yet with a marginal shift towards less water-limitation under I2 in JAS. The strength of the SM-ETA coupling could be seen as indicator for vegetation conditions (Ibrahim et al., 2015) and in terms of a positive coupling reduced moisture availability (Berg and Sheffield, 2019). The strong positive future relationship between the two variables in the basin is also in accordance what is expected roughly on the global scale, when analyzing GCM outputs (Berg and Sheffield, 2018; Berg et al., 2015; Dirmeyer et al., 2013). A limitation which applies not only to our study but to many hydrological impact studies are missing atmoshperic feedbacks which are connected to the ETA-SM coupling, such as evaporative cooling.

For the generally wetter AMJ season we can already identify a positive increase in the SM-ETA relationship under I2 and I3 from near to far future, particularly in the high emission scenarios. This provides evidence that the water-limited regime could further advance in the region, despite precipitation pro-

jection in the far future tend to be positive for I2 as well as I3. In other words, the importance of SM variations to explain ETA variations increases. This shift in the future importance of SM is enhanced by the decreasing dominance of temperature to explain ETA variations under the high emission scenarios. One major deviation between I2 and I3 is found in the magnitude of the SM-ETA feedback in the far future of the high emission scenarios. Under RCP85 the explanatory power of SM replaces the one of temperature as major controlling factor. In contrast, although the importance of SM increases with a simultaneous decline of temperature the final magnitudes in the far future are relatively close to each other (with small values <0.5). We assume that this is caused by the slightly larger increase in precipitation under I3 in AMJ attenuating the increased demand of rising temperatures. However, it is expected that further increasing temperatures would also tip the relationship and approach the conditions of I2.
Stationarity was assumed for the vegetation season, which was used for comparability reasons, although the period can vary with land use (and thus in space) (Zhang et al., 2018; Zhou et al., 2015). As the climatic shifts are likely to shift the vegetative season as well, a constant period represents a simplified manner.
We only included one hydrological model in the assessment. An ensemble of impact models would further improve the robustness of the results. Structural differences in impact models can result in significant deviations in the projections and identify adequateness or inadequacy of impact models for different conditions (Kundzewicz et al., 2018; Wen et al., 2020; Su et al., 2016). Also land use changes were neglected, which are likely to affect processes such as runoff generation, albedo-associated mechanisms or translation and retention.

4.5 Conclusions

Our study shows that despite advances in the forcings of impact models, as a result of GCM, scenario or bias-adjustment developments, estimated impacts are in general concordance.
These are particularly evident in the projected shifts of the hydrological regime and a sharp increase in soil desiccation. We find a shift in peak flow of one month due to an earlier start of the melt season in the high emission scenarios. The earlier onset leads to an increase in discharge in Spring followed by a decline in Summer. Across both GCM generations and all scenarios, the impact assessment further indicates substantial soil moisture deficits, especially in Summer and Autumn.
However, disagreement between I2 and I3 is found in the evapotranspiration estimates, which are higher in the new generation. Temperature projections of I3 are consistent with the general recognition of being warmer than under previous generations (higher climate sensitivity). Rather large deviation between the two generations are found in the precipitation projections. Although consistently positive precipitation anomalies are observed under I3, the projected SM deficits are substantial under both phases, indicating the region's vulnerability. This underlines the sensitive interplay between the different components precipitation, soil moisture and evapotranspiration.
Deviations in the temperature and precipitation projections of I2 and I3 in the two examined seasons (AMJ and JAS) reveal significant effects on the SM-ETA coupling in the basin. Expressed as PCs we find enhanced water-limited conditions in the AMJ season at the end of the century, particularly in the high emission scenarios. The advanced importance of SM is directly shown as increase in the positive relationship between SM and ETA and indirectly through the decrease in the dominance of temperature on ETA. The inherently drier JAS season with its strong positive SM-ETA coupling remains constant under the near and far future of I2, while there is a slight decrease with time under the corresponding I3 scenarios. This indicates that the general and progressive water-limitation among both ISIMIP genera-

tions could be at least slightly weakened under I3, albeit it is still intense. Differences in the magnitudes of effects between I2 and I3 are likely to be caused by mitigating effects present under I3, mostly referred to positive or at least constant precipitation trajectories in the vegetative season.

Acknowledgements

We want to thank the ISIMIP project in which context the guideline for the evaluation test for impact studies was created, as well as for providing us the necessary data for our study. In addition we want to thank the German Research Centre for Geoscience for sharing observational data with us, without which it would not have been possible to conduct our study and Florian Betz for his critical discussions and regional expertise. The authors from TUM also want to thank the BMBF (Bundesministerium für Bildung und Forschung) for the funding of the "OekoFlussPlan" project (grant number 01LZ1802B).

Chapter 5

Sources, Propagation and Contribution of Uncertainty

5.1 Introduction

Complementing Chapter 4, the following chapter addresses several aspects associated with uncertainties in climate impact assessment and the regional effects of global climate change. Regional impacts of global climate change refer to the analysis of different global warming levels, namely 1.5°C, 2°C and 3°C, and the respective impacts on the hydrology of the Naryn Basin. The study provides an indication of what achieving the global warming values set out in the Paris Agreement could imply for the regional catchment hydrology. However, the translation of global warming levels into regional impacts is subject to a large number of uncertainties and therefore offers a different perspective to the general topic of uncertainties in impact assessments within this chapter. Besides, a traditional ANOVA is performed to quantify different sources of uncertainties along the modeling chain. Lastly, an analysis of parameter robustness and, hence, their transferability is conducted. Specifically, the robustness of parameters derived from an independent forcing dataset for calibration is investigated compared to parameterizations using the current GCM reference dataset for calibration.

5.2 Methodology

5.2.1 Uncertainty Associated with Reference Datasets

Supplementing Chapter 4, a more detailed analysis of the reference datasets and their implications on hydrological model calibration is conducted. First, annual basin precipitation and annual mean daily maximum temperature of the basin of both reference datasets are compared, followed by comparing daily precipitation and annual maximum daily precipitation distributions. Daily precipitation is fitted to a gamma distribution and annual maximum daily precipitation to a normal distribution. For this, all daily pixel values from 1979 (starting date of W5E5) to 2016 that are located in the basin are used. Furthermore, daily maximum temperatures of both datasets are compared with each other for each season (summer, winter, autumn, spring). Moreover, spatially distributed precipitation and temperature differences are provided.

This is followed by a discussion of two essential points, 1) implications, even minor, differences between reference datasets can have on calibration and 2), what does it imply, if the calibration of a hydrological

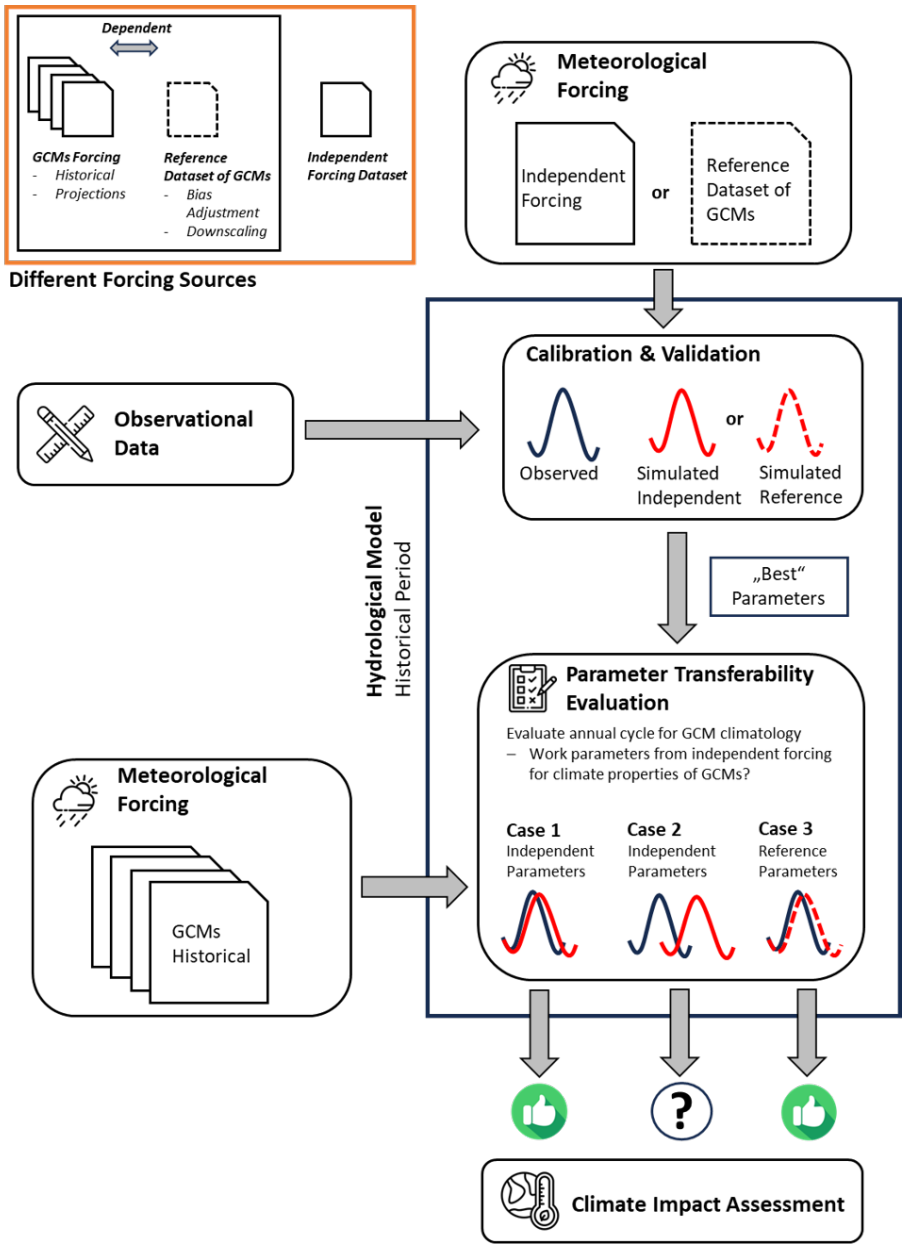

Figure 5.1: Flowchart of the methodology used to evaluate parameter transferability from an independent calibration to climate projections subject to downscaling and bias-adjustment procedures using a differing reference dataset.

model is based on a different independent meteorological dataset compared to the one employed in bias adjustment and downscaling of the climate projections. Point 2) is of tremendous importance and rarely discussed. In many studies, datasets are used to force horological models, which are different from those used by climatologists for bias correcting climate projections, e.g. Senent-Aparicio et al. (2017); Chaemiso et al. (2016); Oo et al. (2020); Zhang et al. (2007); Mango et al. (2011); Sharannya et al. (2018); Samavati et al. (2022); Marhaento et al. (2018); Dessu and Melesse (2012); Ma et al. (2023).

However, parameter transferability of a calibrated model under a new forcing might not hold. This might be especially true, when parameter uncertainties are not considered for impact assessment. This is a common case in impact assessments, which reduces uncertainties to those originating from climate

models and scenarios. The effects of the parameter transfer should, in the opinion of the author, at least be demonstrated and discussed. A straightforward yet reliable method to cross-validate the parameter set under different forcings is to compare the mean annual cycle of a variable of the following three cases: a) the observations, b) the simulations forced with the climate models based on the calibrated parameters not stemming from the reference dataset used for bias-adjustment of the climate models, and c) the simulations based on the actual but independent forcing used for calibration (and correspondingly the resulting parameters). If parameter transferability from c) to b) is a valid assumption, they should be close to each other. A fourth, but optionally, case would be employing the actual reference dataset used for the bias adjustment based on the calibrated parameter set. However, as the bias-adjusted climate data in the historical phase usually obtain the statistical properties of the reference dataset, the cycles over a sufficiently long period of time should be similar. Of course, as climate projections represent climate conditions statistically and not individual events, the time period should be long enough and the approach can be seen as an approximation. If the annual cycle of the simulations driven by the GCMs deviates strongly from the other two cases, it could indicate that parameter transferability is limited.
The idea of the parameter transferability evaluation is explained in Fig. 5.1.

5.2.2 Uncertainty Decomposition - ANOVA

ANOVA is used to decompose the main sources of uncertainty in the projections of hydrological variables. In our example, the sources of uncertainty can be attributed to three main factors, the ISIMIP or CMIP generation (I2 and I3), the GCMs and the scenarios (RCPs/ SSPs), respectively. The study is based on a three-way ANOVA which is widely used in climate impact (Hattermann et al., 2018; Huang et al., 2020a; Ismail et al., 2020; Vidal et al., 2016). The method relies on the total sums of squares (SST), a measure of the total variance of all factors. The SST is derived from the sum between the error sums of squares (SS) within a group and the between sample sums of squares:

$$\mathrm{SST} = \sum_{i=1}^{N_{\mathrm{GCM}}} \sum_{j=1}^{N_{\mathrm{SCEN}}} \sum_{k=1}^{N_{\mathrm{GEN}}} (Y_{i,j,k} - \bar{Y}_{ooo})^2 \tag{5.1}$$

where $Y_{i,j,k}$ represents the values for one particular GCM, scenario and ISIMIP phase, $\bar{Y}_{ooo}$ is the mean of the whole available data and N corresponds to the number of used GCMs, Scenarios and ISIMIP generations, respectively. SST can thus be split into three main effects and four interaction terms

$$\begin{aligned} \mathrm{SST} &= \mathrm{SS}_{\mathrm{GCM}} + \mathrm{SS}_{\mathrm{SCEN}} + \mathrm{SS}_{\mathrm{GEN}} \\ &+ \mathrm{SS}_{\mathrm{GCM*SCEN}} + \mathrm{SS}_{\mathrm{GCM*GEN}} + \mathrm{SS}_{\mathrm{GEN*SCEN}} \\ &+ \mathrm{SS}_{\mathrm{GCM*GEN*SCEN}} \end{aligned} \tag{5.2}$$

where SS_{GCM},SS_{SCEN} and SS_{GEN} represent the sum of squares of the three main effects, four GCMs, two scenarios and 2 climate model generations. The $SS_{GCM*SCEN}$, $SS_{GCM*GEN}$, $SS_{GEN*SCEN}$ and $SS_{GCM*GEN*SCEN}$ correspond to the sum of squares of the nonlinear or non-additive interaction terms. A multi-factor ANOVA distinguishes between main effects and interaction terms. The main effects refer to our three groups and the interaction terms describe nonlinear effects between the main groups. First-order interactions involves the dependency of two main factors while second-order terms describe three-way interactions. Further details can be found in Vetter et al. (2015).
The contribution of each factor is calculated on the annual scale for the two future periods, as well as on a seasonal scale. A subsampling scheme, proposed by Bosshard et al. (2013), is used to remove biases in

the contribution estimates caused by different sample sizes of the main factors (4 GCMs, 2 scenarios, 2 ISIMIP phases). The proposed scheme resamples main factors with a larger sample size (GCMs in our case) to match the factors with a smaller sample size. Thus, GCM pairs are created for the ANOVA calculation.
The factor classification merges scenarios and GCM developments between the two generations. For example, the predecessor and successor GCMs GFDL-ESM2M and GFDL-ESM4 are treated as one climate model in the classification. Correspondingly, RCP85 and SSP85 are considered as the same high emission scenario. However, in particular GCMs belonging to the same family but from different generations can differ greatly from each other. Accordingly, the approach allows to investigate model developments of a GCM family between generations. ANOVA was performed for future discharge, evapotranspiration and soil moisture anomalies.

5.2.3 Regional Impacts of Global Climate Change

It was further investigated how mean global temperature change, ΔT_g, affects the catchment hydrology. Following the global warming levels defined in the Paris Agreement, focus was put on global warming levels of 1.5°C, 2°C and moreover 3°C, a number, which is roughly related to current climate trajectories until the end of the century. An approach also known as *time sampling* (James et al., 2017). For the evaluation of how global climate change is translated into regional impacts, the ensemble mean (MMM) of the temperature anomaly of each GCM and scenario under CMIP5 and CMIP6 was calculated. It should be noted that for this purposes the global time series of the underlying GCMs was used rather than the regional time series, to adequately reflect global climate change. The time series were then smoothed using a 30-year moving average. The date of intersection of the MMM temperature anomaly series with the aforementioned global warming levels was then determined. This date was used to extract the corresponding ETA, SM and discharge values to quantify the regional hydrological response for the different warming levels under each scenario. It is noteworthy that the method is sensitive to the following issues.
First, the definition of the reference period has a substantial impact on the timing when a specific temperature level, such as 1.5°C, is reached and therefore on the hydrological response. Second, no common method exists how the date of exceedance should be exactly chosen. For example, while there is general consensus on using 30 year periods as representation of climate and thus for which a temperature level should be exceeded (expressed as a moving window of 30 years), it is rather arbitrary whether the first point of overshot of the 30 year window should be used or whether the climatology should exceed the global warming level over a longer period. Using an fictive but significant example, there might be a chance that the temperature trend after reaching a warming level is declining and the 30 year window does not reach the threshold temperature anymore. Besides, longer time scales (e.g. 50 years) might offer more robust results. Third, time-lagged hydrological responses are not considered, but only impacts visible at the date of overshoot. This might oversimplify the interaction of the climate and hydrological system. Last, path dependency of climate models is not considered, e.g. exceeding a threshold could be temporary or occur multiple times for a specific GCM trajectory (James et al., 2017). As the date at which a warming level is reached can vary strongly between GCMs not only different time periods are compared, but also the pathway how the warming level is reached is not directly considered (one GCM might slowly and gradually reach the threshold value while another reached it relatively abruptly after a long stable climate phase). The hydrological response could therefore vary significantly.

5.3 Results & Discussion

In the following results are presented for the evaluation of the reference datasets, the uncertainty decomposition and the effects of global temperature change on the regional scale.

5.3.1 Evaluation of the Role of Reference Datasets

In the following a comparison of the two reference datasets W5E5 (ISIMIP3) and EWEMBI (ISIMIP2) is performed, followed by an analysis of how large the hydrological impacts are, when parameter sets obtained from an independent forcing dataset are transferred to the actual reference dataset (used for bias-adjustment and downscaling of the climate models), alongside with the differences in the GCMs-driven hydrological models based on the different parameter sets. For this, the models calibrated for EWEMBI and W5E5 (see Schaffhauser et al. (2023)) are used and the originally-derived parameterizations of each dataset are tested for the respective reference data that deviates from their original calibration.

5.3.1.1 Comparison of W5E5 & EWEMBI

Results for the evaluation of the reference datasets, used in Schaffhauser et al. (2023) are provided in Fig. 5.2 and Fig. 5.3. From Fig. 5.2 it becomes obvious that the datasets are relatively similar and temperature and precipitation representations are comparable. On the annual scale (shown in a) and b)) only minor differences are visible, and both datasets show equal inter-annual precipitation variability expressed by the CV (0.17 for EWEMBI and W5E5). Slightly to moderate differences are recognized for the fitted distributions on daily precipitation (c)) and annual maximum daily precipitation of both datasets. EWEMBI data suggest higher probabilities for lower daily precipitation values and lower probabilities for daily precipitation greater than approx. 3 mm/day. For annual maximum daily precipitation the distribution peak of EWEMBI occurs slightly above 20 mm/day, while it is slightly below 20 mm/day for W5E5. The distributions reflect the continental climate of the basin with relatively low annual maximum daily precipitation values. The scatter plot in e), shows no tendency as to whether a dataset tends to overestimate or underestimate the maximum daily temperatures in a given season. Daily peak temperature, which was chosen to indicate whether differences in the snowmelt potential could be expected between the dataset, therefore indicates no significant discrepancy. The only small deviations between the two datasets would suggest that transferring a robust parameterization from one dataset to the other works relatively well.

This is also confirmed when looking on Fig. 5.3, where the spatial distribution of precipitation and temperature of both datasets is compared. In the upper panels mean annual precipitation sums are illustrated. It becomes evident that both datasets contain a similar declining east-west gradient. The center-left panel shows the corresponding difference map of both datasets for precipitation expressed as percentage. Positive values indicate W5E5 being wetter and vice versa. The mean difference is around ±7%, with a maximum deviation of -24% (EWEMBI wetter than W5E5). Annual mean maximum daily temperature differences (right-center panel) are mostly negative indicating slightly higher temperatures in the basin under W5E5, with an exception in the eastern headwater part in which higher temperatures under EWEMBI can be observed. On average the mean maximum temperature differ around ±0.17 °C. The lower two panels are similar as the ones for precipitation but show absolute values of mean daily maximum temperatures of both datasets. Both maps show agreement in the illustrated west-east gradient, with decreasing temperatures towards the eastern slopes and a second colder area in the south of the basin.

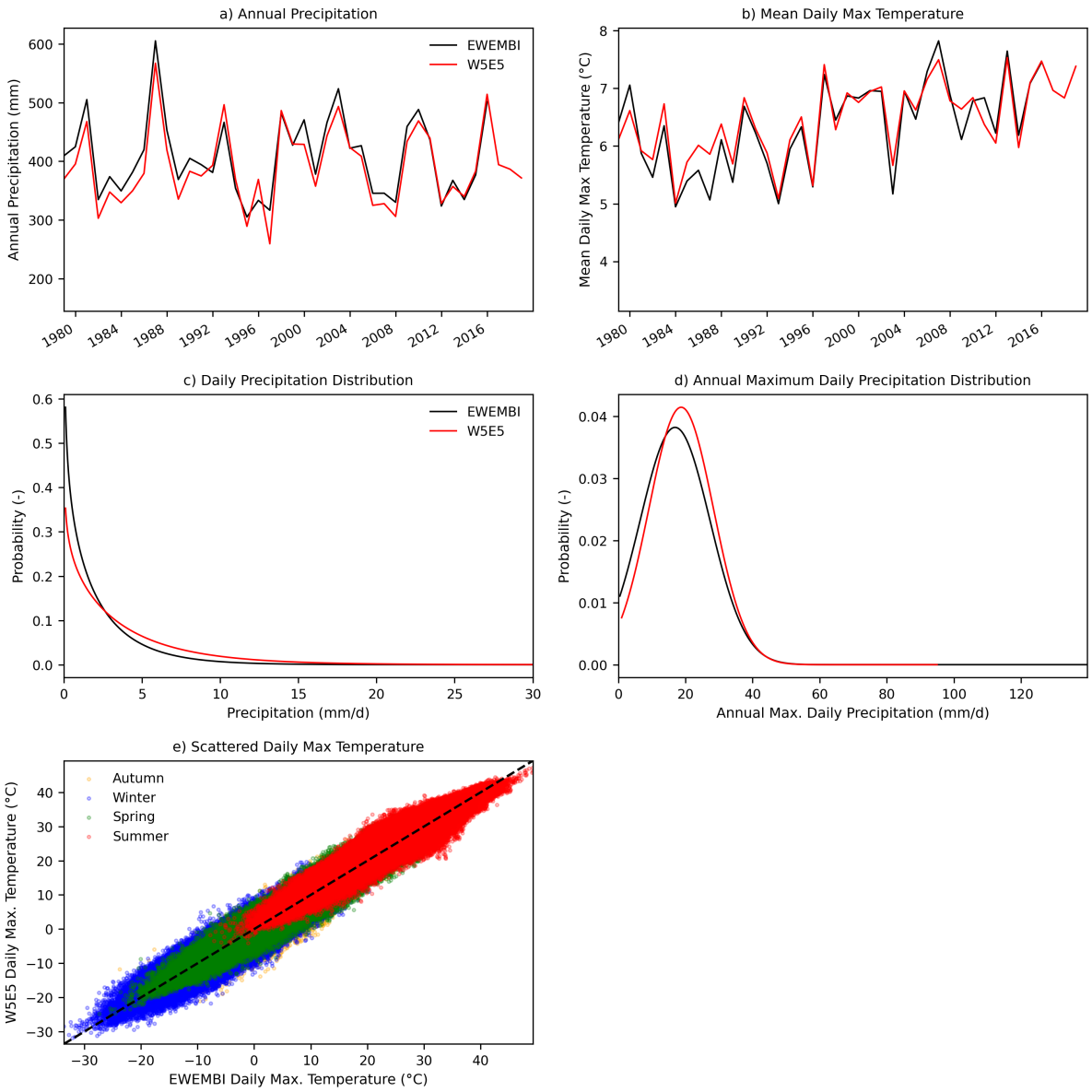

Figure 5.2: Comparison of ISIMIP2 (EWEMBI) and ISIMIP3 (W5E5) reference datasets used in the calibration of the two models in Schaffhauser et al. (2023).

Hence, across all comparisons the differences between W5E5 and EWEMBI are only marginal and a dominant agreement between both datasets can be observed.

5.3.1.2 Parameter Robustness Using Different Reference Datasets

In the following, it is evaluated how robust the parameterizations derived in Schaffhauser et al. (2023) using EWEMBI (and used for the projections of ISIMIP2) and W5E5 (and used for the projections of ISIMIP3) are. Given the dataset similarity shown before, it is assumed that the parameterizations therefore work relatively well under the respective different dataset and thus the projections of the different climate generation. Results are illustrated in Figure 5.4 and show how the annual cycle of the hydrographs change when the EWEMBI-derived parameters are used with the forcing of W5E5 and vice versa (left panel), as well as the impacts when the corresponding parameterizations are used under the

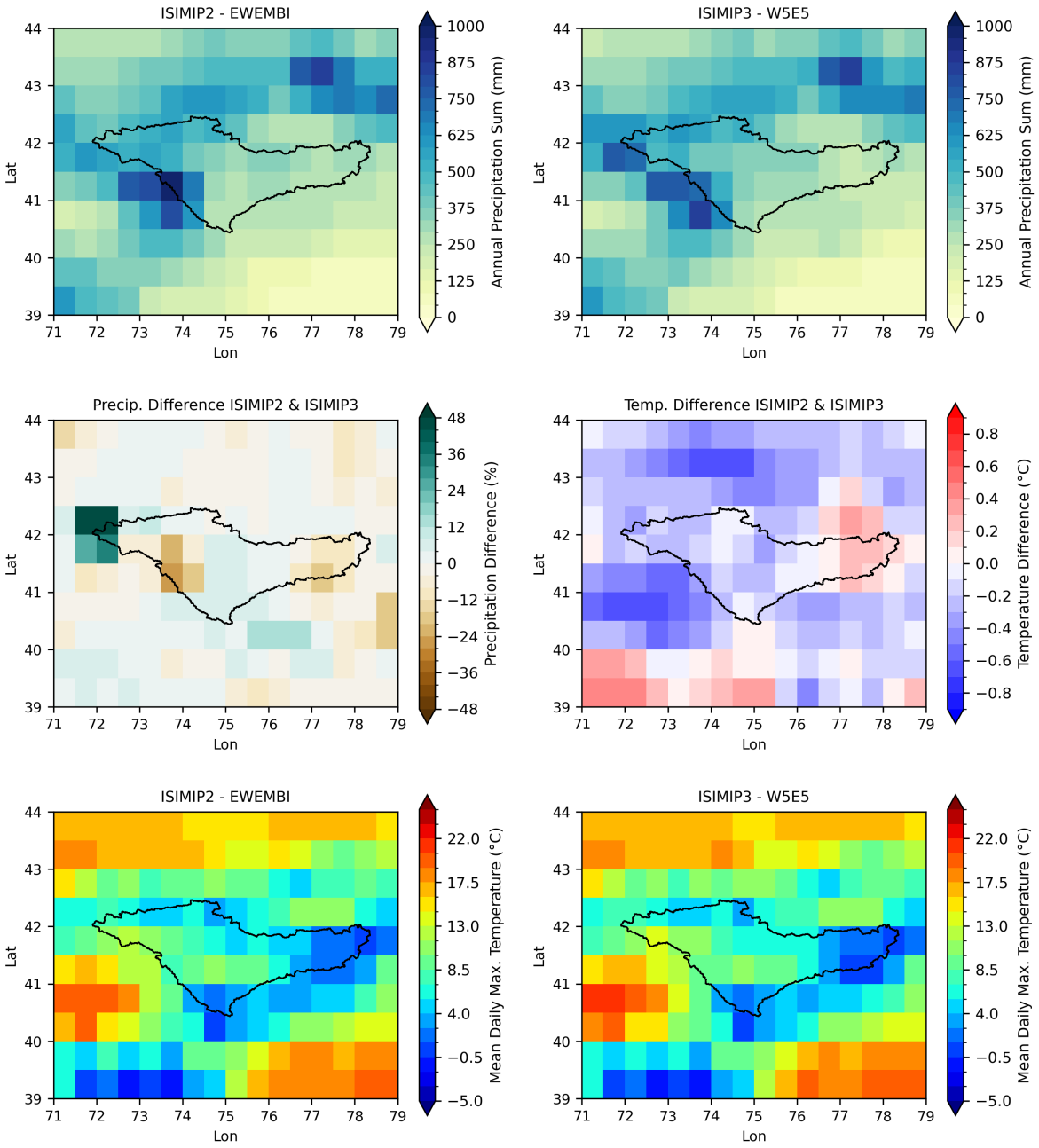

Figure 5.3: Comparison of ISIMIP2 (EWMBI) and ISIMIP3 (W5E5) reference datasets used in the calibration of the two models in Schaffhauser et al. (2023).

historical climate from ISIMIP2 and ISIMIP3 (right panel). The selected periods in Fig. 5.4 refer to the calibration and validation phases explained in Schaffhauser et al. (2023). The evaluation is done for the basin outlet, gauge Toktogul.

From the figure, it can be seen that both calibrations represent observed flows relatively well throughout the year and are almost identical (straight lines in left panel). However, using both parameterizations under the corresponding forcing not-used in the calibration major shifts can be observed (dashed lines). The original mean absolute error between the two datasets is ± 15.5 m^3/s, and the maximum observed

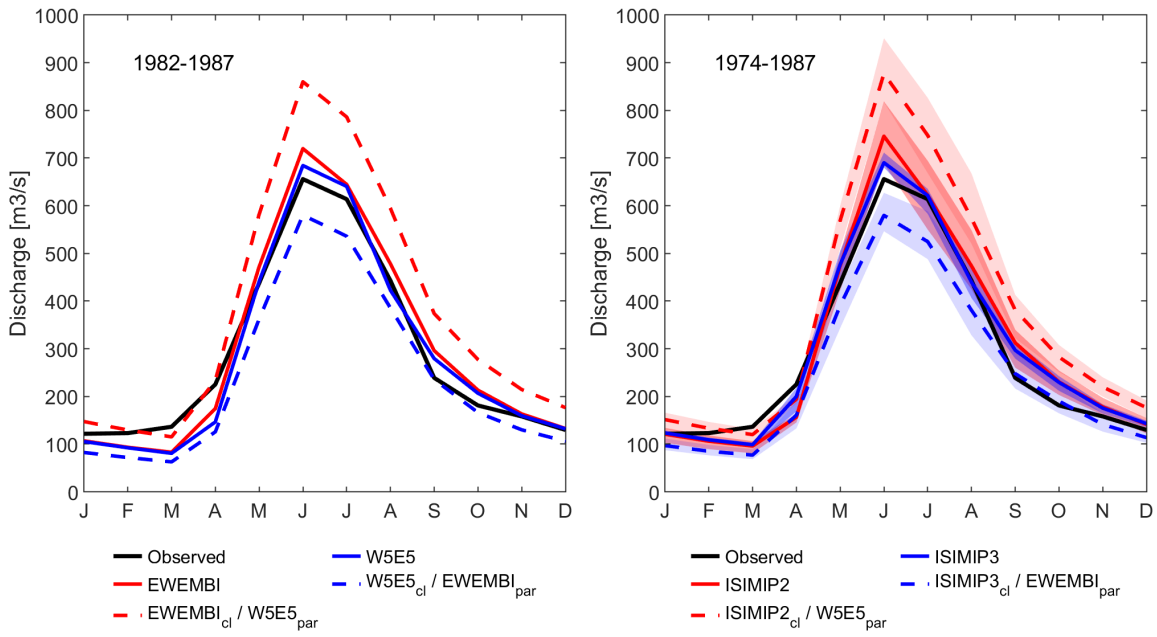

Figure 5.4: Left panel, Effects of parameter transfer between the reference datasets W5E5 and EWEMBI under the climate data of the respective different dataset for the period 1982-1987.Subscript *cl* indicates climate and *par* parameterization, e.g. $\text{EWEMBI}_{\text{cl}}/\text{W5E5}_{\text{par}}$ represents a model driven with the climate data from EWEMBI using the parameterization derived from W5E5. Right panel, same as left but testing the parameterizations for the different historical GCM simulations for the period 1974-1987.

deviation in August corresponds to 11.6%. In contrast, under the different, albeit very similar, climate forcing the mean absolute error increases to ± 137.1 m^3/s with a percentage difference of 32.5% in June where the highest absolute difference occurs. The difference can also be expressed as change in performance metric with respect to observed daily flows. The EWEMBI parameterization forced with W5E5 (daily flows of dashed blue line) leads to a NSE of 0.81 compared to 0.84 before (daily flows of straight red line). However, the KGE drops even from 0.92 to 0.75. The NSE of the W5E5 parameterization using the climate from EWEMBI (daily flows of dashed red line) aggravates to 0.67 from originally 0.87 (daily flows of straight blue line). The KGE is reduced from 0.93 to 0.67, hence underlining the significance of the different baseline one might obtain using independent datasets for calibration, even under comparable climatic conditions.

The effects of the parameter change are directly reflected in the historical period of the climate models (right panel). It is apparent that the climate models of both ISIMIP generations match the mean annual cycle of observed flows well (both straight lines), while using them under the respective different climate generation lead to similar hydrograph shifts observed for the actual reference datasets. The results suggest that even minor climatological changes can highly impair the potential robustness of a calibrated parameter set. Hence, the potential baselines (indicated in the right panel) used for the actual impact assessment can diverge largely and might affect the final impact statement. If the changes are reflected consistently throughout the projections, relative changes with respect to the baseline might be less affected. However, caution is advised, at least when working with absolute values of projected impacts. As a minimum, studies should evaluate and demonstrate the effects of using independent calibration datasets in the historical period. This would not only increase transparency but also allow for a better validation of the the robustness of the impact model. A large deviation of the baselines can at least be interpreted as

an indication that the impact statement may contain uncertainties, particularly with regard to absolute values.

5.3.2 Uncertainty Decomposition - The Role of Climate Model Generations

The results of the uncertainty decomposition for discharge are illustrated in Fig. 5.5. Results of ETA and SM are provided in Fig. 5.6. With respect to the annual discharge projections, we observe a shift in the contribution of the three main factors in favor of the interaction terms from near to far-future. Overall, the interaction terms have a strong contribution in both, the near and far-future. The strongest contribution of the main factors stems from the ISIMIP generation. The strong contribution of the interactions arises mainly from the relationship between the ISIMIP generation and the GCMs (GEN*GCM). For example up to 75% of the total variance can be apportioned to the interaction terms. All other effects are rather small.

Surprisingly, the seasonal results show significant disparities. While the interaction term GEN*GCM,

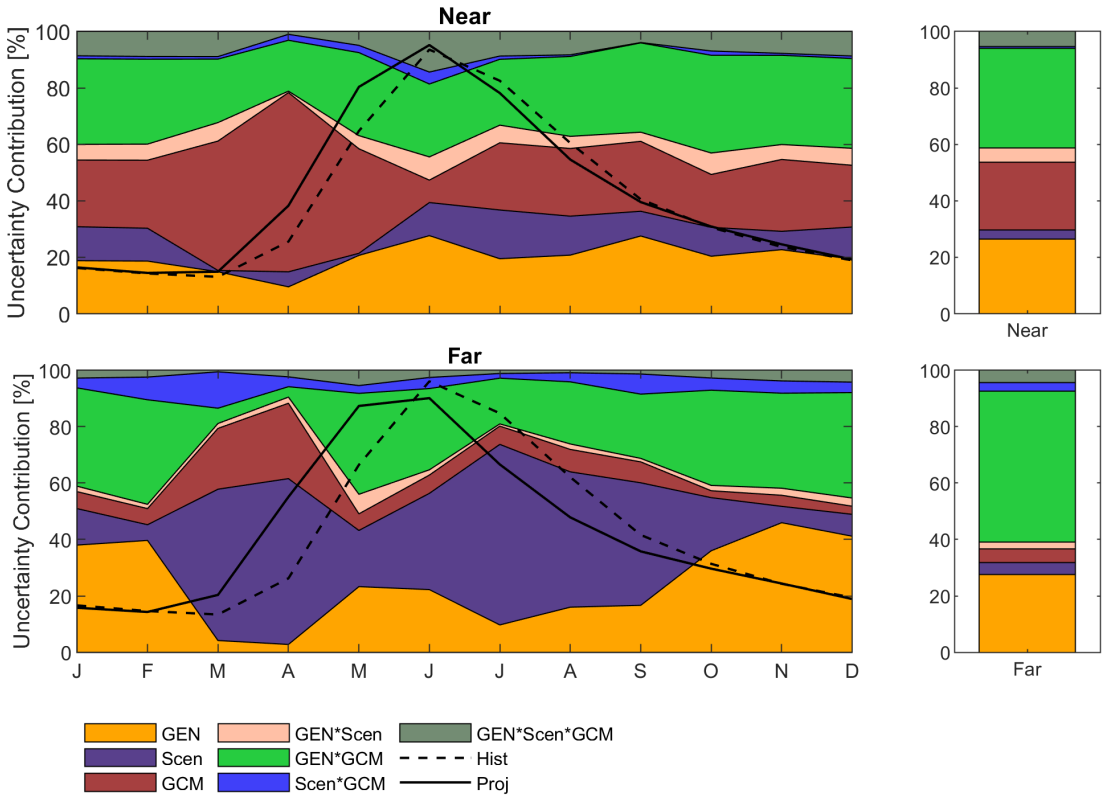

Figure 5.5: Analysis of variance of mean seasonal (left panels) and annual (right panels) discharge anomalies at Gauge Uchterek across 2 ISIMIP generations (ISIMIP2, ISIMIP3 denoted as GEN), 2 emission scenarios (high emission - identified with SSP5-85 and RCP85, low emission identified with SSP1-26 and RCP26 denoted as Scen) and 4 climate models (GFDL-ESM2M identified with GFDL-ESM4, HadGEM2-ES identified with UKESM1-0-LL, IPSL-CM5A-LR identified with IPSL-CM6A-LR and MIROC5 identified with MIROC6 denoted as GCM) for 2 periods (near-future, far-future). The dashed and solid lines of the seasonal ANOVA plot refer to the seasonal discharge of the historical period and the corresponding future projection period (averaged over all years, models and scenarios). Seasonal discharge curves are normalized.

again shows the largest contribution of the secondary effects and does not change much between the two periods, there are large temporal changes in the contributions of the main factors. In the near-future,

the ISIMIP generation (analogous to the annual scale) constitutes the biggest uncertainty contribution. However, in the far-future period a strong increase in the scenario's contribution is observed from March to October (Spring, Summer). At the same time the uncertainty in the discharge projections depends less on the ISIMIP phase. The two phases clearly represent the seasons where the biggest impacts on discharge are expected and the largest overall uncertainties are observed.

In general, high interaction terms relate to the dependency of one factor on another. A high variance caused by an interaction indicates a non-additive relationship. If we pick the dominant source of uncertainty for mean annual discharge changes (GEN*GCM), it is suggested that a change can't be explained by GCMs alone but through its linkage with the other components. The contribution of the individual GCMs depends on the respective ISIMIP phase. A GCM might have a strong effect, expressed as high positive discharge anomaly, on the dependent variable under I2, while it is strongly negative under I3. This also points at least partly to strong developments on the GCM level.

We observe that the GCM term is smaller than its interaction with the generation. This indicates the

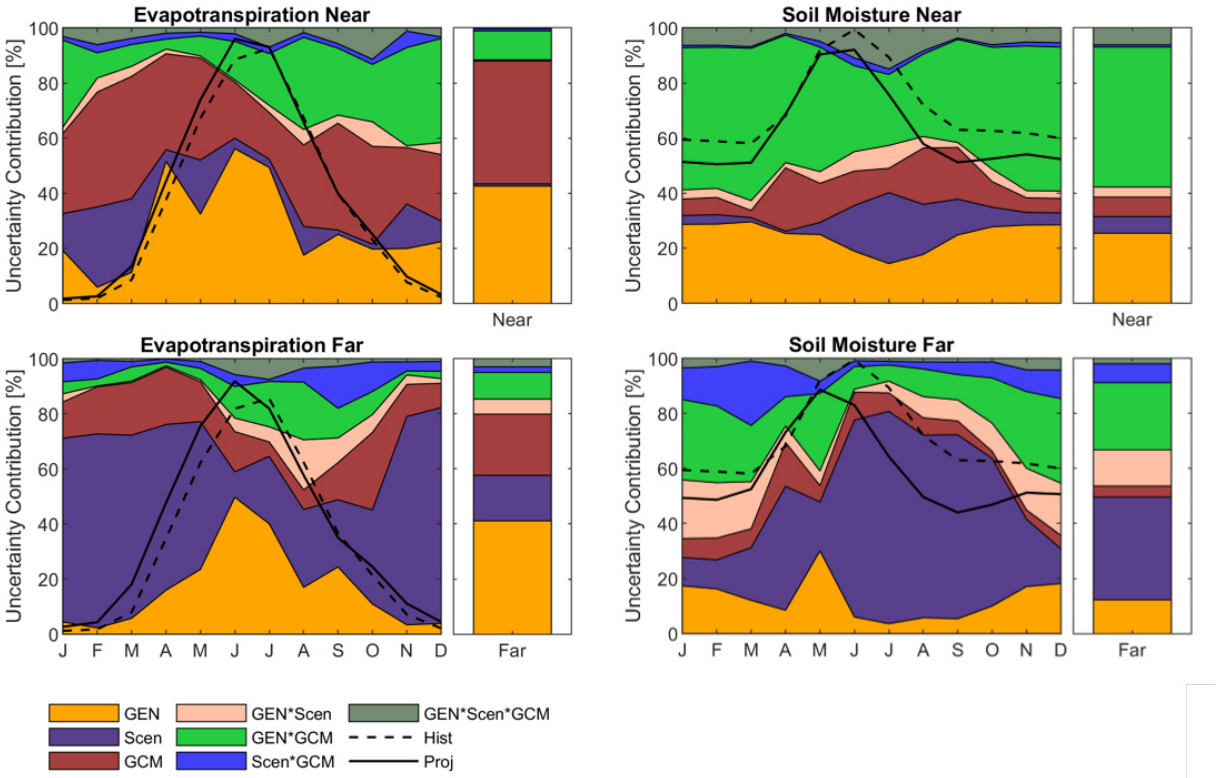

Figure 5.6: Same as Fig. 5.5, but for evapotranspiration (ETA) and soil moisture (SM) at the basin scale. The 4 panels to the left refer to the seasonal and annual analysis of variance (ANOVA) results of ETA, respectively. The 4 panels to the right demonstrate the ANOVA results for SM, again on the seasonal and annual scale respectively. The dashed and solid lines of the seasonal ANOVA plot refer to the seasonal ETA or SM estimates of the historical period and the corresponding future projection period (averaged over all years, models and scenarios). Seasonal SM and ETA curves are normalized.

strong development of the GCMs of the same family. It could further mean that advances within a GCM family of two generations exceed differences between GCMs of the same generation but different families (smaller uncertainty from GCM term than GCM*GEN term). The importance of the scenarios is most evident on the seasonal scale and the far future. We detect larger shifts between the main sources of uncertainty of the near and far future on the seasonal scale compared to the annual scale.

Fig. 5.6 contains the results for ETA and SM. For SM largest annual shift are observed with respect to the low and high emission scenarios. Compared to SM, the annual ETA estimates only have a strong scenario discrepancy under the far-future of SSP85, the remainder are relatively similar. However, the increase in the contribution of the scenarios is also visible for ETA during winter in the far-future period. The period is characterized by strong relative ETA changes which are likely the reason for the increase in the importance of the scenario term. For SM the scenario becomes again crucial in the periods where the biggest changes are observed in the far-future (June to October). The annual as well as seasonal ETA projections are characterized by a huge contribution of the GCMs (and not by the interaction term). This indicates that the distinct GCMs reveal a specific tendency in their estimates regardless of other factors such as the ISIMIP phase.

By using an ANOVA, we showed that differences between climate model generations contribute more to the overall uncertainty than differences between climate models of the same generation. The substantial impacts associated with advances in climate modeling are further reinforced by the strong contribution of the interaction terms that include the generation of climate models. Especially with respect to RQ 1 and 2, it can be concluded that developments in climate modeling can significantly affect impact statements and carry a large part of the total uncertainty. Besides, the scenarios showed an increasing importance in their contribution with time, which was particularly pronounced in Spring and Summer. As we considered GCMs of different generations in one group, the GCM term might be weaker as it would be when the ANOVA would have been applied for the ISIMIP phases individually. In studies where ANOVA was applied for only one generation of climate models the uncertainty resulting from the GCMs is often reported as one of the main sources, while obviously the interaction terms are smaller as in our study (Vetter et al., 2015; Kay et al., 2008). Studies built on that could explore how the uncertainty is attributed if GCMs and scenarios are taken separately per climate model generation rather than lumped as in our example. We further have to admit that one important source of uncertainty, hydrological models, is missing in our study. The inclusion would lead to a more comprehensive picture of the uncertainty sources and could be included in future work. Moreover, parameter uncertainty, which is rarely considered in impact studies, was not taken into account.

5.3.3 Regional Impacts of Global Climate Change

In the following the impacts of global climate change, expressed as mean global temperature change, on the regional catchment hydrology of the Naryn Basin are presented.

The results for the low and high emission scenarios are presented in Figure 5.7. The strongest effects are found for ETA and SM of the high emission scenario. ETA increases linearly with ΔT_g under both I2 and I3, but with a stronger catchment response of ETA under CMIP6. Evaluating the effects of the global thresholds defined in the Paris Agreement, 1.5°C and 2°C, the new scenarios show an increase in ETA for both thresholds, which is around twice of the increase under I2 (+10% under I3 to +5% under I2). After a global temperature increase of 2°C, RCP85 and SSP85 deviate largely, which is true for all variables of the high emission scenarios.

For SM, the impact of a global temperature increase of up to 2°C translates into a similar catchment response under both I2 and I3, after which the decrease in SM for I3 abruptly stops and remains constant. In contrast, under I2 the SM decrease is persistent even for higher levels of global warming although the rate of decrease is dampened. Under the low emission scenario, especially the discharge and SM patterns show no clear trend and tend to show hysteresis effects. Nevertheless, the higher warming levels projected under CMIP6 cause generally larger negative SM anomalies, larger positive discharge and ETA

anomalies. It should be noted, that under the low emission scenario only the 1.5°C level is exceeded in the MMM of both generations.

Discharge shows a very different picture for I2 and I3. However, there is strong agreement that a global temperature rise of 3°C represents a tipping point, at which the direction of changes is reversed. This applies for both, I2 and I3. Interestingly, the pattern of the reversal is flipped under I2 and I3. While the slope of the discharge and ΔT_g relationship is negative under I2 until it becomes positive at a value of 3°C, the relationship is positive (discharge increases with global warming) until it is reverted at a 3°C warming level. Of course the magnitude of change (as shown before) is less compared to ETA and SM.

In summary, the basin demonstrates a hydrological response linked to global warming levels of 1.5°C,

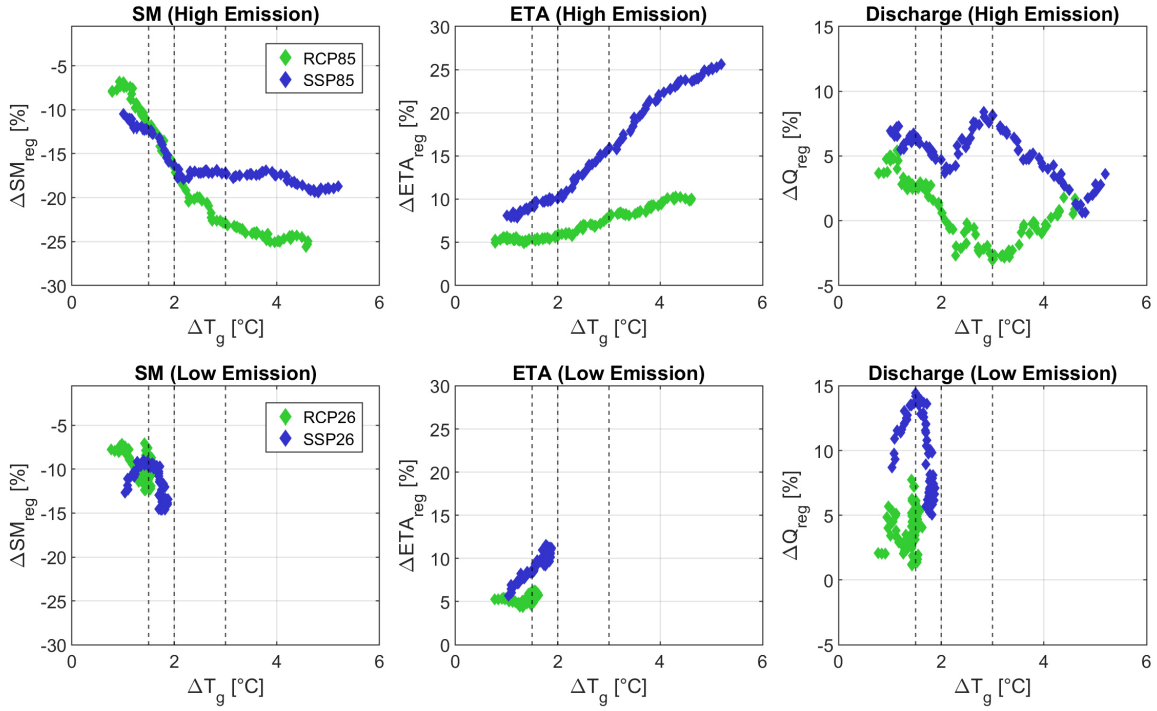

Figure 5.7: Impacts of annual mean global temperature change on regional soil moisture (SM), actual evapotranspiration (ETA) and discharge (Q). Dashed vertical lines refer to the temperature changes which are associated with the Paris Agreement, namely 1.5°C, 2°C and 3°C. Mean global temperature changes are calculated with respect to the 1921-1950 climatology.

2°C, and 3°C, with a slightly stronger effect observed under CMIP6. Moreover, the importance of putting regional impacts into a global context is shown, however, in the example a further temperature increase of 0.5°C is not associated with strong additional hydrological impacts. This could vary significantly for other regions, and half a degree can have significant hydrological implications, which is for example globally shown in Schleussner et al. (2016). Using the warming levels from the Paris Agreement, allows comparability between regional impact studies, albeit the approach is subject to uncertainties partly resulting from a certain degree of subjectivity. Among the related uncertainties, which are also comprehensively discussed in James et al. (2017) for multiple methods, are for example the effects of baseline choices (Liersch et al., 2020), which apply also for the current study here. Moreover, the time sampling approach is not capturing path dependency, i.e. how a model reached the temperature increment. A drawback, also not further explored in this small investigation. However, the regional response of the hydrological system might vary whether greenhouse gas emissions alone are responsible for the global warming level

or if different factors are contributing. Besides, the time sampling approach assumes the response to climate change to be time-invariant, however, hydrological systems can show strongly lagged responses, particularly elements such as glaciers or aquifers might take significantly longer to respond. What many methods have in common is that they have difficulties in capturing the complex relationships between impacts, global temperature rise, greenhouse gas emissions or climate variability. In other words, the regional effects derived from different levels of global warming, entail the uncertainties of this cascade, albeit ignored by the time sampling approach where emphasize is put on the warmer world itself. Nonetheless, approaches such as the relatively long-standing time shifting from James et al. (2017) are important components in communicating the regional effects of global warming. Ultimately, regional impacts attached to different global temperature increments can aid the development of mitigation measures, something on which should be put as much effort as possible.

5.4 Conclusions

In this supplementary chapter, common uncertainties associated with climate impact assessment were investigated using relatively simple and computationally efficient methods. Uncertainties were considered from different perspectives, e.g. uncertainties related to the forcing datasets used to calibrate the model were examined and what implications the choice of dataset may have on model performance and subsequent climate impact assessment. In addition, an ANOVA was carried out to investigate the contribution of different sources of uncertainty along the modeling cascade to the total uncertainty of the impact assessment. Main emphasis, as rather understudied, was put on the contribution of the climate model generation. Finally, and especially in contrast to the ANOVA approach as not focusing on uncertainties along the model cascade, different and widely acknowledged global warming levels were assessed with respect to the respective regional hydrological response.

There is not only a wide range of uncertainties associated with climate impact assessments or methods to analyze them, but also a significant need for transparent discussions and the application of these methods. It is also illustrated that it is valuable to view uncertainties from different perspectives to receive a more comprehensive picture. This is the base for a solid and profound communication of uncertainties present in impact statements. Ultimately, a proper investigation and communication of uncertainties increases the credibility of impact statements that are often founded on the base of ensemble means or medians subject to large model spreads. From the communication perspective, using existing and widely acknowledged global warming levels as from the Paris Agreement can contribute making global warming more tangible at the local level. In particular, the effects of using independent forcings in the calibration period with respect to the reference data set of the climate models can be demonstrated without much computational effort and are, as shown here, enormous. However, this is only one example of many exploration possibilities in this direction. In contrast, approaches that focus on uncertainties along the modeling cascade are, albeit no less important, more common and frequently applied. However, these rarely address the influence of climate model generations and give therefore less indications on the GCM developments within a model family. This, again, underlines the importance to consider different generations of climate modeling in impact assessment, to differentiate between them and in best case to identify reasons for sometimes large differences.

Chapter 6

SWAT-GL: A New Glacier Routine for the Hydrological Model SWAT

The following chapter is based on the publication Schaffhauser et al. (2024b)[1].

Abstract

The hydrological model SWAT is widely used in water resources management worldwide. It is also used to simulate catchment hydrology in high-mountainous regions where glaciers play an important role. However, SWAT considers glaciers in a simplistic way. Although some efforts were done to overcome this limitation, there is no official version available that considers glaciers adequately. This strongly impairs its applicability in glacierized catchments. In this technical note, we propose a novel version of the traditional SWAT, called SWAT-GL, which introduces 1) a mass balance module and 2) a glacier evolution routine to represent dynamic glacier changes. Mass balance calculations are based on a conceptual degree-day approach, similar to the snow routine implemented in SWAT. Glacier evolution is realized using the delta-h (Δh) parameterization, which requires a minimum of data and is thus suitable in data-scarce regions. The approach allows users to simulate spatially distributed glacier changes. Annual mass balance changes are translated to distributed ice thickness changes depending on the glacier elevation. We demonstrate how SWAT-GL is technically integrated into SWAT and how glaciers are merged with the existing spatial units. Model code and test data is freely accessible to promote further model development efforts and a wide application. Ultimately, SWAT-GL aims to make SWAT easily applicable in glacierized catchments without the need of additional tools.

6.1 Introduction

The Soil Water Assessment Tool (SWAT) (Arnold et al., 1998) is a widely-used and recognized tool in water resources management. It is used for a vast number of applications ranging from the evaluation of land management practices (as originally developed for) (Ullrich and Volk, 2009; Himanshu et al., 2019) and sediment transport (Betrie et al., 2011; de Oliveira Serrão et al., 2022) to water quality studies (Nazari-Sharabian et al., 2019; Qi et al., 2020) or the assessment of the impacts of climate

[1] **Schaffhauser, T.**, Tuo, Y., Hofmeister, F., Chiogna, G., Huang, J., Merk, F., & Disse, M. (2024). SWAT-GL: A new glacier routine for the hydrological model SWAT. In JAWRA Journal of the American Water Resources Association. Wiley. https://doi.org/10.1111/1752-1688.13199

change (Schaffhauser et al., 2023; Schürz et al., 2019). SWAT is a physically-based and semi-distributed model mainly applied on the watershed scale. These applications include a large number of studies in mountainous catchments, which are dominated by snow and glacier processes (Tuo et al., 2018; Xu et al., 2015; Rahman et al., 2012; Schaffhauser et al., 2023; Marahatta et al., 2021; Omani et al., 2017; Khan and Koch, 2018; Shukla et al., 2021; Andrianaki et al., 2019). While snow processes are considered by a degree-day approach, glacier processes are not specifically represented in an official release yet.

However, some efforts have already been made in the past to address this topic. One of the most widely used extension of SWAT that considers glacier processes is the approach proposed by Luo et al. (2013). Here a volume-area scaling was integrated into SWAT, and then applied several times (Gan et al., 2015; Luo et al., 2018; Changkun et al., 2015; Shafeeque et al., 2019; Wang et al., 2018) and recently transferred to the successor of SWAT, namely SWAT+ (Yang et al., 2022).

Most of the existing approaches represent glacier dynamics in SWAT either by coupling it to an external glacier model, using static approaches with unrestricted glacier melt that does not allow to represent glacier retreat or by using a volume-area scaling to simulate glacier retreat. Another way is to increase the initial snow storage to compensate for the missing glaciers (Rahman et al., 2012). Besides, none of the approaches has been considered in any of the official releases, nor are these methods always easily or openly available.

To overcome these limitations, modified versions of the traditional hydrological model SWAT are required. We thus propose a modified version which we call SWAT-GL. It consists of two features to model glacier dynamics a mass balance module as well as a glacier evolution module. Like most of the hydrological models (Tiel et al., 2020), SWAT-GL simulates mass balance variations based on a simple degree-day approach. However, in contrast to existing approaches glacier evolution is simulated by implementing the Δh-parameterization first proposed by Huss et al. (2008) and further developed by Huss et al. (2010). The Δh-parameterization is an empirical approach, where glaciers are distributed in elevation zones. The elevation zones are normalized for the minimum and maximum elevation and each zone receives a specific ice thickness change. The idea is that lower elevated zones are subject to stronger ablation than higher zones. The method is mass-conserving and can be applied with a minimum of input requirements (Li et al., 2015), making it also suitable in data-scarce regions. While the method found its way already to hydrological models such as HBV (Seibert et al., 2018a; Li et al., 2015), or WASA (Duethmann et al., 2015) the community did, to the best of our knowledge, not yet integrate the approach in SWAT. However, an alternative approach, which is based on glaciological response units, is for example shown by Wortmann et al. (2016) for SWIM, a model which is very similar to SWAT. The glaciological response units are based on elevation zones and aspect and ice flow can occur between elevation zones.

Our technical note aims to illustrate the technical implementation of the glacier routine in the SWAT code and how glaciers can generally be incorporated into hydrological models such as SWAT. Hereby we take up the point made by Seibert et al. (2018a) who encouraged the community already to implement the approach in other hydrological models. To increase the interoperability, reproducibility and accessibility based on the FAIR principles (Wilkinson et al., 2016) our model code is well-documented and openly available via a Git repository. The technical note of the glacier routine is structured in 4 parts. 1) we provide details on input and preprocessing requirements in combination with the spatial integration of glaciers in SWAT. In 2), the mass balance module is described, followed by 3), the implementation of the glacier evolution module. 4) Last, we provide information where users can access, download or further develop SWAT-GL, as well as where users can obtain the data and model of our model application.

6.2 SWAT-GL

SWAT-GL is a revised version of SWAT considering glacier mass balance as well as glacier evolution. The core of the model is the discretization of glaciers in the aforementioned increments to allow for spatially distributed glacier changes according to the Δh-parameterization. From now on, we use the term Elevation Section (ES) for these increments in which glaciers are divided. ES differ from the traditional elevation bands (EB) of SWAT in which snow processes are calculated and climatic inputs can be adjusted. ES on the other hand are integrated in the existing approach and only used for glacier-specific computations. Climate information here are also received from the individual EBs.
Fig. 6.1 gives an overview on the general workflow of SWAT-GL and how it fits in the actual SWAT code.

6.2.1 Preprocessing & Spatial Integration in SWAT

To apply the glacier model, two additional input datasets (compared to the standard version of SWAT) are required: initial glacier outlines and initial ice thickness values (from which glacier mass can be inferred). Several openly available datasets can be acquired for that task, for example the ice thickness estimates from Farinotti et al. (2019) or Millan et al. (2022) and the glacier outlines from the RGI (RGI Consortium, 2017) might serve as appropriate for the model setup. A further approach to determine initial glacier thicknesses is to use geodetic approaches such as GlabTop (Linsbauer et al., 2012). The approach makes use of an empirical relationship between the elevation range of a glacier and average basal shear stress. However, this assumes an adequate DEM and glacier outlines to be available for the purpose and in best case represent the model starting year, which is probably seldom the case.
As a first preprocessing step before setting up the SWAT model, users have to define ES spacing on the basin scale. A narrow spacing allows for a more detailed representation of glacier evolution. Then, users have to merge their original land use map with the glacier outlines and prepare a land use class for each ES rather than one land use class for glaciers. In the current version of SWAT-GL, the land use class is static, which means glacierized HRUs keep its land use class regardless whether the glacier receded or not. The following descriptive example emphasizes the preprocessing step: If we assume an example in which an ES increment of 50 m would results in 10 ES, the user would need to define 10 separate land use classes. This would refer to an adaption of the land use map, its look up table and the SWAT crop database, in which all 10 classes must be defined accordingly. Typical values for the increment can be 50-200 m (Duethmann et al., 2015; Seibert et al., 2018a). Section 6.3 illustrates the step for a real world example. This step assures that glaciers and especially the corresponding glacierized HRUs get a distinct location on the glacier scale. An HRU with a land use class that corresponds to the lowest ES thus receives an indirect altitude information (as an ES covers an elevation range). In other words, there is a distinct assignment of HRUs to ESs which ensures the applicability of the Δh-parameterization.
Area and ice thickness information are provided in a separate initialization file called *swat_gles_full.txt*. The file contains the initial values for each ES and subbasin. This is important as glaciers are defined on the subbasin scale in SWAT-G. In turn, this means that users define the initial information for all glacierized subbasin in the input file. ES are defined glacier-wide, however, if we for example assume 3 glacierized subbasins and 10 ES, one needs to determine ice thickness and glacier area for each ES separately per subbasin. Detailed information on how to format the new files so that they are correctly read by SWAT-GL, as well as an example, can be found in the manual of our GitLab repository (see Section 6.4). Besides, a second input file with the initial parameterization for glacierized HRUs has to be provided. We hereby follow the approach to put the information of all HRUs in one file rather than

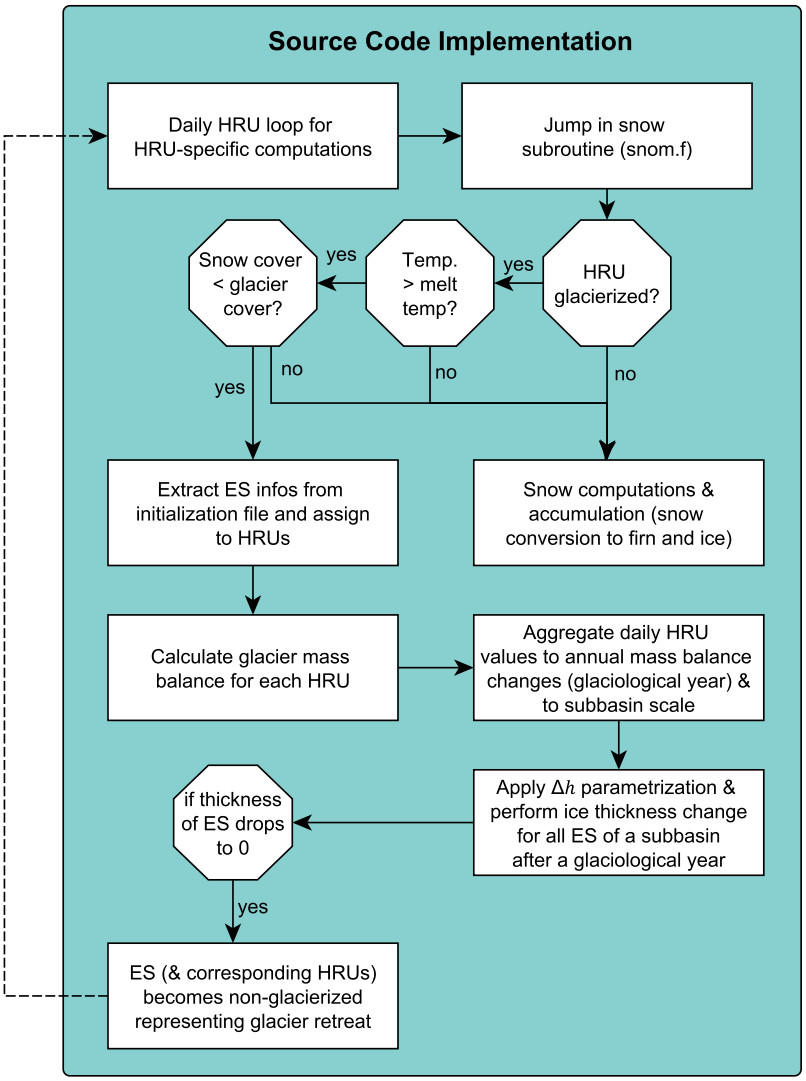

Figure 6.1: Flowchart of the new glacier routine implemented in SWAT-G. Hydrological Response Units (HRUs) are differentiated in glacierized and non-glacierized, its initialization is based on the intersection of the land use map and the glacier outlines. Glacier-specific calculations, except accumulation, only take place when the corresponding HRU is not fully snow-covered (indicated by snow cover $<$ glacier cover). Accumulation takes place when snow is present in an HRU. HRUs are distinctly assigned to a specific ES. Daily glacier calculations on the HRU-scale are aggregated to annual mass balance estimations at the end of the ablation period upon which the Δh-parameterization takes place. After the annual mass balance calculations are distributed over the ES, the HRU values are updated accordingly.

having one file per HRU (as standard in SWAT). Parameters refer to glacier melt, accumulation and sublimation, which are introduced in the next chapter. When the files are prepared and the model is set up as for the standard SWAT, it can be run. An overview of all new input and output files as well as the new parameters introduced can be found in Table 6.1.

6.2.2 Glacier Routine

When SWAT-GL is executed, first the mass balance module is called to estimate melt, accumulation and sublimation on the daily scale. The calculations are performed on the HRU scale. Due to the unique allo-

Table 6.1: Overview of new input files required by SWAT-GL, as well as new or modified output files generated when applying SWAT-G.

Name	Type	Description
Files		
swat_gles_full.txt	Input (New)	Contains the ES details of glaciated subbasins, such as ice thickness and area.
gl_hru_par.txt	Input (New)	Contains the parameterization of glacier HRUS. In total values for 5 new parameters must be assigned.
gl_mb_aa.txt	Output (New)	Summarizes annual (glaciological year) glacier mass balance changes for ES of glaciated subbasins.
output.hru	Output (Modified)	Can be used to print HRU-specific mass balance details on a daily time step.
Parameters		
GLMTMP	Input	Threshold temperature for glacier melt [°C]
GLMFMX	Input	Melt factor for ice on June 21 [mm H_2O/(°C·day)]
GLMFMN	Input	Melt factor for ice on December 21 [mm H_2O/(°C·day)]
β_f	Input	Refreezing factor of glacier melt [-]
f_{acc}	Input	Conversion factor of snow to firn and ice [-]

cation of HRUs to ESs within a subbasin, the initial values of an ES (specified in the *swat_gles_full.txt*) are uniformly assigned to the HRUs that belong to this ES. For this purpose, a new subroutine *readglsubs_t.f* was created. Daily mass balance estimates are aggregated for each glaciological year (01.10. - 30.09.) before they are used in the Δh subroutine, which refers to the second step of the routine.

6.2.2.1 Mass Balance Model

The mass balance calculations are merged with the existing snow-based computations (*snom.f*), with an exception for sublimation which is estimated in the original *etact.f* routine, where also snow sublimation is considered. The mass balance can be formulated as

$$EW_t = EW_{t-1} - M_t \cdot (1 - \beta_f) - S_t + C_t \tag{6.1}$$

where EW_t is the water equivalent of ice [mm H_2O] at day t, M_t is the melt rate [mm $H_2O \cdot d^{-1}$], S_t is the sublimation rate [mm $H_2O \cdot d^{-1}$], C_t is the accumulation rate [mm $H_2O \cdot d^{-1}$] and β_f refers to a refreezing factor of melt, for example also introduced in Luo et al. (2013) where a value of 0.2 from Jiang et al. (2010) was stated. In our case β_f is a variable calibration parameter that can be used to reduce high melt rates. It has been reported that refreezing of melt water in glaciers or ice sheets can have a considerable contribution (Munro, 2005; Bonekamp et al., 2019; Abrahim et al., 2023).
Glacier melt can be written as

$$M_t = \begin{cases} (T_{mx,t} - T_{gmlt}) \cdot b_{gmlt}, & \text{if } T_{mx,t} > T_{gmlt} \text{ and } A_{sc} < A_{gc} \\ 0, & \text{if } T_{mx,t} < T_{gmlt} \text{ or } A_{sc} > A_{gc} \end{cases} \tag{6.2}$$

with $T_{mx,t}$ being the daily maximum temperature [°C], T_{gmlt} being the threshold temperature [°C] glacier melt to occur, b_{gmlt} refers to the degree-day factor of ice [mm $H_2O \cdot °C^{-1} \cdot d^{-1}$] and A_{sc} and A_{gc} refer to the snow and glacier covered fractions, respectively. Glacier melt is only initiated when the daily maximum temperature reaches the critical threshold temperature and a glacier HRU is not fully snow covered.

The degree-day factor of ice is calculated analogously to the one of snow, which is is already included in SWAT and is time varying. The corresponding equation is

$$b_{gmlt} = \frac{(b_{gmlt,mx} + b_{gmlt,mn})}{2} + \frac{(b_{gmlt,mx} - b_{gmlt,mn})}{2} \cdot \sin\left[\frac{2\pi}{365}(t - 81)\right] \tag{6.3}$$

where $b_{gmlt,mx}$ and $b_{gmlt,mn}$ correspond to the maximum and minimum degree-day factors for June 21 and December 21 [mm $H_2O \cdot °C^{-1} \cdot d^{-1}$]. In theory degree-day factors of ice are usually higher than those of snow. An overview of the magnitude of degree-day factors can be found in Singh et al. (2000). If degree-day factors of snow exceed those of ice, SWAT-GL automatically corrects the latter and uses the factors of snow for ice as well in order to preserve the physical relationship.
Accumulation is formulated as

$$C_t = SWE_t \cdot f_{acc} \tag{6.4}$$

with SWE_t being the snow water equivalent of the current day [mm H_2O] and f_{acc} being the accumulation rate coefficient [-], which varies seasonally according to Luo et al. (2013). The described accumulation process is a very simple representation of the transition from snow to firn and ice (Seibert et al., 2018a; Luo et al., 2013). Moreover we want to mention that we introduced a simple snow redistribution scheme, which is standard in many other hydrological models such as WASA (Duethmann et al., 2015). If the SWE of an elevation band reaches a critical threshold the difference between the actual SWE and the threshold SWE is shifted to the next lower elevation band.
Lastly, sublimation is described as function of potential evapotranspiration

$$S_t = ETP_t \cdot \alpha \tag{6.5}$$

where ETP_t is the potential evepotranspiration rate [mm $H_2O \cdot d^{-1}$] and α represents a sublimation coefficient [-]. The sublimation coefficient is seasonally varying.

6.2.2.2 Glacier Evolution Model

The Δh relationship between normalized ice thickness change and normalized glacier evolution was first determined for 48 Swiss glaciers from which three general relationships have been inferred based on different glacier sizes (Huss et al., 2010). The three general relationships are illustrated in Fig. 6.2 and serve as the core of the glacier evolution module. The parameterization is based on the assumption that ice thickness changes are stronger in lower elevated zones (ablation zone) and small in the accumulation zone. The normalization of the ES is based on the minimum and maximum glacier elevation in the basin:

$$E_{norm,i} = \frac{E_{max} - E_i}{E_{max} - E_{min}} \tag{6.6}$$

where $E_{norm,i}$ is the normalized elevation of ES i [-], E_{max} and E_{min} refer to the maximum and minimum glacier elevation [m] and E_i is the actual elevation of ES i [m].

The Δh-parameterization uses annual mass balance changes on the subbasin scale (EW values from Eq. 6.1 are multiplied with subbasin area to obtain volume changes). Technically, to be consistent with the original SWAT code, it is implemented in the *virtual.f* subroutine, where other subbasin-specific aggregations also take place. Annual hereby refers to the glaciological year. At the end of each glaciological year the mass balance change serves as input to the Δh module. Therefore, a corresponding *deltah_ t.f* subroutine is called from the *virtual.f* file, on the 1st of October of each year. The mass balance change

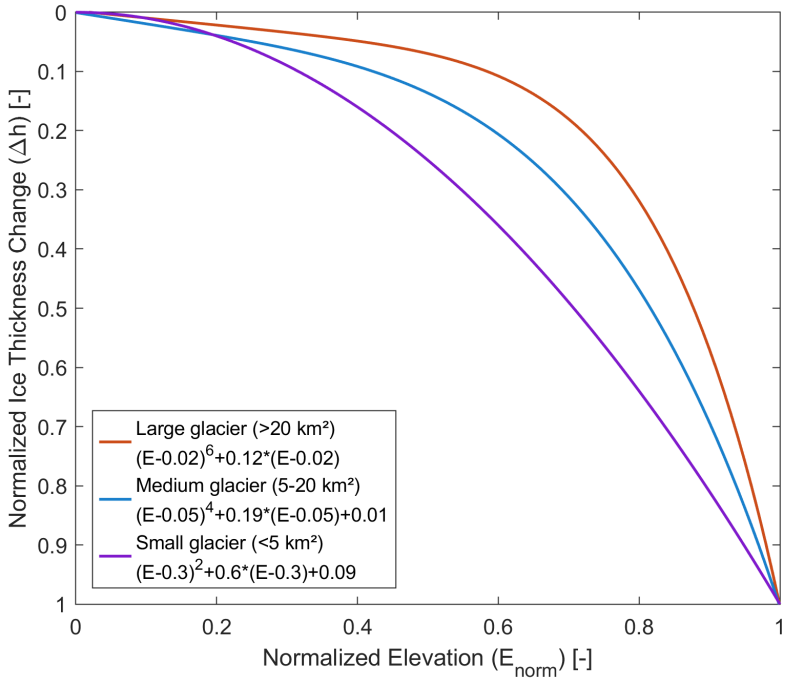

Figure 6.2: Empirical relationship between ice thickness change and elevation based on Huss et al. (2010). The mathematical description of the relationship is based on the glacier size and three cases can be distinguished: large glaciers - red; medium size glaciers - blue; small glaciers - purple.

is distributed over the ESs of a subbasin to simulate either ablation or accumulation (of existing ES). Glacier advance is not yet considered, similar to the original description by Huss et al. (2010). Accumulation is therefore restricted to glacierized areas of the corresponding time step. However, upcoming releases of SWAT-GL will consider advance, e.g. following the approach of Seibert et al. (2018a).
The ice thickness change follows one of the three equations of Fig. 6.2, which in its general form is written as

$$\Delta h_i = (E_{norm,i} + a)^y + b \cdot (E_{norm,i} + a) + c \tag{6.7}$$

with a, b, c, y being coefficients varying for the glacier size classes and Δh_i being the normalized ice thickness change for ES i. Users should note that specific coefficients, different from the ones proposed by Huss et al. (2010), can theoretically be determined to describe local conditions more adequately. However, the required information to derive the coefficients must be available for a specific glacier, usually two DEMs from different points in time. In the next step, a scaling factor f_s [m] which converts the dimensionless ice thickness change Δh_i into an actual change is determined. The scaling factor, in combination with the sum of the individual ES areas multiplied with the respective Δh values must equal the calculated annual mass balance change of a subbasin. To obtain the annual varying scaling factor the following equation is used

$$f_s = \frac{V_a}{\sum_{i=1}^{n} A_i \cdot \Delta h_i} \tag{6.8}$$

where V_a refers to the annual glacier volume change expressed as water equivalent [m^3], determined by

multiplication of the annual aggregated EW values (Eq. 6.1) with the respective subbasin area. A_i refers to the area of ES i and n is the total number of ES.
The scaling factor is then used to update the actual ice thickness of each ES

$$h_{i,1} = h_{i,0} + f_s \cdot \Delta h_i \tag{6.9}$$

$h_{i,1}$ hereby refers to the updated ice thickness [m water equivalent] after a glaciological year in ES i, while $h_{i,0}$ represents the ice thickness [m water equivalent] in the same ES before the Δh-parameterization was applied. If $h_{i,1}$ becomes zero the ES is assumed to be ice-free and the glacier extent is updated accordingly.
After the ice thickness and the glacier area are updated, the HRU ice water equivalents are updated accordingly. SWAT-GL was tested and evaluated in the small and glacierized Martelltal catchment in South Tyrol in Italy. Our test catchment has an area of approx. 77 km^2 with a glaciated area of 11.7 km^2 (approx. 15%) based on the Randolph Glacier Inventory V6 (RGI Consortium, 2017). The elevation ranges from 1,800 m to over 3,700 m. 10 ES were defined based on a spacing of 100 m with the first ES starting from roughly 2,666 m. 10 out of 32 delineated subbasins were (partly) glaciated. We used discharge data of two gauges as well as glacier mass balance and area data to calibrate and validate SWAT-G. While daily discharge data of the Zufall Hut was mostly available during spring and summer, the discharge data at the basin outlet (Zuftritt Reservoir, see Fig. 6.3) was artificially generated from the elaboration of the hydropower station data. This discharge time series was already used in a previous study (Puspitarini et al., 2020) to calibrate and validate a hydrological model in Martelltal. Details on the discharge data of the Zuftritt Reservoir can be found in Puspitarini et al. (2020). Glacier mass balance measurements were available for the Langenferner Glacier from 2004 to 2021 which were here used to exemplify the capabilities of the glacier routine (Galos et al., 2017). Mass balance observations from further glaciers were not available. However, the model setup included all glacierized parts of the basin and was not restricted to the Langenferner. See Fig. 6.3 for the location of the Langenferner and the other glacierized parts in the basin. Meteorological data was acquired from the Autonomous Province of Bozen/Bolzano - South Tyrol.
Calibration in the example application was performed in a multi-objective way for glacier mass balances of the Langenferner and discharge at the two gauges. An automatic scheme using multiple iterations of Latin-Hypercube samples were used to identify reasonable parameter ranges. The automatic procedure was partly extended by manual calibration runs based on expert-knowledge. Used parameters varied between the calibrated subbasins, but included all snow and (introduced) glacier-related parameters. Performance metrics provided in the following paragraphs refer to one of the best solutions of the last iteration and serve only as an example.
Glaciers were initialized using the ice thickness estimates from Farinotti et al. (2019) and glacier outlines from RGI (RGI Consortium, 2017). The assumption is that this combination might be representative in our application that spans a relatively short period in the 2000s. However, for simulations starting decades ago this assumption is likely to not represent former conditions adequately and approaches such as from Linsbauer et al. (2012) can offer valuable solutions.

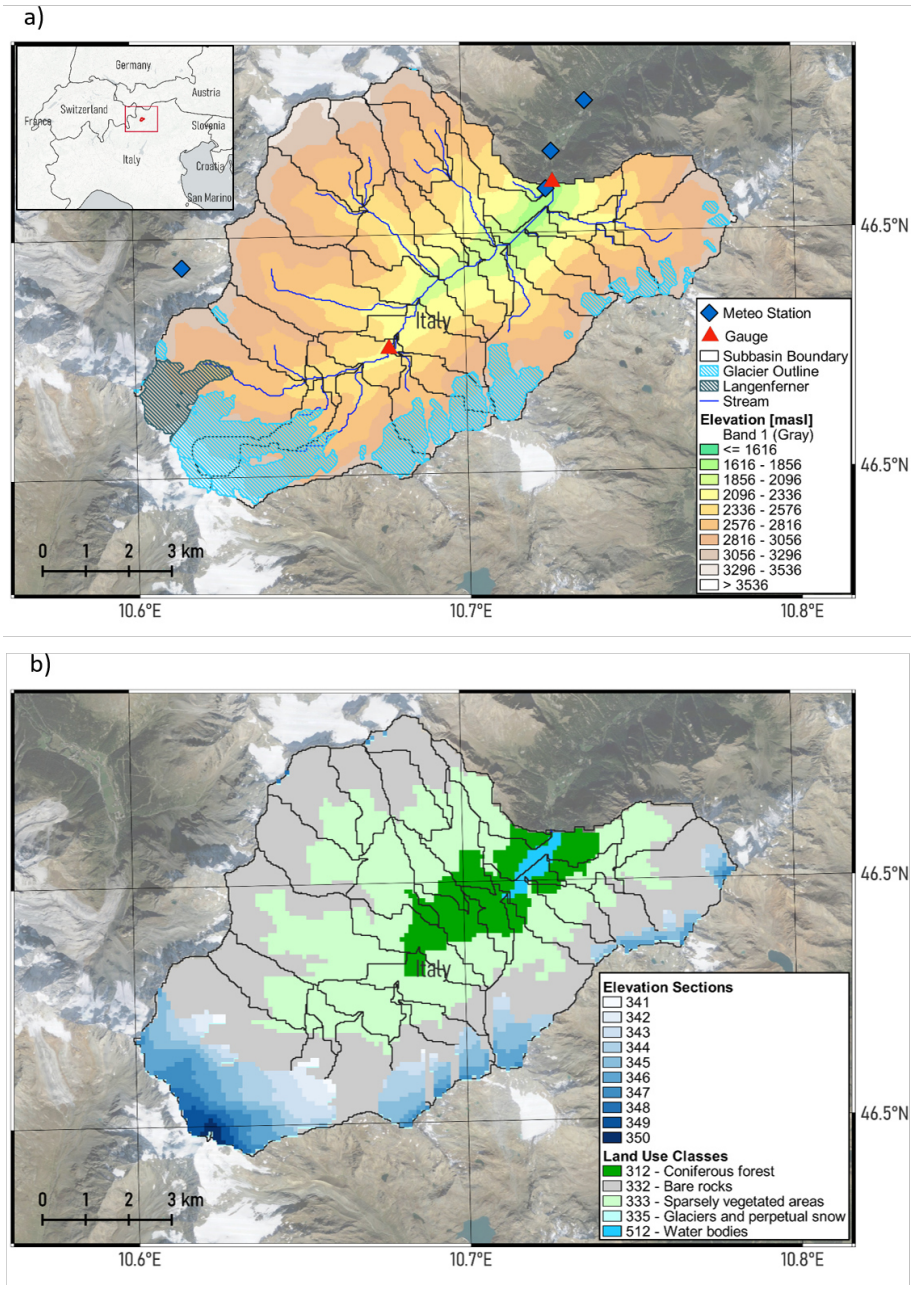

Figure 6.3: Overview of the study area, a) elevation distribution, b) land use map showing an example where 10 ES (indicated as 341 to 350 in the legend) are separately considered as distinct land use classes. Originally, there was only one global land use class for glaciers and no separate land use class for a specific elevation range was considered.

6.3 Application Example

6.3.1 Model & Test Setup

6.3.2 Test Results

Figure 6.4 summarizes the results of an example simulation in which the model was calibrated and validated for discharge and glacier mass balances. In our example the model was evaluated by means

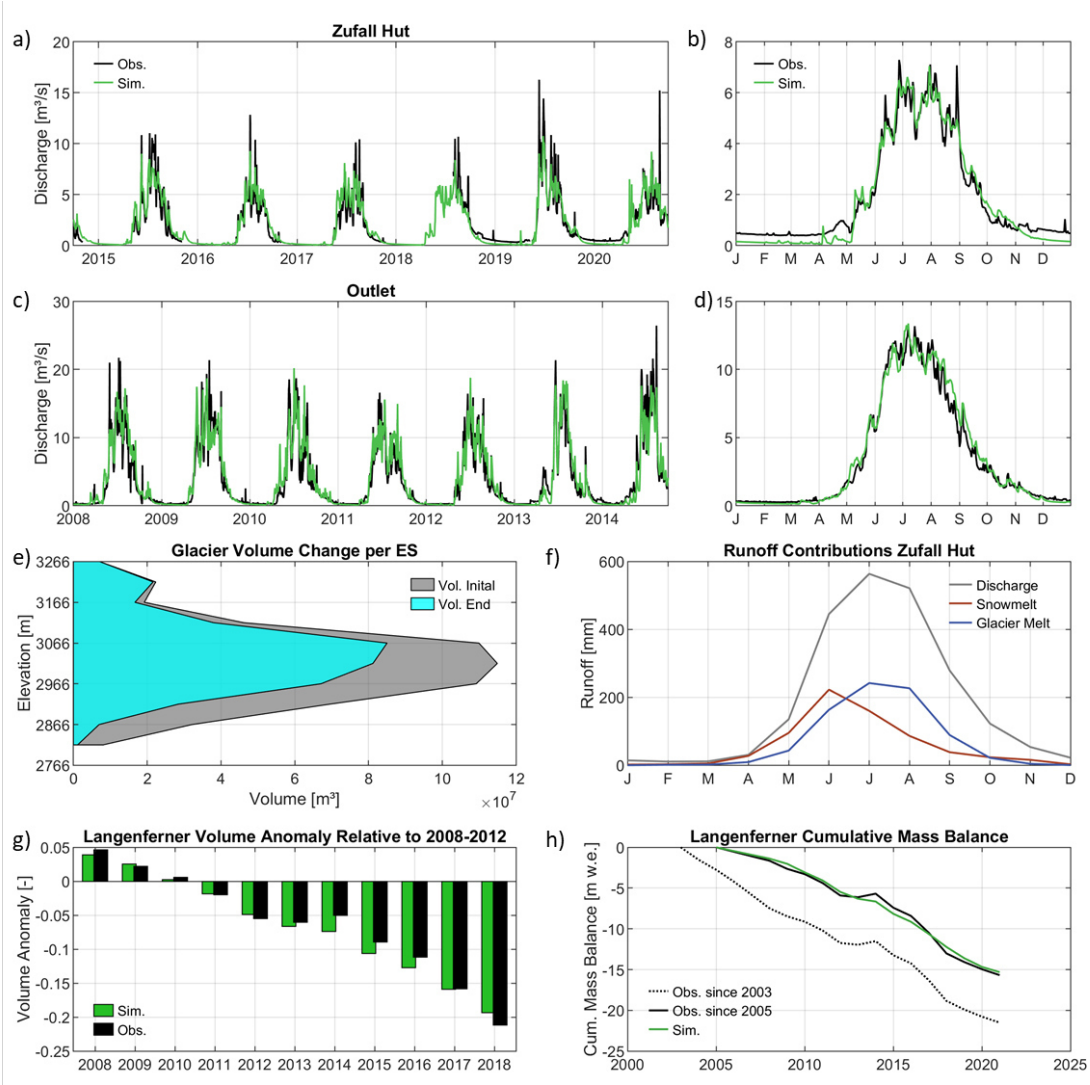

Figure 6.4: Example application of SWAT-GL in the Martelltal: a) Daily discharge simulations at the Zufall Hut for both, the calibration and validation period. Gaps in the time series indicate missing observations; b) Mean seasonal discharge simulations based on the combined calibration and validation period; c) Same as a) but for the basin outlet; d) Same as b) but for the basin outlet; e) Elevation-dependent glacier volume change for the whole simulation period. Changes are shown as the volume at the beginning (grey) and end (cyan) of the simulation for the different elevation sections. The initial volume was derived from the ice thickness estimates of (Farinotti et al., 2019); f) Mean monthly simulated runoff contributions of snow and glacier melt on the basin scale (no observations); g) Model performance of the glacier routine evaluated as glacier volume changes for the Langenferner glacier. Results are shown as annual volume anomalies relative to the period 2008-2012, as the subbasin covering the Langenferner exhibits a larger glacier extent as the Langenferner glacier alone. h) shows the cumulative mass balance change expressed as surface elevation change for the Langenferner and the covering subbasin.

of the NSE (Nash-Sutcliffe Efficiency), KGE (Kling–Gupta efficiency) and PBIAS (Percentage Bias). Due to differences in data availability the two gauges were evaluated for different periods. SWAT-GL calculates mass balances on the subbasin scale, which means that actual observed mass balance values of a glacier can only be used directly if a subbasin almost perfectly matches the outline of that glacier, which is difficult in practice. In our watershed delineation we ended up with a subbasin which is larger than the outline of the Langenferner and thus also included adjacent glaciers. We thus assessed glacier

mass balance changes using annual mass balance anomalies rather than absolute mass balance values. As our subbasin includes adjacent glaciers the mass balance observations that only cover the Langenferner would mismatch. The assumption is therefore that the relative changes of the Langenferner observations and the corresponding subbasin simulations should be in the same order of magnitude. Anomalies were calculated with respect to the period of 2008-2012.

Discharge calibration and validation results (Fig. 6.4 a) - d)) for KGE are 0.90 (Zuftritt) and 0.79 (Zufall Hut), 0.89 (Zuftritt) and 0.84 (Zufall Hut) for NSE and the PBIAS ranged from -11.6% to 5.5% at both gauges (calibration and validation), respectively. In general, the results were slightly better for the Zufall Hut compared to the Zuftritt Reservoir. The degraded results at the reservoir outlet stem mainly from an overestimation in simulated discharge at the beginning of the high flow period in May. However, as the observed time series was artificially reconstructed, mismatches are inherent. We further can see that the glacier routine is able to reproduce the general catchment behaviour, with Spring flow being dominated by snow melt while glacier melt is dominant in late Summer (Fig. 6.4 f)) at the Zufall Hut. At the Zufall Hut the mean annual contribution of the cumulative snow and glacier melt to the total runoff is around 70%.

With respect to the glacier mass balance anomalies of the Langenferner (Fig. 6.4 g)) the general pattern was relatively well represented in our routine. Besides, the model was able to slow down the present recession trend from 2012 - 2014 (Fig. 6.4 g) and h)). However, the glacier routine nearly produced an equal cumulative mass balance loss over the simulation period as shown in the observations (Fig. 6.4 h)). The correlation (Pearson) of the anomalies is around 0.99. The core of the Δh-parameterization in SWAT-GL is visible in Fig. 6.4 e). It is shown that the lower elevated glaciated parts of the basin are subject to a stronger glacier retreat while upper parts of the basin remain stable. It can be seen that lower ES faced a decrease in ice volume of more than 80% in this example.

6.4 Code & Data Availability

In order to follow the FAIR principles (Wilkinson et al., 2016), the model code is made publicly available. Interested users can access the model via a GitLab repository `https://gitlab.com/lshm1/swat-g`. Here, we also provide users a complete documentation of SWAT-G. Updates will be announced there as well. In addition, our example application can be accessed via Zenodo and contains the full model setup we used here. The archive is found via `https://zenodo.org/record/8068725` (Schaffhauser, 2023).

6.5 Conclusions & Outlook

We present a new and openly available version of the hydrological model SWAT, which is not only capable to simulate glacier mass balances on the catchment scale, but considers glacier evolution based on the Δh-parameterization.

The mass-conserving approach allows users to represent spatially distributed glacier changes. It overcomes limitations of hydrological models where glaciers are either not integrated at all, or areal glacier changes are not taken into account (e.g. glacier evolution). The latter case can lead to unrealistic runoff contributions, especially when substantial glacier retreat phases are simulated. In contrast to volume-area scaling, the Δh-parameterization explicitly considers basin-wide ice thickness distributions and enables elevation-dependent ice thickness changes which are translated into distributed area changes. Volume-area scaling only provides the average ice thickness for each glacier and cannot account for a number of parameters that have a large influence on glacier volume, such as surface slope or climatic conditions

that affect mass balance (Helfricht et al., 2019). Furthermore, Δh-parameterization can be applied to individual glaciers, which is not recommended for volume-area scaling (Bahr et al., 2015).
Moreover, our application demonstrates the capabilities of SWAT-GL, although the calibration scheme does not explicitly consider snow processes, e.g. in the form of snow cover maps. Even though our small test showed generally good agreements in the mass balance patterns as well as the cumulative changes of the Langenferner, weaknesses were discovered in the accumulation phase of 2012 - 2014 where the model slowed down the recession but without reversing it. The general evaluation and benchmarking of SWAT-GL needs clearly more and comprehensive investigations which is an on-going task, while we here wanted to focus on the technical details. We highly recommend the application of any of the existing solutions to represent glacier processes in glaciated catchments, when SWAT is the favorable model to be used. In recognition of the barriers to use SWAT in these catchments, such as the inaccessibility of existing solutions or the requirement to use and manually couple additional (glacier-specific) software, our work can contribute to circumvent these obstacles and expand the applicability of SWAT. We also want to encourage the community to make model code available to promote model development, so feedback, reuse or further development of our own code is highly appreciated. Generally improved access to existing and new software would facilitate reproducibility and promote benchmarking in hydrological studies.
However, compared to Luo et al. (2013) glacier advance is not yet included but will be included in one of the next releases. A promising approach is to follow the suggestion of Seibert et al. (2018a), which allows to include glacier advance within the initial glacier outline. Care should thus be taken if SWAT-GL should be used for long-term simulations in the past where glacier advance might be significant. Besides, the current study focuses only on one example in Italy. A comprehensive benchmarking and verification of SWAT-GL is thus missing so far. As the focus in this technical note is on the presentation and implementation of the new module, future studies will focus explicitly on the question of SWAT-GL's applicability in and transferability to other glaciated catchments across the world. A current limitation of the approach is, that particularly for narrow ES spacing, the land use thresholds in the HRU definition are recommended to be close to 0 to allow a best-possible representation of all ESs and to avoid an accidental removal of ES. However, this in combination with the general approach to define one land use class per ES could result in relatively high numbers of HRUs, especially in large catchments. Besides, land use classes are static in our current version, whereas changing land use classes in periods of full glacier retreat would be more realistic. An approach could be further developed to define two land use classes for glacier areas, one for periods of glaciation and one for ice-free times in future studies. However, parameter stationarity assumption is a general issue that should be given more consideration in hydrology, but which is beyond the scope of this work. Further difficulties refer to the data availability for the initialization as those information is rather scarce and usually not available at multiple points in time for a specific region. However, global available datasets such as the ice thickness dataset from Farinotti et al. (2019) or Millan et al. (2022) provide reasonable solutions. Besides, fragmented glaciers within a subbasin are treated as one construct, a limiting but necessary simplification in semi-distributed hydrological models. One solution for the consideration of debris-covered glaciers would be the adjustment of the melt factor of ice. Different studies (e.g., Muhammad et al. (2020)) concluded that the degree-day-factor for ice melt is reduced through debris cover. However, a sound sensitivity study for glaciers with different degrees of debris cover is necessary for defining an appropriate parameter range. This was (also) not within the scope of this study, but we encourage the scientific community to test the new glacier routine under different conditions.
We also recognize that future model development efforts should be put into the newer successor model SWAT+ (Bieger et al., 2017). Future versions of SWAT-GL that address current shortcomings will thus

increasingly focus on SWAT+. However, the spatial integration of glaciers in SWAT+ will likely differ from our current implementation, due to the revised spatial structure of SWAT+.

Acknowledgements

We gratefully acknowledge support from the Autonomous Province of Bozen/Bolzano - South Tyrol and the Institute of Atmospheric and Cryospheric Sciences (ACINN) - University of Innsbruck for providing the hydrometeorological station data. This work was supported by the DFG (Deutsche Forschungsgemeinschaft CH981/3-2 and DI639/5-1) Research Group (FOR2793/1 and FOR2793/2) “Sensitivity of High Alpine Geosystems to Climate Change since 1850” (SEHAG); under Grants [DI639/5-1 and CH981/3-1]. The authors also want to thank the BMBF (Bundesministerium für Bildung und Forschung) for the funding of the "OekoFlussPlan" project (grant number 01LZ1802B).

Chapter 7

Merits and Limits of SWAT-GL: Application in Contrasting Glaciated Catchments

The following chapter is based on the publication Schaffhauser et al. (2024a)[1].

Abstract

The recently released SWAT-GL aims to overcome multiple limitations of the traditional hydrological model SWAT (Soil Water Assessment Tool) in glaciated mountainous catchments. SWAT-GL intends to increase the applicability of SWAT in these catchments and to reduce misapplication when glaciers have a significant role in the catchment hydrology. It thereby relies on a mass balance module, based on a degree-day approach similar to SWAT's snow melt module, extended by a glacier evolution component which is based on the delta-h (Δh) parameterization. The latter one is a mass conserving approach which enables the spatial distribution of ice thickness changes and thus dynamic glacier retreat. However, the extended SWAT version was not yet comprehensively benchmarked. Hence, our paper aims to benchmark SWAT-GL with four different benchmark glaciers which are part of the USGS (United States Geological Survey) Benchmark Glacier Project. The benchmarking considers a comprehensive evaluation procedure, where the routine is optimized on glacier mass balance and hypsometry as well as snow cover. Snow cover is included to consider snow-glacier feedbacks appropriately. Besides, a sensitivity analysis using Elementary Effects (or Method of Morris) is performed to give a detailed picture on the importance of the introduced glacier processes, as well as the relevance of the interactions with the already-existing snow routine. We intentionally did not include discharge in the optimization procedure to fully demonstrate the capabilities of SWAT-GL in terms of glacier and snow processes. Results demonstrate that SWAT-GL is able to perform reasonably well in contrasting glaciated catchments, which underlines SWAT-GL's applicability and transferability. We could further show its strong (non-linear) interactions with the existing snow routine suggesting a simultaneous calibration of the snow components. While snow and glacier processes were adequately represented in the catchments, discharge was not necessarily represented sufficiently

[1] **Schaffhauser, T.**, Hofmeister, F., Chiogna, G., Merk, F., Tuo, Y., Machnitzke, J., Alcamo, L., Huang, J., and Disse, M.: Merits and Limits of SWAT-GL: Application in Contrasting Glaciated Catchments, Hydrol. Earth Syst. Sci. Discuss. [preprint], https://doi.org/10.5194/hess-2024-89, in review, 2024.

when excluded in the optimization procedure. However, SWAT-GL has been shown to be easily capable of reproducing discharge when used in a stand-alone optimization, although this may come at the expense of model consistency.

7.1 Introduction

We recently submitted a paper that introduces a new glacier routine to SWAT to overcome current limitations in its applicability especially in glaciated catchments (Schaffhauser et al., 2024b). The work is built on previous efforts of multiple groups which intend to address common constraints of conceptual and physically-based hydrological models in glacier-dominated catchments. Examples include the work of Seibert et al. (2018a) or Li et al. (2015) for HBV (Hydrologiska Byråns Vattenbalansavdelning), Wortmann et al. (2016) for SWIM (Soil and Water Integrated Model) or Shannon et al. (2022) for the DECIPHeR model (Dynamic fluxEs and ConnectIvity for Predictions of HydRology). However, to our knowledge many hydrological models do not include glacier routines by default and glacier-focused extensions are often only available to the developing groups, although trends are clearly towards publishing model code and making it openly accessible. Despite these improvements, applications in glaciated basins remain challenging due to missing (or very simple) glacier representations, whereby modelers might rely on external couplings of glaciological and hydrological models (Adnan et al., 2019; Wu et al., 2015; Stoll et al., 2020; Du et al., 2022; Naz et al., 2014; Wiersma et al., 2022; Chen et al., 2017b). As it is commonly known, glacio-hydrological models applied to small and highly glacierized catchments often have no or a rather rough representation of additional hydrological components (e.g., evapotranspiration) (Hassan et al., 2021; Ali et al., 2017; Pradhananga et al., 2014), potentially leading to an integration problem, at least when the model domain is extended (Tiel et al., 2020; Wortmann et al., 2016). A further problem is that it is not always clear in the existing literature whether, for example, a glacier routine is coupled with or integrated into an hydrological model, if this hydrological model by default does not take glacier processes into account (e.g., SWAT, VIC (Variable Infiltration Capacity), HBV) and is used to simulate glacio-hydrological processes. The terms *integrated* and *coupled* seem to be used interchangeably, thus impairing reproducibility. From our perspective, integration should suggests a model expansion, while coupling suggests the use of an additional model.

Past SWAT-specific efforts to improve capabilities in glacier-fed catchments other than those from Schaffhauser et al. (2024b) for example refer to the work of Luo et al. (2013), who implemented a volume-area scaling (VA scaling) for glacier evolution along with a degree-day-based mass balance module. The modified SWAT version was further applied in several studies, mostly focusing on China (Gan et al., 2015; Luo et al., 2018; Changkun et al., 2015; Shafeeque et al., 2019; Wang et al., 2018) and recently integrated in SWAT+ (Yang et al., 2022). However, to the best of our knowledge in none of the publications the model code was made publicly available. Besides, Ji et al. (2019) for instance implemented an ice melt routine based on a degree-day approach but did not account for glacier evolution explicitly. Unfortunately, none of the approaches was included in any of the official SWAT revisions.

SWAT-GL was developed to tackle these issues and to provide a freely available and user-friendly SWAT version for glaciated catchments (Schaffhauser et al., 2024b). Besides, the chosen approach, namely the Δh-parameterization from Huss et al. (2010) has proven to be a robust method to simulate glacier evolution in glaciated catchments (Huss and Hock, 2015). Mass balance simulations are similar to most glacio-hydrological models performed with a simple temperature-index approach (Tiel et al., 2020). The empirical Δh-parameterization is called annually to translate the cumulative mass balance change to a change in glacier geometry. The concept assumes that lower elevated areas closer to the glacier terminus

receive stronger ablation than higher elevated ones (Huss et al., 2008, 2010). Therefore, glaciers are divided into different elevation sections (ES) for the application. In addition to its spatial distributed applicability the method is mass-conserving and can be applied with glacier outlines and glacier thickness data only (Li et al., 2015; Seibert et al., 2018a).

However, no comprehensive evaluation of SWAT-GL has been conducted to date. As glaciers and high-mountainous catchments are usually rather data-scarce (Tuo et al., 2016), testing the performance of glacio-hydrological models for long observed time series of good quality is challenging. Moreover, in many cases the available variables to calibrate and validate the model are limited to discharge only, with these gauges often located much further downstream and not close to the glacier. Evaluating the glaciological routines of glacio-hydrological models by discharge alone, representing a superposition of multiple signals, might be problematic, as it might not reflect the signatures which would be visible in the glaciological components. In other words, a satisfactory representation of discharge used to evaluate catchment glaciology (or other processes), albeit often done, might be inadequate. Nevertheless, a sound evaluation of newly introduced schemes in glacio-hydrological models (e.g., glacier components into a hydrological model) should be desired and aimed for. If a mass balance module is implemented together with an evolution module (as in SWAT-GL), in the best case both are evaluated individually and complemented by discharge assessments.

The USGS Benchmark Glacier Project (O'Neel et al., 2019) is a promising attempt to overcome current limitations in data accessibility and modelling efforts of high-mountainous and glaciated basins. The project involves five glaciers, four where long-term measurements are available and one for which the project expanded more recently. The glaciers, namely Gulkana, Wolverine, Lemon Creek, South Cascade and Sperry Glaciers, are located across the Northern United States and thus characterized by various climate regimes (O'Neel et al., 2014; O'Neel et al., 2019). As each glacier is situated within the catchment of a close-by discharge gauge, they are well-suited for glacio-hydrological studies. Long-term hydrological, meteorological, glaciological as well as geodetic measurement are available for each glacier, which range back to the 1950s in terms of mass balance and glacier area observations. O'Neel et al. (2019) found that mass loss is not only present from the beginning of the measurements but has actually increased for four of the five glaciers since the 1990s. A trend which is likely to continue under global temperature projections (Tebaldi et al., 2021).

This paper thus aims to benchmark the recently developed SWAT-GL with respect to its glaciological components. Thus, a sensitivity analysis (SA) using the method of Morris (Morris, 1991), or Elementary Effects (EE), is conducted for screening and ranking of the new input factors under different conditions. Besides, the model is evaluated against long-term glacier mass balance and glacier area-altitude (hypsometry) measurements. Due to the feedbacks between the snow and glacier routine, the model evaluation also considers the performance to simulate snow cover. Discharge is used to cross-validate SWAT-GL under the hypothesis whether a well-performing snow and glacier routine in alpine catchments is sufficient to reproduce discharge in this kind of environments. The benchmarking is performed for four highly glaciated catchments across the US based on the USGS Benchmark Glacier Project (O'Neel et al., 2019).

7.2 Materials & Methods

In the following we will briefly introduce the USGS Benchmark Glacier Project, the chosen datasets for the benchmarking, the study area as well as the benchmarking and SA approach.

Table 7.1: Overview of datasets used. P represents precipitation (mm), T temperature (°C), Q discharge (m^3/s) and SC snow cover (%). Glaciological data is a merged representation of annual net mass balance change (B_{gl} in m w.e.), total annual glacier area (A_{gl} in km^2) and annual glacier hypsometry (H_{gl} in km^2 at a specific elevation range). *var.* indicates that measurements stem from various locations or refer to the whole glacier. The elevation in the glaciological dataset section refers to the total glacier elevation range.

Glacier	Basin Site	Variable	Time Step	Lat	Lon	Elevation [masl]	Temporal Coverage	Missing [%]
Meteorological								
GG	On-site	P	Daily	63.26	-145.41	1,480	1964-2022	9
		T						12
WG	On-site	P	Daily	60.39	-148.94	990	1964-2022	12
		T						14
LCG	Juneau Airport	P	Daily	58.35	-134.56	6	1936-2022	<1
		T						<1
SCG	Diablo Dam	P	Daily	48.71	-121.14	272	1914-2022	<1
		T						<1
Hydrological								
GG	Phelan Creek	Q	Daily	63.24	-145.47	1,127	1966-2023	19
WG	Wolverine Creek	Q	Daily	60.37	-148.9	366	1964-2023	57
LCG	Lemon Creek	Q	Daily	58.39	-134.42	204	1951-2023	41
SCG	SF Cascade	Q	Daily	48.37	-121.07	1,613	1957-1993	28
Snow								
All	Basin mean	SC	Monthly	-	-	-	2002-2023	0
Glaciological								
GG	On-site	B_{gl}	Annual	*var.*	*var.*	1,185-	1966-2022	0
		A_{gl}				2,420	1965-2022	0
		H_{gl}						0
WG	On-site	B_{gl}	Annual	*var.*	*var.*	466-	1966-2022	0
		A_{gl}				1,653	1965-2022	0
		H_{gl}						0
LCG	On-site	B_{gl}	Annual	*var.*	*var.*	543-	1953-2022	0
		A_{gl}				1,550	1946-2022	0
		H_{gl}						0
SCG	On-site	B_{gl}	Annual	*var.*	*var.*	1,619-	1959-2022	0
		A_{gl}				2,439	1950-2022	0
		H_{gl}						0

7.2.1 Datasets & USGS Benchmark Glacier Project

The study is based on the USGS Benchmark Glacier Project (O'Neel et al., 2019) which provides data for five long-term monitoring glaciers across the Northern United States (McNeil et al., 2016; Baker et al., 2018). The five sites are distributed over Alaska, Washington and Montana and thus represent coastal as well as inland locations. Long-term meteorological, geodetic and glaciological measurements starting from the 1950s or 1960s onward are available for four of the glaciers. For the relatively new Sperry Glacier in the program only short time series (from 2005 on) are available: therefore it was excluded in this study. In the following we only refer to the Gulkana (GG), Wolverine (WG), South Cascade (SCG) and Lemon Creek (LCG) glaciers. Seasonal mass balance estimates are derived from geodetically calibrated, conventional glacier-wide mass balance observations (McNeil et al., 2016). The project combines measurements with homogeneous data processing methods to allow for inter-glacier comparisons. An overview of the acquisition years of the geodetic surveys can be found in O'Neel et al. (2019). Glaciological field visits of each glacier take place every spring and fall. Summarizing, the following glaciological variables were used from the USGS Glacier Benchmark Project, total annual glacier area (km^2), annual net mass balance change (m w.e.), annual glacier hypsometry (km^2 at a specific elevation range) (see Table 7.1). Glacier hypsometries hereby represent the area-altitude distribution of

the glacier.

Continuous daily meteorological time series (precipitation & temperature) are directly available on-site for the GG & WG. However, although on-site measurements are also available for the LCG & SCG, the time series are rather short and show a relatively high amount of missing values. For reasons of comparability we follow the approach of O'Neel et al. (2019) and use the closest representative station, which is Juneau Airport (LCG) and Diablo Dam (SCG), respectively. The latter two are also part of the official data of the USGS Benchmark Glacier Project (Baker et al., 2018).

However, as SWAT-GL needs minimum daily (Tmin) and maximum daily temperature (Tmax) which was not continuously available in the project for GG (starting 1995) and WG (starting 1997), a regression model was established to produce continuous daily Tmin and Tmax time series. In detail, daily mean temperature was used as predictor of either Tmax or Tmin in the period where all three variables were available. Subsequently, the regression model was used to predict Tmax and Tmin backwards for the periods before 1995 (GG) and 1997 (WG), respectively. Data gaps of up to three days were linearly interpolated and longer gaps were regressed using daily data from the closest meteorological station of each glacier. The approach is similar to O'Neel et al. (2019), with the only difference that they used monthly regression for longer gaps. We also investigated the potential of ERA5-Land (Copernicus Climate Change Service, 2019) data for gap-filling which was inadequate especially due to a significant precipitation excess throughout the year compared to the station data (not shown in this study).

Hydrological data was obtained from the USGS National Water Information System (U.S. Geological Survey, 1994). We used the closest available gauge for each glacier to determine the total basin area. In detail, the discharge data of the Phelan Creek (representing the GG basin), Wolverine Creek (WG basin), Lemon Creek (LCG basin) and the SF Cascade (SCG Basin) was used. Details about the meteorological and hydrological sites and time series are found in Table 7.1.

Snow cover (SC) data was derived from the MOD10A1 & MYD10A1 V061 NDSI (Normalized Difference Snow Index) products (500 m resolution). The NDSI is based on optical sensors from MODIS (Moderate Resolution Imaging Spectroradiometer) and is calculated as the difference between the reflection in the green spectrum (GREEN) and the shortwave infrared (SWIR) divided by the sum of the two (Dozier, 1989).

$$I_{NDSI} = \frac{B_{GREEN} - B_{SWIR}}{B_{GREEN} + B_{SWIR}} \tag{7.1}$$

where I_{NDSI} is the NDSI and B_{GREEN} and B_{SWIR} are the green and SWIR bands, respectively. For the classification of snow or no-snow pixels a NDSI threshold of 0.4 was used, where values above 0.4 (Hofmeister et al., 2022) indicate snow pixels and smaller values are classified as snow-free. Daily fractional SC (%) on the basin and subbasin scale was then calculated as the average of snow-covered pixels within each basin. Subsequently monthly aggregates were produced. MODIS NDSI data was available from 2002 up to now. A full overview of all datasets is given in Tab. 7.1.

Auxiliary datasets used were elevation data from the Shuttle Radar Topography Mission (SRTM) (NASA JPL, 2013a), the Randolph Glacier Inventory V6 (RGI) (RGI Consortium, 2017) as well as ice thickness estimates from Farinotti et al. (2019).

7.2.2 Study Area

The location of the USGS Benchmark Glaciers combined with the location of the corresponding hydrological and meteorological stations used for each glacier is shown in Fig. 7.1. Note that the representative meteorological stations which were used to force the SWAT-GL models of the SCG and LCG are not

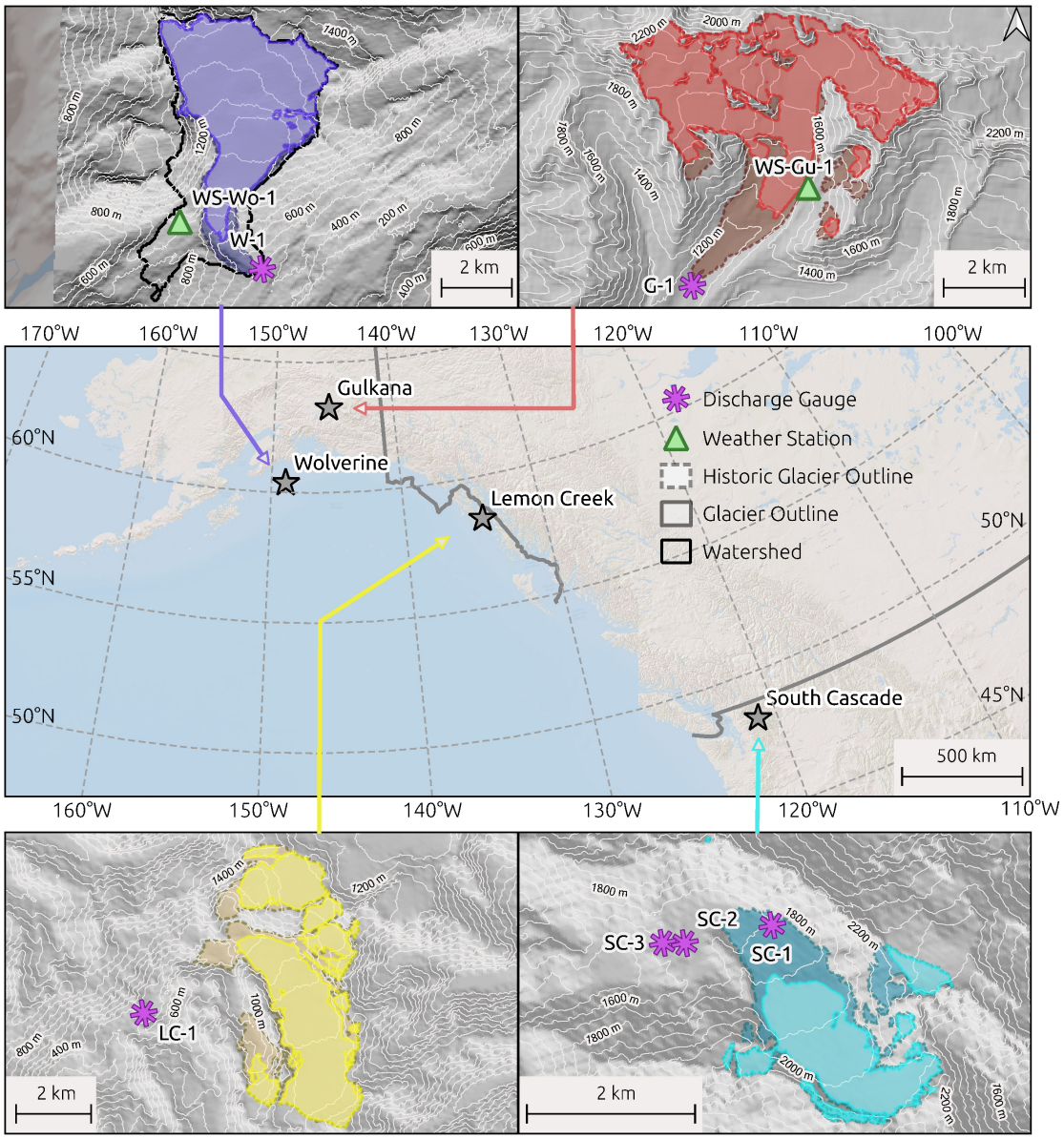

Figure 7.1: Overview of the four USGS Benchmark Glaciers used in this study. Note that the SCG & LCG meteorological stations that were used are remote stations which is why they are not visible in the map. The transparent outline refers to a historical date, the filled outline to a recent date. The dates are: 1957/2021 GG, 1950/2020 WG, 1948/2021 LCG, 1958/2021 SCG.

shown as they are situated outside of the basin boundaries (see O'Neel et al. (2019) for more details). Besides, the map contains the basin boundaries for each glacier which were used as model domain. The total glacier area in each catchment is slightly higher than the individual glacier area of each glacier, as the basins can include several adjacent glaciers. However, the main glacier fraction can be accounted to the four benchmark glaciers in each basin.

The basins have an area of 28.4 km^2 (GG, 64% glaciated 2009), 23.9 km^2 (WG, 69% glaciated 2006), 29.3 km^2 (LCG, 50% glaciated 2005) and 5.9 km^2 (SCG, 58% glaciated 1958). Basin-wide glacier fractions were determined using Randolph Glacier Inventory data (RGI Consortium, 2017). Each glacier hereby represents a distinct climate regime, where the most northward located GG is characterized by a continental (high-latitude) climate (O'Neel et al., 2014; O'Neel et al., 2019). WG, in contrast, is characterized

by a maritime (high-latitude) climate regime (O'Neel et al., 2014; O'Neel et al., 2019). LCG represents another high-latitude maritime glacier, while SCG represents a mid-latitude maritime glacier (O'Neel et al., 2019; Horlings, 2016). All glaciers are retreating, where the SCG shows the strongest relative recession with a glacier area loss of more than 40% (1.3 km²) since 1950. GG lost around 18% (3.3 km²) of its area since 1965, LCG decreased by 16% (3.3 km²) from 1946 up to now and WG receded around 12% (2 km²) compared to 1965. Glacier recession magnitudes reveal a gradient from North to South (O'Neel et al., 2019). Mass balance rates are provided in Fig. 7.3, which show an increasing negative (statistically significant) trend at all sites (O'Neel et al., 2019). According to O'Neel et al. (2019) total uncertainty, consisting of a geodetic and glaciological component, in the mass balance estimates is around 0.2 m w.e. a^{-1} except for GG where it it higher with 0.4 m w.e. a^{-1}.

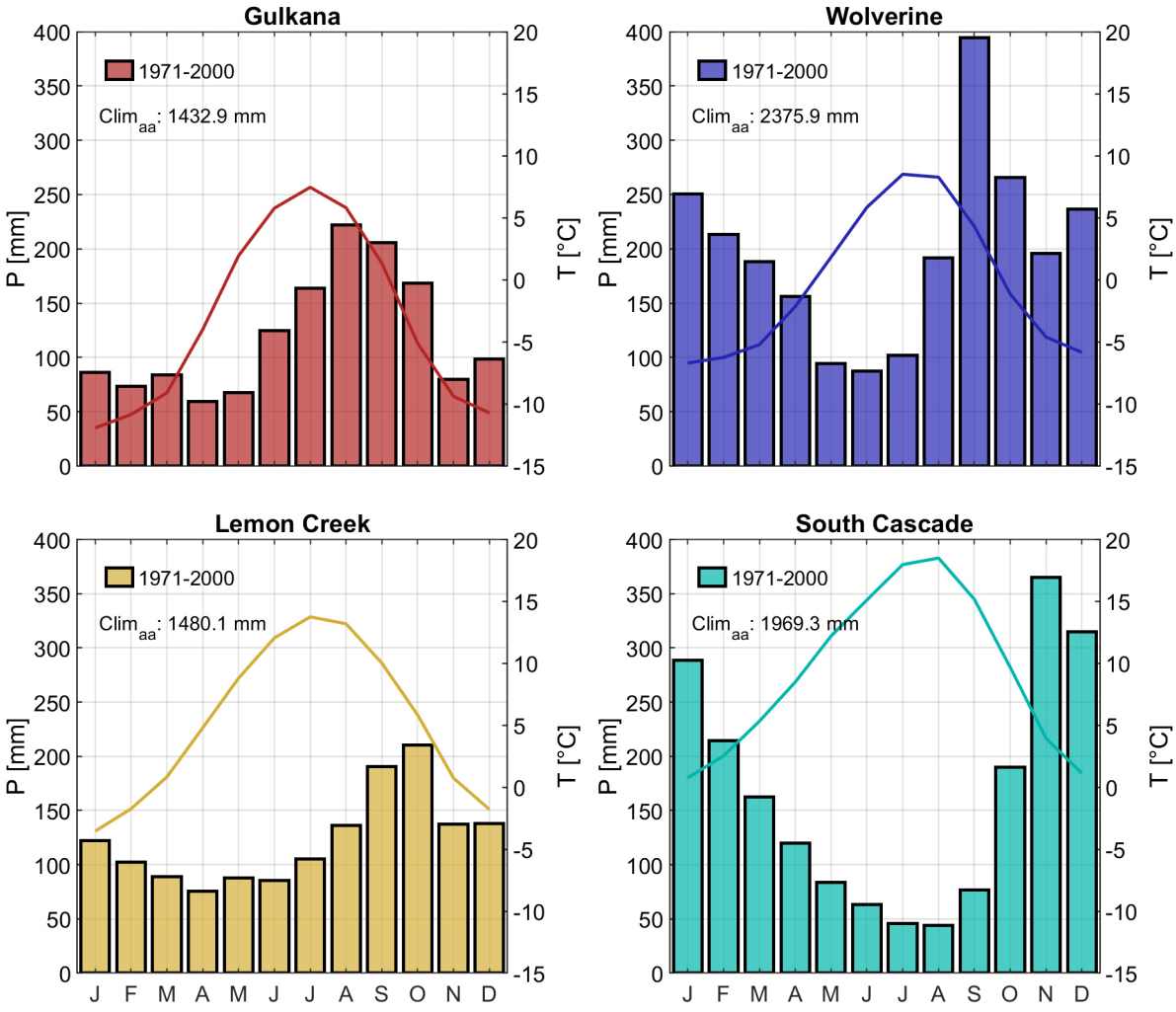

Figure 7.2: Overview of the mean climate regimes of the four glaciers according to the stations of Tab. 7.1 and the reference period 1971-2000. Lines refer to monthly mean temperatures and bars to mean monthly precipitation sums. It has to be noted that no lapse rates were applied, what causes the high monthly mean temperatures of LCG and SCG as remote and lower-elevated stations were used (20 and 40 km apart, respectively). The letters from J to D correspond to the months January to December.

The following mean climate characteristics of each basin were evaluated based on the meteorological

stations listed in Tab. 7.1 and the period 1971-2000. The continental GG with an annual average precipitation of 1,480 mm and a peak in August/ September is significantly drier than its maritime counterparts. SCG with an annual average of 1,970 mm and WG with an annual mean of 2,375 mm show the highest precipitation values among the four. While WG has its precipitation peak roughly in September/ October it also shows consistent high precipitation during the Winter months and a dry period in summer. SCG is also characterized by a summer low, followed by an increase in precipitation during Autumn and ending in a strong late Autumn and Winter peak (November - January). The annual precipitation totals of the LCG are around 1,480 mm with a less pronounced peak in Autumn. The glacier generally has a lower gradient between wet and dry period, which makes precipitation more evenly distributed throughout the year. In terms of temperatures, all glaciers reach their maximum in either July or August and their yearly minimum in January. The mean climates are illustrated in Figure 7.2. However, it should be noted that no lapse rates have been applied for the climate classification, the meteorological station used for LCG and SCG come with an elevation difference of more than 500-1,500 m (LCG) and 1,300-2,100 m (SCG). Lapse rates (temperature and precipitation) were later calibrated through the optimization procedure (see section 7.3.3).
All gauges belong to intermittent streams, which can fall dry during the winter months (Fig. 7.3). While the corresponding streams of GG and WG had almost no flow in the available time series (see Tab. 7.1) from December to April/May, the streams of SCG and LCG carried water sporadically during these months. Annual average flows are 2.4 m^3/s (GG), 2.7 m^3/s (WG), 5.5 m^3/s (LCG), 28.1 m^3/s (SCG) evaluated for the period 2002-2022 for all glaciers except SCG, where the years from 1972-1992 had to be chosen. Inter-annual variability is highest at GG (Coefficient of Variation (CV) of 0.21) and lowest at SCG (CV of 0.13). Except in the SC basin, we can see a tendency of a slight shift in the flow period towards an earlier onset of the melt season.

7.2.3 SWAT-GL

The recently developed SWAT-GL (Schaffhauser et al., 2024b) is a modified version of the traditional hydrological model SWAT (Arnold et al., 1998), which includes glacier dynamics based on the Δh approach developed by Huss et al. (2010). It basically consists of two modules, a mass balance and a glacier evolution module. Mass balance estimations are based on a degree-day approach, similar to the already existing snow routine of SWAT. Glacier evolution is implemented by means of the Δh approach (Huss et al., 2008, 2010). For detailed technical explanations we refer to Schaffhauser et al. (2024b), as we only provide a short summary of the main points here.
In general, the mass balance is formulated as:

$$W_t = W_{t-1} - M_t \cdot (1 - \beta_f) - S_t + C_t \tag{7.2}$$

with W_t being the water equivalent of ice [mm $H_2O \cdot d^{-1}$] at day t, M_t represents the melt rate [mm $H_2O \cdot d^{-1}$], S_t represents the sublimation rate [mm $H_2O \cdot d^{-1}$], C_t refers to the accumulation rate [mm $H_2O \cdot d^{-1}$] and β_f is an adjustable refreezing factor of ice during melt periods. M_t is calculated analogously to snow melt in the standard SWAT using a distinct melt factor.
The physically-based Δh-parameterization is able to simulate spatially distributed glacier retreat. The core of the approach is that glaciers are discretized in elevation sections (ES), where each has an inherent storage and receives distinct ice thickness changes. The ES are normalized for the glacier elevation range and a characteristic (normalized) ice thickness change is assigned to each zone, according to:

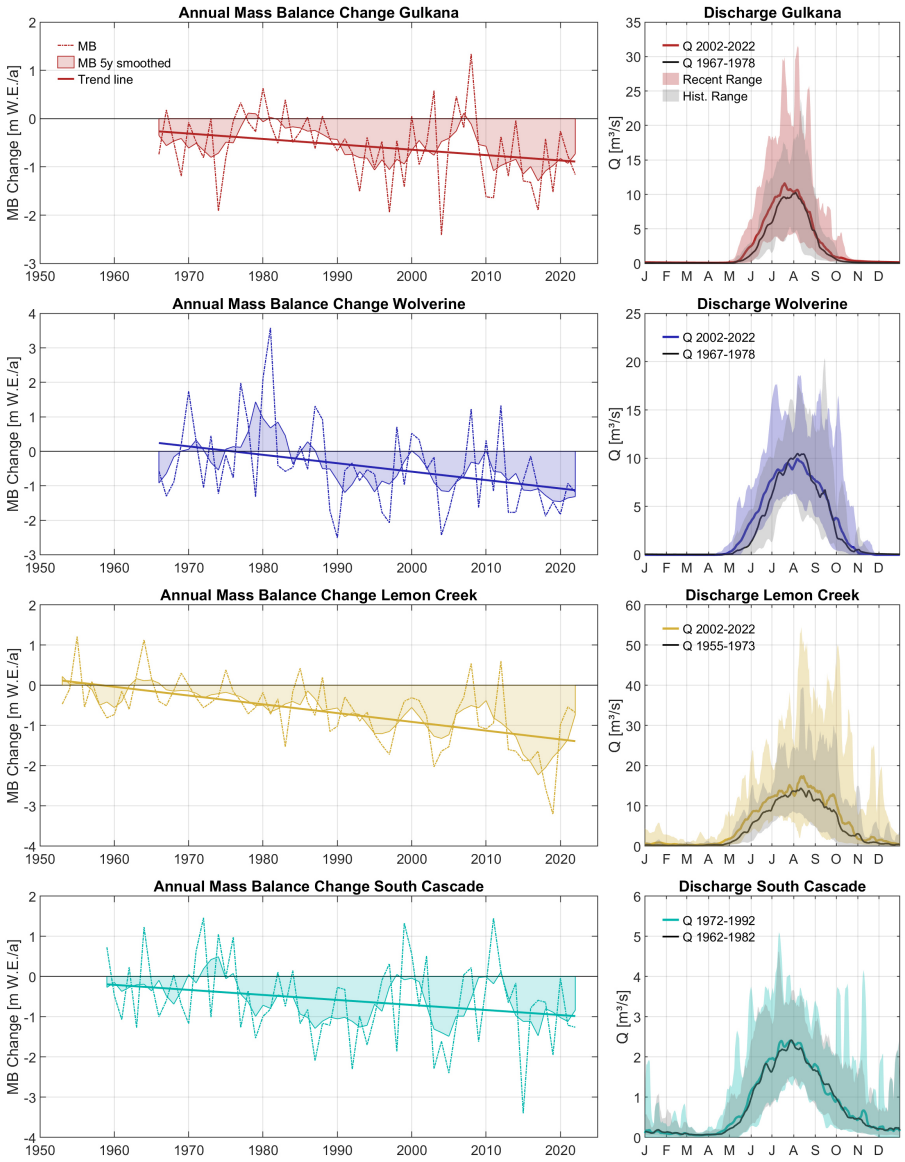

Figure 7.3: Overview of the annual mass balance rates of all glaciers merged with the mean daily discharge of two periods for each glacier. The recent periods refer to 2002-2022 (GG, WG, LCG) and 1972-1992 (LCG) and the older periods refer to 1967-1978 (GG, WG), 1955-1973 (LCG) and 1962-1982 (SCG) as indicated in the day of the year plots.

$$E_{norm,i} = \frac{E_{max} - E_i}{E_{max} - E_{min}} \tag{7.3}$$

with $E_{norm,i}$ is the normalized elevation of ES i [-], E_{max} and E_{min} refer to the maximum and minimum glacier elevation [m] and E_i is the actual elevation of ES i [m].

Lower altitudes hereby receive stronger ablation than higher ones. The characteristic ice thickness change for each normalized elevation varies with glacier size. One of three parameterizations is used separately for each glacier, which are thus classified as small (<5 km²), medium (5-20 km²) or large (>20 km²). The empirical relationship is illustrated in Fig. 7.4. It is important to note that the Δh-parameterization is called annually (at the end of a glaciological year) to redistribute (lumped) annual

mass balance changes over the individual ES of a subbasin to simulate glacier retreat. The normalized ice thickness change formulas of Fig. 7.4 follows the general form

$$\Delta h_i = (E_{norm,i} + a)^y + b \cdot (E_{norm,i} + a) + c \tag{7.4}$$

where a, b, c, y are coefficients which vary for glacier size and Δh_i represents the normalized ice thickness change for an Elevation E_i. We use the parameters based on Huss et al. (2010). Theoretically the parameters could be derived specifically for any glacier if the required data is available (e.g. two DEMs at different dates). The dimensionless ice thickness change Δh_i is rescaled using a scaling factor f_s [m] to receive the change in meters for every glaciological year (Eq. 7.5).

$$f_s = \frac{V_a}{\sum_{i=1}^{n} A_i \cdot \Delta h_i} \tag{7.5}$$

with V_a referring to the annual glacier volume change expressed in water equivalent [m^3], that is calculated by multiplication of annual EW values (see Eq. 7.2) and the subbasin area. A_i is the area of ES i with n being the total number of ES. Annual ice thickness changes are then calculated via

$$h_{i,1} = h_{i,0} + f_s \cdot \Delta h_i \tag{7.6}$$

where $h_{i,1}$ is the updated ice thickness [m water equivalent] after each glaciological year of ES i, $h_{i,0}$ is the ice thickness [m water equivalent] in ES i before the application of Δh parameterization. If $h_{i,1}$ is ≤ 0 the ES is assumed ice-free causing an update of the glacier extent.

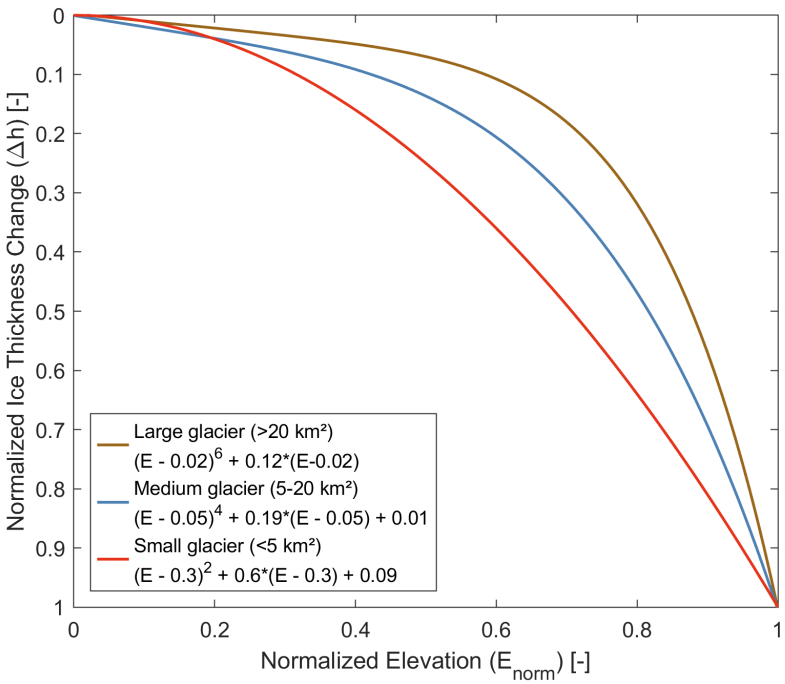

Figure 7.4: Empirical relationship of normalized glacier elevation and the normalized ice thickness change based on Huss et al. (2010).

SWAT-GL classifies glaciers on the subbasin scale, meaning that a simplified assumption is used where all glaciated areas within a subbasin are considered as one glacier object. The implementation of the glacier routine in SWAT introduces five new parameters which control glacier melt (and refreezing) and accumulation. Besides, one new output file containing annual glacier mass balance information (for each glaciological year) as well as two new input files which require some preprocessing are introduced. The two input files refer to the parameterization on the HRU scale as well as the glacier initialization with respect to hypsometry, ice thickness and volume. The source code of SWAT-GL together with an example is freely accessible via GitLab `https://gitlab.com/lshm1/swat-g`.

7.2.4 Sensitivity Analysis

In order to provide a comprehensive picture of SWAT-GL we also performed a global sensitivity analysis using the Method of Morris or Elementary Effects (EEs) (Morris, 1991; Saltelli et al., 2008). The method which is based on a multiple-starts perturbation approach and thus belongs to the one at a time (OAT) methods is able to determine approximate sensitivities at a relatively low computational cost (Saltelli et al., 2008; Pianosi et al., 2016). The EEs test has thus been established as a robust method for screening and ranking of the input factors (Pianosi et al., 2016).
For sampling we used the radial design, where r sample points from a Latin hypercube sampling serve as well-spread starting points in the input space (Campolongo et al., 2011). The total sample size N follows the form $r(M+1)$, with M being the number of input factors. For r a value of 500 was chosen, resulting in 7,500 model evaluations ($M = 14$) (Sarrazin et al., 2016). The basic idea of the radial design is that from a starting point, one factor M is varied while keeping all others fixed. This results in M steps (varying each factor once) that are performed for each sampling point (r). In general, r EEs are calculated per input factor which are then averaged to provide a global sensitivity metric μ_i for each input factor i. The calculation itself is based on finite differences. To account for non-monotonic effects in the model, μ_i^* is used based on the absolute values of the EEs (Campolongo et al., 2011). The formulation is:

$$\mu_i^* = \frac{\sum_{i=1}^{r} |\mathrm{EE}_i|}{r} \tag{7.7}$$

An EE of an input factor i can be calculated as follows:

$$\mathrm{EE}_i = \frac{Y(X_1, X_2, ..., X_{i-1}, X_i + \Delta, ..., X_M) - Y(X_1, X_2, ..., X_M)}{\Delta} \tag{7.8}$$

with $X = (X_1, X_2, ..., X_M)$ being the individual values of the factors and Δ the step (or perturbation). We also calculated the standard deviation σ_i of the EEs as a proxy of the interaction of input factor i with the other factors. It further describes whether the model output is linearly or non-linearly affected from an input factor (Garcia Sanchez et al., 2014; Merchán-Rivera et al., 2022). To avoid confusion with other sigmas used in this work we define sigma of an EE as $\sigma_{EE,i}$. The ratio of $\sigma_{EE,i}/\mu^*$ also provides insights on whether the effect of a factor is monotonous or almost monotonous. This is especially important to derive information whether the interaction of snow and glacier processes are represented (adequately). Due to the different scales of the input factors, the standardization from Sin and Gernaey (2009) was applied. The SA is used to support and complement the benchmarking procedure to i) determine the importance of the newly introduced parameters under different conditions, ii) rank the parameters to get insights into how the most influential parameters differ between the catchments and iii) identify whether the glacier routine interacts appropriately with the snow routine. For the SA the combined calibration period from 2002-2015 and validation period from 2016-2022 was used.

7.2.5 Calibration and Validation Procedure

The core of the study is to evaluate the capabilities of SWAT-GL to simulate multiple glaciological components in glaciated catchments. This not only involves the assessment of multiple performance criteria in the four glaciated catchments, but is also complemented by a SA in order to demonstrate the need of a robust glacier representation. In addition, the benchmarking is also used to evaluate potential model weaknesses of SWAT-GL that need to be should be addressed in future developments of the model.
Although hydrological models traditionally focus on the simulation of discharge, which is also one of the goals of SWAT-GL, we will evaluate the glacier routine mainly in terms of representing glacier and snow processes. Being aware of the relevance of discharge, the variable will be presented for cross-validation purposes throughout the paper under the hypothesis that an adequate (at least monthly to annual) discharge representation in glacier and snow-dominated catchments can be achieved by a reasonable representation of the snow and glacier components alone. In detail, we calibrated and validated each of the four models based on snow cover (monthly), glacier mass balance variations (annual) and glacier hypsometries (annual). For an adequate representation of the snow and glacier routine, monthly snow cover estimates were considered to be appropriate. As mentioned, the effects on discharge (daily to annual) are provided in addition to evaluate model consistency. Thus, only parameters affecting the newly introduced glacier routine as well as SWAT's snow processes, due to its strong interactions with the glacier component, were used (see Tab. 7.3). As a result 14 parameters were considered. A model run comprised the full available glaciological time series of each catchment, due to differences in the availability of snow cover and the glaciological components. For example, the glaciers were initialized for the starting year of the mass balance time series (GG 1966, WG 1966, LCG 1953, SCG 1959), while MODIS SC was available from 2002 onward. SC was therefore calibrated from 2002-2015 and validated for the period (2016-2021). For glacier mass balances and glacier hypsometries two calibration phases were used, one which was the same period as for snow cover (2002-2015) and one at the beginning of the model run (GG 1971-1985, WG 1971-1985, SCG 1962-1976, LCG 1953-1967). The remainder of the time series was then used as validation phase (one period matching the one of snow cover (2016-2021) and one covering the remaining time series at the end of each glaciers first calibration phase up to 2001). The second validation phase was used in order to make full use of the available information and to assess SWAT-GL over a long time scale. The summary of the temporal settings is given in Table 7.2. Note that the periods for discharge were intended to match those of the other variables, data restrictions however did not allow for a perfect match.

Table 7.2: Overview of the calibration and validation phases. Note: As discharge was used for cross-validation purposes only, it has two validation phases rather than a calibration and validation phase. For the SCG no discharge data was available in the 2000s leading to only one validation period. The asterisk for the WG indicates that validation period 1 is rather poor due to mainly missing values.

Glacier	Mass Balance		Snow Cover		Hypsometry		*Discharge*	
	Calibration	Validation	Calibration	Validation	Calibration	Validation	Validation I	Validation II
Gulkana	1971 - 1985	1986 - 2001	2002 - 2015	2016 - 2021	1971 - 1985	1986 - 2001	1971 - 1978	1990 - 2001
	2002 - 2015	2016 - 2021	-	-	2002 - 2015	2016 - 2021	2002 - 2015	2016 - 2021
Lemon Creek	1953 - 1967	1968 - 2001	2002 - 2015	2016 - 2021	1953 - 1967	1968 - 2001	1953 - 1967	1968 - 1973
	2002 - 2015	2016 - 2021	-	-	2002 - 2015	2016 - 2021	2002 - 2015	2016 - 2021
South Cascade	1962 - 1976	1977 - 2001	2002 - 2015	2016 - 2021	1962 - 1976	1977 - 2001	1962 - 1976	1977 - 1992
	2002 - 2015	2016 - 2021	-	-	2002 - 2015	2016 - 2021	-	-
Wolverine	1971 - 1985	1986 - 2001	2002 - 2015	2016 - 2021	1971 - 1985	1986 - 2001	1971 - 1981	1986 - 2001*
	2002 - 2015	2016 - 2021	-	-	2002 - 2015	2016 - 2021	2002 - 2015	2016 - 2021

We used a Multi-Objective Automatic optimization (MOO) procedure, where each of the three prescribed variables referred to one objective. The optimization was performed using the widely used evo-

lutionary NSGA-II (Nondominated Sorting Genetic Algorithm) algorithm (Deb et al., 2002). Based on nondomination sorting and the introduction of a crowding distance operator to favor solutions which are less-crowded (high crowding distance), NSGA-II iteratively finds solutions which are uniformly spread at the Pareto front. The population size of a generation was 100 and the maximum number of generation was set to 100. We used Simulated Binary Cross Over with a cross over probability of 0.9 and Polynomial Mutation with a mutation probability of 0.3.

For all variables a normalized form of the Root-Mean-Square Error (NRMSE), based on the standard deviation of the observations of each variable, was consistently used as objective function (OF) (Eq. 7.10). The NRMSE increases comparability between the individual OFs and allows to minimize the residuals between observed and simulated values of all variables. For SC we excluded winter months in the optimization procedure to put more weight on the months where snow cover is dynamic as snow cover is usually 100% from December to at least April. The effect is more pronounced in basins with less snow cover dynamics and therefore a relatively high minimum summer SC. The simulation of a permanent snow cover then leads to good OF values. We will further discuss this issue during the paper. The standard form of the RMSE can be defined as follows:

$$\mathrm{RMSE}_x = \sqrt{\frac{\sum_{t=1}^{n}(O_{x,t} - S_{x,t})^2}{n}} \tag{7.9}$$

with $O_{x,t}$ being observed and $S_{x,t}$ simulated components of variable x, which is either snow cover, glacier mass balance or hypsometry, t refers to the time step (monthly or annual depending on the variable) and n represents the number of available data points. The standardization follows the form:

$$\mathrm{NRMSE}_x = \frac{\mathrm{RMSE}_x}{\sigma_x} \tag{7.10}$$

where σ_x is the standard deviation of the observations of each variable. However, as glacier hypsometries provide areal time series for multiple glacier elevations, the individual RMSE of each elevation was calculated and then averaged to obtain one RMSE value which was standardized in a last step using the standard deviation of observed total glacier area. This gives a more equal weight to all elevations to get rid of solutions where individual elevations might have a high non-standardized error. In other words, if we would use an average of the NRMSE of all individual elevations, those with small observed standard deviations (e.g., higher-elevated and less dynamic ones) could lead to an excessive degradation of the overall OF.

SWAT-GL needs distributed glacier thickness and glacier area information as input for each ES and subbasin. However, this data is usually not easily available for various years. In most cases only globally and openly available datasets such as glacier areas from the RGI (RGI Consortium, 2017) and ice thickness from Farinotti et al. (2019) or Millan et al. (2022), representing a fixed point in time, are available to use. Alternatively, if geodetic information is available at different times glacier thickness can also be directly inferred for each of them. As the USGS Benchmark glacier project provides geodetic data for several years we have chosen the best available DEM closest to the mass balance observation start of each glacier. Best hereby refers to a full coverage of the glacier basins along with a minimum of missing values. The thickness was then estimated using the GlabTob2 model (Linsbauer et al., 2009, 2012; Frey et al., 2014). Glacier outlines were also available at several times in the Benchmark project and the year closest to the chosen DEM was selected to initialize SWAT-GL. If a mismatch between DEM and glacier outline acquisition year and mass balance observation start was present, we corrected the initial glacier volume by adding the mass loss or gain that happened since the observation start to the DEM acquisition year. E.g., if mass balance measurements started 1966 and the best DEM and glacier outline was available from

1977, the cumulative mass balance estimates until 1977 were added to the initial volume. The volume was then distributed to the individual ESs while maintaining the original volume fractions of the bands in the total volume. ES sections were defined with a spacing of 100 m.
Results for all variables will particularly focus on the last generation of the optimization and the best simulations of each variable. Apart from the described methodology we tested SWAT-GL's ability to reproduce observed mass balance nonstationarities, inter-annual variability and the monotonic relationship between simulated and observed mas balance. For this purpose the full time series was used, homogeneity was tested with the Wilcoxon-Rank Sum Test (WRS) (Wilcoxon, 1945), the Pettitt Test (Pettitt, 1979) and trends were detected using a modified Mann-Kendall version from Hamed and Ramachandra Rao (1998) which considers autocorrelation. For the WRS the first calibration and the last available validation periods were used. Lastly, for demonstration purposes we performed a Single-Objective Optimization (SOO) using Differential Evolution (DE) (Storn and Price, 1995) with the adaptations described in Dawar and Ludwig (2014) for the WG and two variables, namely mass balance and discharge. The results exemplify what potential users could expect for pure discharge or mass balance calibrations, which is often (albeit questionable) the case in hydrological studies.

Table 7.3: Parameters and their relative ranges used for the benchmarking of SWAT-GL.

Parameter	Description	Minimum	Maximum
SFTMP	Snowfall temperature [°C]	0	4.5
SMTMP	Snowmelt temperature [°C]	0	4.5
SMFMX	Melt factor for snow on June 21 [mm H_2O/(°C·day)]	0.1	7
SMFMN	Melt factor for snow on December 21 [mm H_2O/(°C·day)]	0.1	7
TIMP	Snow temperature lag factor [-]	0	0.5
SNOCOVMX	Snow water equivalent threshold where 100% snow cover occur [mm]	2	75
SNO50COV	Fraction of SNOCOVMX at which 50% snow cover occur [-]	0.1	0.9
TLAPS	Temperature Lapse Rate [°C/km]	-9	-5
PLAPS	Precipitation Lapse Rate [mm/km]	550	1800
GLMLTMP	Threshold temperature for glacier melt [°C]	0	4.5
GLMFMX	Melt factor for ice on June 21 [mm H_2O/(°C·day)]	3.5	13
GLMFMN	Melt factor for ice on December 21 [mm H_2O/(°C·day)]	3.5	10
β_f / f_{frze}	Refreezing factor of glacier melt [-]	0.001	0.01
f_{accu}	Conversion factor of snow to ice [-]	0.1	0.6

7.3 Results

In the following the results of the SA and optimization procedure of SWAT-GL are presented. First, the results of the SA are shown followed by the sections related to the optimization procedure. The results of the optimization focus on the simulations of the last generation (with N=100 simulations/evaluations) unless otherwise stated. Table 7.4 provides a summary of the performance metrics for all glaciers based on the last generation of the optimization. Discharge is discussed separately from the results of SWAT-GL's glacier and snow performance as it was only used for cross-validation purposes.

7.3.1 SWAT-GL's Glacier and Snow Parameter Sensitivity

The SA is based on the EEs method. Our results for all catchments are presented as scatterplot between μ^*, the mean sensitivity of a factor (parameter), and σ as proxy for the interactions of a factor (Figure 7.5).
A common pattern that all catchments share, albeit to varying extents, is their spread around the 1:1

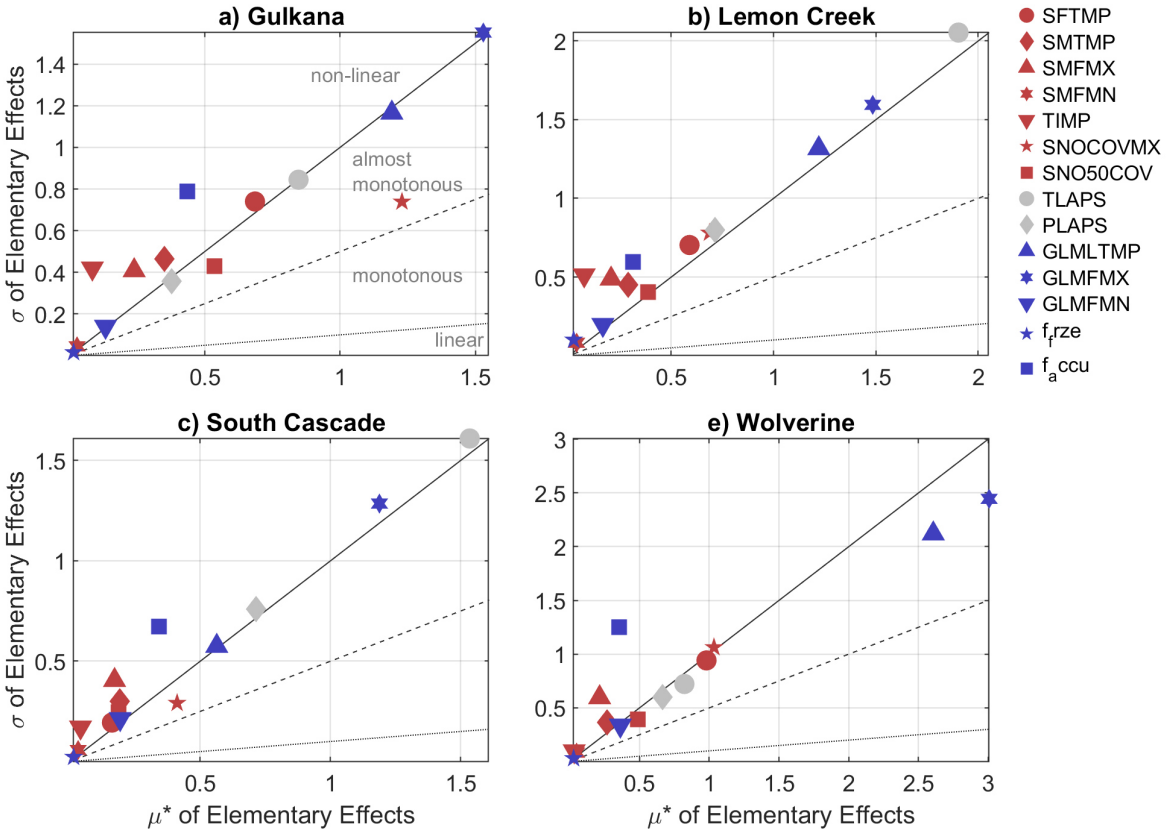

Figure 7.5: SA results based on the EE method for all four catchments and 14 parameters. The slopes (σ_{EE}/μ^*) of the different lines that classify parameter effects on the model outputs as linear, monotonous, almost monotonous and non-linear are: 0.1 (dotted line), 0.25 (dashed line), 0.5 (solid line).

line that differentiates between non-linear and almost-monotonous effects. Moreover, it is shown that in general the more sensitive parameters (larger μ^*) tend to have higher interactions as well as stronger potential non-linear model responses. It is also shown that the model response of all catchments strongly depends on GLMFMX that controls the maximum value of the degree-day factor of ice and thus the amount of glacier melt that can occur at a specific day of the year. In terms of factor ranking GLMFMX is either the most or the second-most influential factor. It is the most important parameter in the WG & GG basins where the respective meteorological stations are located directly at the glacier. However, GLMFMX is substituted by the temperature lapse rate (TLAPS) at the SCG & LCG, where the respective meteorological stations are located outside of the catchment and at a significant lower elevation than the glaciers. Due to the difference in altitude (which is part of the precipitation correction formulation) between station and elevation band centers it is inherent that the lapse rates become more important. Due to the temperature dominance of both, snow and glacier processes, the sensitivity of the precipitation lapse rate (PLAPS) is less pronounced and strongest at the high-elevated SCG. Among the four most important factors is the threshold temperature of glacier melt (GLMLTMP) which controls the onset of melt and has an effect on the timing of melt events as well as the amount of melt. The temperature lapse rate and glacier melt temperature can favor similar conditions or act contradictory (decrease of melt temperature favors earlier melt onset and small or no temperature lapse rate as well). In general, SWAT-GL is strongly temperature-dominated in all catchments.

However, the relevance of precipitation in the SCG basin might be a special characteristic (with regard

to the PLAPS ranking in the other basins). Besides, (with an exception for the SCG) the SNOCOVMX parameter is ranked among the most sensitive parameters in the catchments. The parameter determines a threshold of snow water equivalent (SWE) that is required to cause a 100% coverage of snow. As glacier melt can only occur when the glacier is snow-free the parameter directly affects glacier melt, which explains its relevance in the catchments.
With respect to potential interaction and non-linear model responses the accumulation factor (f_accu) that is responsible for the snow metamorphism (or turnover from snow to ice) takes a dominant role at the WG, GG and SCG.

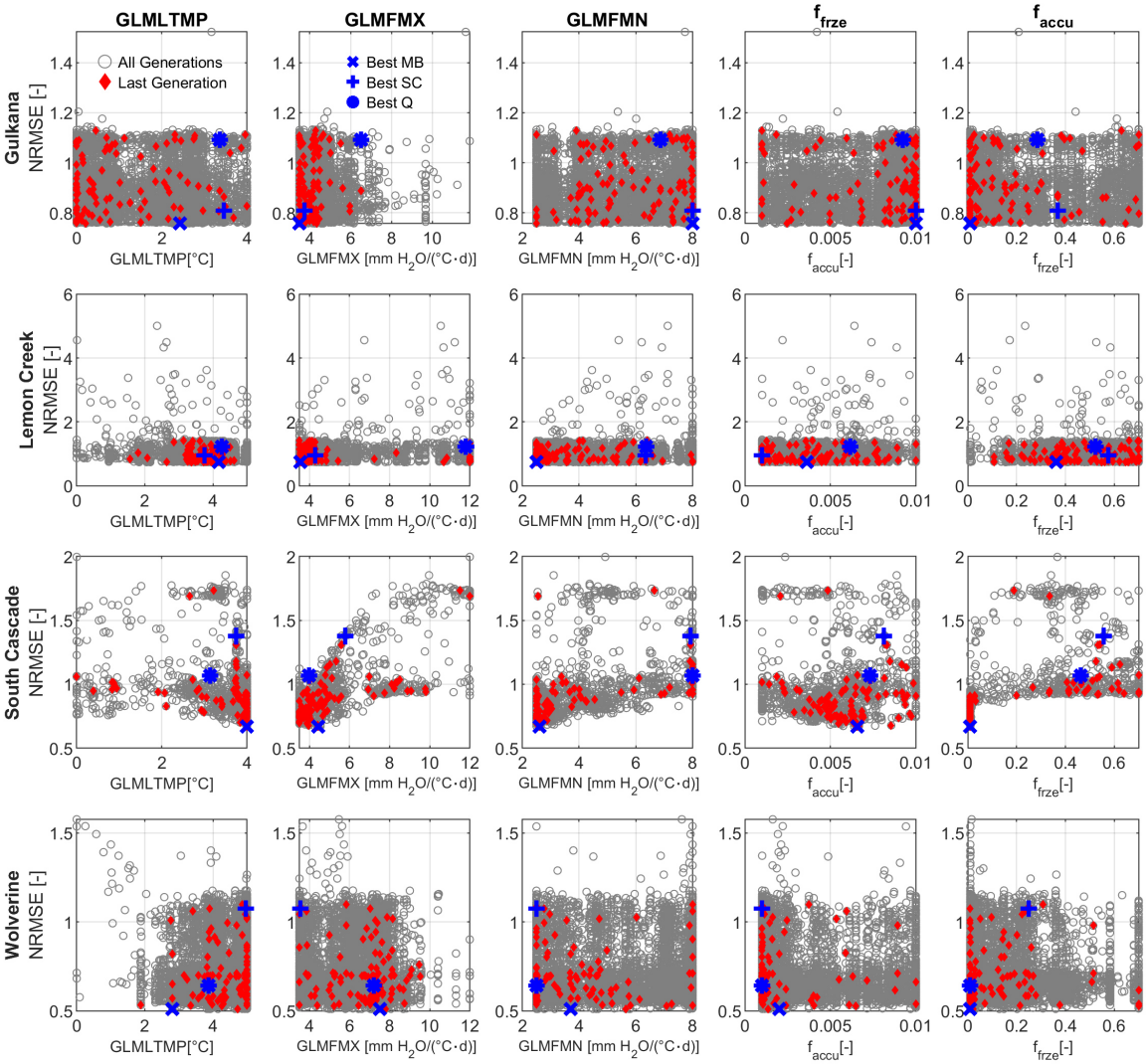

Figure 7.6: Parameter space illustrated for all glacier-related parameters for all generations (grey) and the final generation (red). The shown NRMSE values refer to the results of annual glacier mass balance simulations. The three blue symbols refer to the individual best solutions of mass balance, snow cover and (cross-validated) discharge in the last generation.

This is plausible as it couples snow and glacier processes by transforming a specific fraction of snow lying on the glacier to ice and thus affecting both storages. Although it is not among the most sensitive factors, it can have a high significance in certain situations due to its possible interactions. Although the most influential parameters receive high σ_{EE} values they do not necessarily fall in the non-linear area

(GG, WG). However, all models show generally a non-linear or monotonous behavior and are potentially characterized by interactions rather than a linear relationship.
Furthermore, we can identify at least 6 to 8 less or non-influential parameters which would reduce the dimension of the parameter space to a 6 or 8 dimensional problem in the different models, respectively.

7.3.2 Inter-Basin Comparison of Optimized Glacier Parameters

A comparison between the values of the final parameter sets of all catchments is shown in Fig. 7.6. As the main purpose is to evaluate the glacier routine introduced in SWAT-GL, the comparison is limited to the five glacier parameters. Results are presented for the RMSE values of the annual glacier mass balance only. The parameter values of the GG are relatively well-spaced in the parameter space with an exception for the GLMFMX parameter, which controls the maximum amount of glacier melt. The parameter tends to cluster at its lower boundary for GG, LCG and SCG. The lower bound of GLMFMX is associated with a reduction of strong negative mass balance rates which might lead to an overestimated ablation. Analogously to the GG, the final parameterizations of the LCG are generally well-spread. An exception exhibits here the glacier melt temperature (GLMLTMP). In contrast to the maximum melt factor, the glacier melt threshold temperature groups at its (relatively high) upper bound for the LCG, WG and SCG. The two patterns indicate the necessity of high melt rates which should not occur too early. The final parameter distribution of the WG is more narrow compared to the other catchments. It is shown that especially for the accumulation rate (f_{accu}), small values are desired to avoid large positive mass balance simulations.
Table B.1 provides a detailed overview of the ranges and median values for each glacier parameter and each catchment at the end of the optimization.

7.3.3 Evaluating SWAT-GL's Representation of Glacier & Snow Processes

Table 7.4: Performance of SWAT-GL for all variables and glaciers with respect to the best simulation of the last generation of the optimization procedure. Note: Discharge was not calibrated but is only shown for cross-validation purposes. Discharge thus has two validation phases following the periods assigned to the glacier mass balance evaluation of each glacier (see Section 7.3.3). Cum. B_{gl} refers to the mismatch between observed and simulated cumulative mass balance at the end of the time series and is therefore not attributed to any of the calibration or validation periods. Negative values of Cum. B_{gl} indicate that the model is underestimating mass balance losses.

Glacier	NRMSE [-]						KGE		PBIAS [%]
	Mass Balance		Snow Cover		Hypsometry		Discharge		Cum. B_{gl}
	Calibration	Validation	Calibration	Validation	Calibration	Validation	Validation 1	Validation 2	Complete
Gulkana	0.76	0.68	1.08	0.98	0.34	0.28	0.62	0.62	-21.68
Lemon Creek	0.74	0.70	0.43	0.44	0.70	0.75	0.29	0.19	-1.60
South Cascade	0.67	0.69	0.35	0.36	1.13	0.90	0.82	0.59	-55.17
Wolverine	0.51	0.56	0.87	0.99	2.29	1.79	0.64	0.64	-11.16

The performance of the optimization procedure is shown in Table 7.4. Statistical results for discharge are presented alongside the other variables for demonstration purposes, although it was not part of the optimization procedure. Results for discharge comprise two validation periods which were chosen analogously to the ones of glacier mass balance, if data was available.
With respect to glacier mass balance estimates, lowest NRMSE values were found for the WG model (both, calibration and validation), followed by the SCG model. In contrast, the WG model shows the worst performance w.r.t. hypsometries, for which best performance was reached in the GG basin. Concerning snow cover, the LCG and SCG model achieve the best performance metrics in both the calibration and

validation period (0.43 and 0.44 for LC model, 0.35 and 0.36 for SC model). No significant degradation was observed between the calibration and validation phases for any of the objective functions of all variables included in the optimization procedure. It should be noted that a direct inter-comparison of the absolute NRMSE values between the catchments is difficult, as the standard deviation of the observations used for the normalization has a dominant effect on the values. For example, the standard deviation of the mass balance observations of the WG is a factor of 1.6 - 9.4 higher than that of the other glaciers (calibration period). The same applies for the snow cover results of the SCG, where observed snow cover standard deviations are 1.3 - 2.4 times above those of the other glaciers.

A clearer picture emerges from the graphical representation of the optimization results in Figure 7.7.

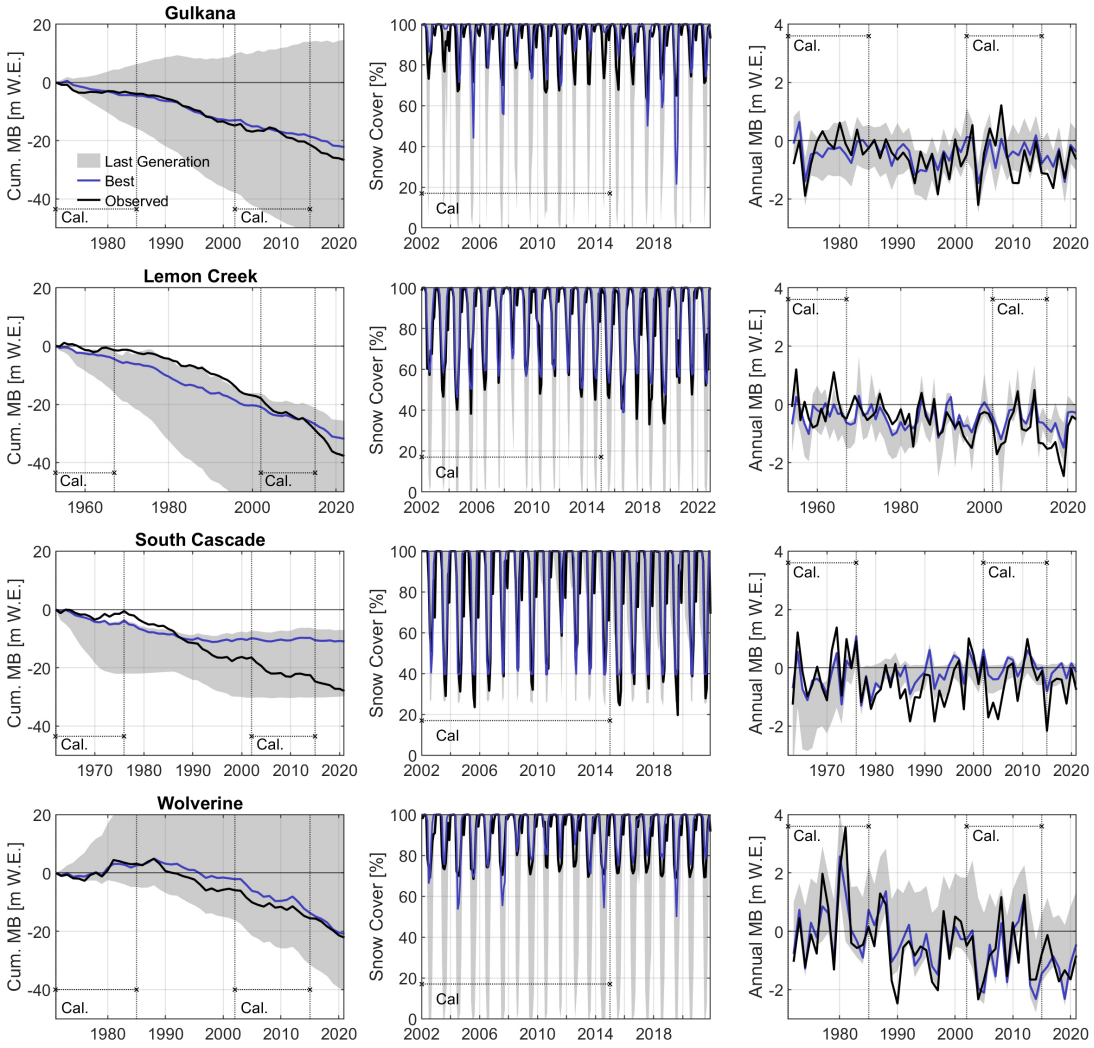

Figure 7.7: Simulation results of the last generation of the optimization procedure for the cumulative mass balance (Column 1), monthly snow cover (Column 2) and annual mass balance (Column 3). Each row corresponds to one glacier. Blue represents the best evaluation of the last generation for the respective variable, black refers to observations and grey shadings indicate the range of all evaluations part of the last generation. The dashed lines with the Cal. annotation indicates the individual calibration phases of each study area. The remainder of each time series was used for validation.

The discrepancy at the end of the simulation periods of the cumulative mass balance ranges from -1.6% to -55.17%. However, the relatively large outlier of -55.17% arises from the SCG, where the mass balance

loss stagnates in the 1990s. The abrupt change in the mass balance is also indirectly reflected in the cross-validated results of discharge. While the KGE in the beginning of the simulation is around 0.82, a significant drop to 0.59 can be observed in the second evaluation period. Problems in the SCG model become especially apparent when focusing on the lower bound of the model range (grey shading in Fig. 7.7 of cumulative mass balance). Here, we can notice a very poor representation of the inter-annual signal as the simulations show two very abrupt drops and long periods of stagnation. It is likely that the glacier has retreated to altitudes which are not subject to temperature-induced melting. The wide range visible in the cumulative mass balance can be caused by only a few solutions and does not allow for conclusions about the real distribution of the final simulations. We thus show in Fig. B.4 the individual cumulative mass balance representations of all optimized solutions together with the distribution of the cumulative mass balance at the end of the simulation period (all 100 values of the last year of simulation for each glacier). It is found, that particularly the upper bound of the GG and WG, as well as the lower bound of the LCG and SCG are caused by a small subset of solutions. The LCG with an almost perfect fit at the end of the simulation period, however, is overestimating ablation over a large part of the simulation period. Overall, the models are underestimating ablation rates after the 2000's (with an exception for the WG). This becomes even more evident by looking at the annual mass balance rates (last column Fig. 7.7). All models perform well in simulating monthly snow cover, in both, the calibration and validation phase. The spread of the models (grey shading) is relatively large and includes simulations with almost no snow cover in summer. The WG consists of a period of positive mass balance in the 1980s which is likely causing the upward tendency of the simulation range (the share of simulations with a positive cumulative mass balance until the end of the simulation).
Summarizing, we assess SWAT-GL's results in general as satisfactory to very good, with the exception of SCG, which shows sharply declining accuracy of the results over the course of the simulation period.

7.3.4 SWAT-GL's Ability to Capture Mass Balance Inhomogeneity and Variability

As long simulation periods were used, we evaluate whether SWAT-GL is capable to capture potential inhomogeneities which are present in the mass balance observations. Furthermore, it was investigated how basic summary statistics of mass balance simulations, such as the Coefficient of Variation (CV) as proxy for inter-annual variability, differ from observations. Inhomogeneities were not only determined by trend detection based on the modified Mann-Kendall, but also using the Pettitt and Wilcoxon-Rank Sum Test (WRS). The Null Hypothesis of both methods is that the time series contain no change.
While the Spearman Correlation suggests a satisfying to good agreement in the monotonic relationship between simulated and observed mass balance in all catchments, the models failed to represent the trend detection results of the observations. Based on the modified Mann-Kendall test that takes serial correlation into account, SWAT-GL was only in one catchment able to correctly classify the trend statistic. The Sen's slope estimator, used to represent trend magnitudes, differed especially for the outlier SCG model where the sign was mismatched. In the WG and LCG models the simulated trend component was underestimated, while it was overestimated for the WG. The CV, with generally large values mostly above 1, was underestimated in two as well as overestimated in two cases. While it was general in an acceptable range, the SCG again exhibited an outlier. The annual mass balance time series, which were trend-corrected in the presence of a trend, have been further tested on inhomogeneities based on the non-parametric Pettitt and Wilcoxon-Rank Sum Tests. In summary, the simulations agreed relatively well with the observations at the 0.05 significance level. SWAT-GL was not capturing the shift in median

detected in the observed time series of the WG (WRS) and further rejected the Null Hypothesis of the Pettitt Test in case of the SCG. However, it must be noted that the poor simulations of the SCG affect the meaningfulness of the test results.

Table 7.5: Summary of statistical results for the simulated and observed mass balance time series over the whole ismulation period. The summary table consists of the Spearman Correlation (ρ), the modified Mann-Kendall after Hamed and Ramachandra Rao (1998) considering autocorrelation (MMK_H), the Sen's Slope estimator, the Coefficient of Variation (CV) as well as the Pettitt and Wilcoxon-Rank Sum (WRS) Test.

Glacier	Mass Balance										
	ρ	MK_H		Sen's Slope		CV		Pettitt		WRS	
	-	Sim	Obs	Sim	Obs	Sim	Obs	Sim	Obs	Sim	Obs
Gulkana	0.67	0	1	-5.142e+4	-2.479e+5	1	1.30	0	0	0	1
Lemon Creek	0.62	0	1	-5.846e+4	-2.107e+5	0.88	1.21	0	0	0	0
South Cascade	0.79	1	0	3.281e+4	-1.113e+4	2.75	1.72	1	0	0	0
Wolverine	0.83	1	1	-5.311e+5	-3.931e+5	2.52	2.72	1	1	0	0

7.3.5 Cross-Validation of Discharge

Discharge in all catchments was cross-validated on the daily scale, assuming that a reasonable fit is achievable when glacier and snow-related processes are well represented in the heavily glaciated catchments. The performance, evaluated based on the KGE, can be found in Table 7.4. The temporal coverage of the validation periods of each catchment is found in Tab. 7.2.

First, we see that the GG and WG models show no difference in the KGE values of the respective calibration and validation periods. Second, a significant drop in quality for the SCG model (from calibration to validation) is found. Lastly, the LCG model acts as a strong outlier with KGE values <0.3. In contrast, the other three catchments almost entirely show KGE values >0.6, results often considered satisfactory in hydrological studies. Interestingly, the worst-performing glacier during the glacier-based optimization (SCG) exhibits the best overall discharge performance. However, its good quality in the first validation period stems mainly from an overestimated ablation that reduces the underestimation of available water for discharge compared to the other glaciers. Given the large glacier influence in the catchment, the good performance w.r.t. discharge is likely caused for wrong reasons. This stresses the fact that models that are evaluated only for discharge are questionable for comprehensive hydrological investigations. The other glaciers are characterized by a high PBIAS towards the observations (simulations have less water than observations). Besides, the second validation phase of the SCG model, in contrast to the other glaciers, does not cover the 2000s which are associated with even higher instationarities. Covering the 2000s would likely further reduce the performance metric. The large PBIAS is most significant in the LCG model, where streamflow is underestimated by 45% or 51% in the calibration and validation phase, respectively.

This behavior is emphasized when looking at Fig. 7.8, which illustrates mean annual discharge, together with two separate periods of simulated mean daily discharge (averaged discharge for each day of the year over the indicated periods). For the WG, LCG and GG models a distinct underestimation of annual flows is present in the simulations. When we center (correct the time series by its mean) the annual discharge time series (see Fig B.1), we find a good monotonic relationship for the WG model. For the SCG, the deviation of observed and simulated annual discharge increases over time. This further indicates that a temporal coverage of the 2000s of the SCG model would degrade its results. In general, simulated annual flows of SCG show a decreasing tendency with time, which could be caused by the strong recession

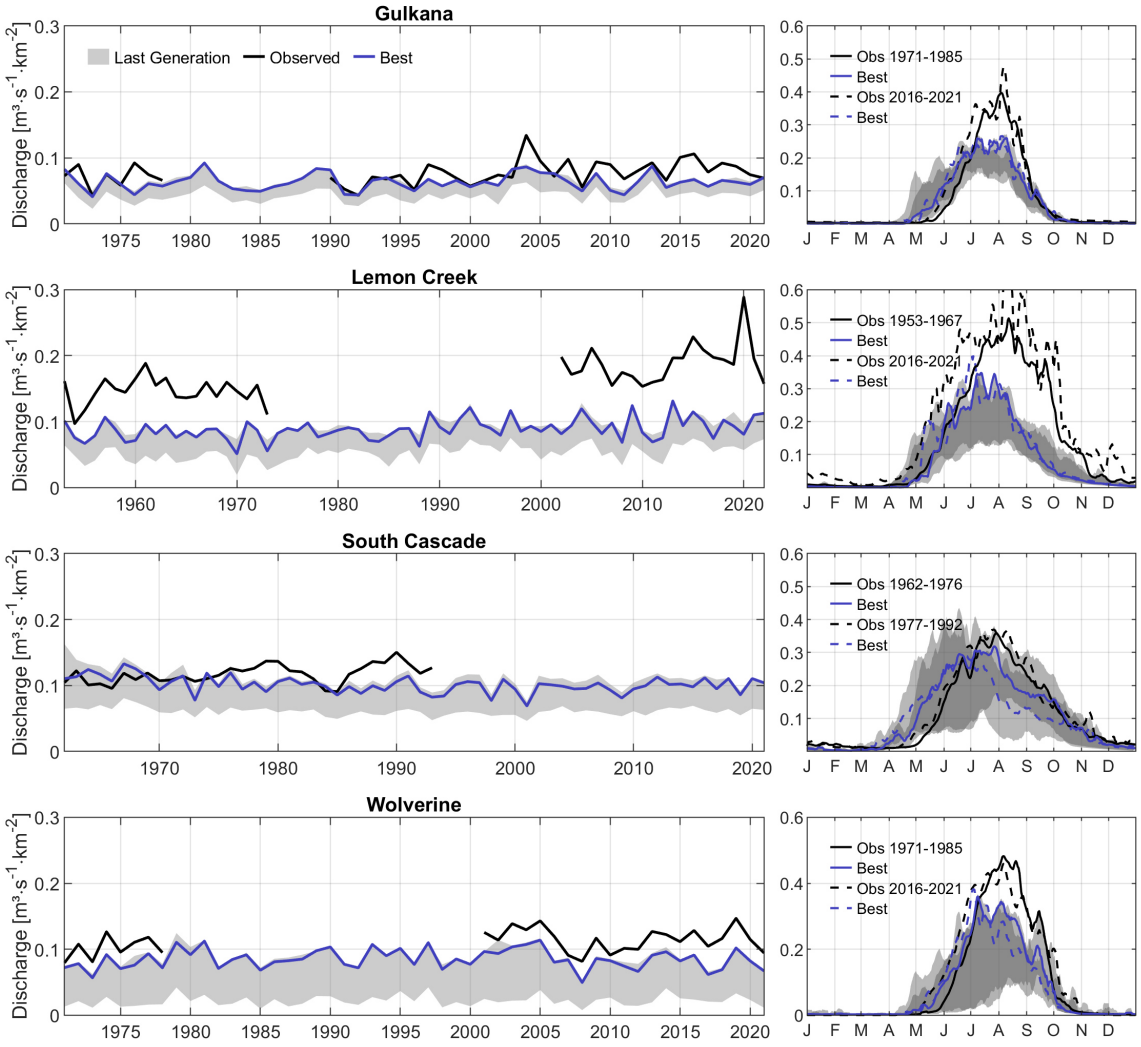

Figure 7.8: Simulation results for the cross-validation of discharge in all four catchments for mean annual flows (Column 1) and average flows of each day of the year (Column 2) at the end of the optimization. Mean daily flows over the year (from day to 366) are averaged for the earliest available slice of validation period 1 (solid black line represents observation; solid blue line the best simulation) (see Tab. 7.2) and the latest available slice of validation period 2 (dashed black line represents observation; dashed blue line the best simulation). Blue lines (simulation) cover the same period as indicated for the corresponding black lines.

initiated at an early stage of the simulation period as shown before (Sec. 7.3.3). This could then cause a pronounced underestimation of glacier melt contribution especially at end of the simulations as the actually contributing elevations disappeared already.

Evaluating mean daily discharge of the different periods further stresses the substantial undercatch of flow in the simulations (Fig. 7.8 all glaciers). The periods were chosen so that they lie from each other to highlight potential model deficiencies in the representation of nonstationarities. The LCG and GG models do not show any significant change in the amount of discharge between the early and late period. This points to relatively stable glacier conditions. In contrast, flows in the SCG basin essentially decrease in the later period, which is in line with the aforementioned results. A similar, albeit not as pronounced, pattern is found for the WG model. The model of the SCG shows the largest range of simulated flows

over the year (grey shadings). The share of simulations with a positive mass balance for the WG (see Section 7.3.3) is likely causing the very low simulation bound of discharge in the basin, with almost no flow until August.

7.3.6 Comparison of Single-Objective and Multi-Objective Optimization Results of Discharge and Mass Balance

To illustrate the capabilities of SWAT-GL in terms of discharge and mass balance simulations, two SOOs were conducted. One for each of the two variables. The SOO of discharge was based on KGE and that of mass balance on NRMSE to be consistent with the MOO counterpart. The test case was conducted for the WG only and should replicate a typical hydrological modelling case where people rather use discharge or mass balance only than multiple objectives. It should be emphasized here that we are not saying that the approach is desirable or good practice, but that despite known shortcomings of single-objective studies, the approach is still common practice. For the SOO of the mass balance the parameter choice and ranges are similar to Table 7.3. However, as we did not consider any lags for discharge in the MOO we added SWAT's SURLAG and CN2 parameter for the SOO of discharge to maximize SWAT-GL's capabilities in the representation of streamflow.

Fig. 7.9 illustrates the results. The first line of plots (a) and b), respectively) refers to the SOO model of discharge and the second to the one of MB (c) and d)). The first column shows the SOO results for the variable which was optimized (a) and c)) and the second column for the variable that was not used in the respective SOO (MB in b), Q in d)). The SOO results clearly demonstrate the sharp increase in KGE values when discharge is used directly as objective, compared to KGE values from the MOO. The former best KGE of 0.64 (Tab. 7.4) is substituted by a relatively high upper bound of 0.92. The results are reached after 40 generations already and lead to a median KGE shift from 0.5 to 0.9. Besides, the median KGE of the MOO represents the lower bound of the SOO. While the simulated mean annual flow of the MOO procedure showed a significant bias compared to the observed flow (see Fig. 7.8), the SOO results in PBIAS values of -0.78% (not shown) and is capable to bring reasonable amounts of water in the system. The SOO of discharge shows a substantial degradation in representing annual mass balance compared to the MOO (b)). The best solution of the SOO of discharge exhibits has a NRMSE value which is 0.37 above the MOO counterpart (or 78%).

When looking at annual mass balances, although we see a much better representation already after 20 generations, the "best" performing simulations smaller or equal to the 10% percentile are similar between MOO and SOO. This even goes so far that the minimum NRMSE of 0.51 achieved in the MOO, is not beaten by the SOO procedure after convergence. This indicates the appropriateness of the MOO using SWAT-GL in complex glaciated and snow-fed catchments. However, the MOO shows slightly poorer results for discharge compared to those resulting from the SOO for mass balance (d)). Above the 20th percentile the performance of the SOO has on average a KGE approximately 0.05 above the one from the MOO. Below the 20th percentile the performance of the MOO is largely degraded.

7.4 Discussion & Outlook

As a profound evaluation of SWAT-GL's performance in different glaciated catchments was missing so far the intention was to contribute to close this gap with our work. Note that further information especially regarding technical details of SWAT-GL and future plans about model improvements are found in Schaffhauser et al. (2024b).

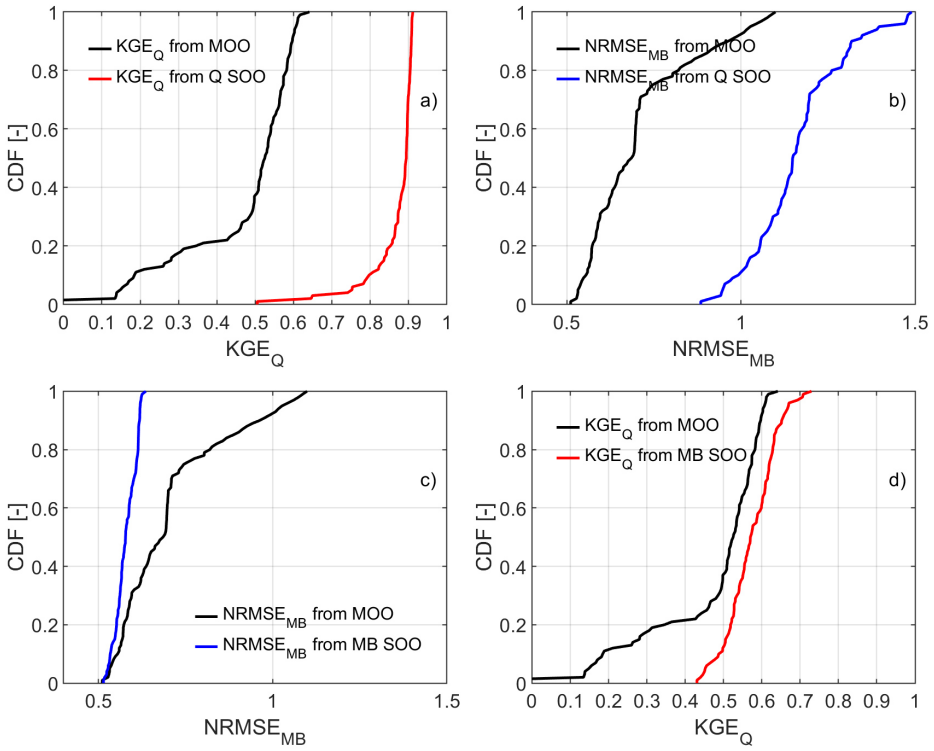

Figure 7.9: Cumulative distribution function (CDF) plot of the objective functions for the single-objective optimization (SOO) and multi-objective optimization (MOO) results of daily discharge and annual mass balance for the Wolverine Glacier. a) and b) compare the MOO results with the SOO model optimized for discharge. c) and d) compare the MOO results with the SOO model optimized for mass balance. a) and c) show the results for the optimized variable of the SOO, b) and d) show how the corresponding non-optimized variable (MB for b) and Q for d)) perform. The CDF of the MOO refer to the last generation (Generation = 100), while the selected SOO results refer to generations at a relatively early stage of the optimization close to the point where convergence is reached. In detail, the selected generation of the discharge optimized model refers to generation 40, and generation 20 for the mass balance optimized model. Black indicates the CDF from the MOO, red the one from the discharge SOO model and blue the mass balance SOO model. The subscripts Q and MB indicate discharge and mass balance, respectively. The sample size N of the CDFs is 100 which refers to the general sample size in the optimization. Results focus on the calibration period of the WG.

7.4.1 Glacier Parameterizations & Process Representation

We used the Method of Morris to identify (screen) and rank glacier and snow parameters in the four basins (Song et al., 2015; Pianosi et al., 2016; Sarrazin et al., 2016).

In general, it was shown that lapse rates together with parameters controlling the maximum degree-day factor for ice (and thus glacier melt) are among the most sensitive parameters in all catchments. The strong temperature-dependence inherent in SWAT-GL is further emphasized through the relevance of the threshold temperature that triggers glacier melt. An important role plays the SWE threshold that determines when a (sub-)basin is fully snow-covered. The parameter links the old snow routine and the newly integrated glacier routine, as glacier melt can only occur under snow-free conditions on the glacier. SWAT's snow cover received growing attention in multi-objective calibration studies that try to improve model consistency (Tuo et al., 2018; Grusson et al., 2015). The fraction of snow cover is directly affecting the amount of daily snow melt (lower fractions reduce the amount of snow melt) and indirectly glacier melt. As any degree of snow cover can be achieved with any SWE, there is also the risk to accomplish

good snow cover results with implausible amounts of snow. This circumstance is, contradictory to its importance, rarely discussed in the literature. SWE measurements are, although seldom available, of tremendous importance as they can be used for plausibility checks of snow amounts. Reanalysis or remote sensing-based SWE product resolutions as well as a relative large spread in product comparison studies are still an obstacle. Further, SWE would probably be more suitable to draw conclusions on precipitation inputs compared to snow cover only. We want to further emphasize that the intermediate sensitivity of precipitation lapse rates might be misleading. The objectives chosen might not be valid to allow for precipitation-related conclusions as none of the three variables is based on absolute volumes (of snow or ice). For example, a separate consideration of summer ablation and winter accumulation would provide a more realistic picture of system in- and outputs.

As SWAT-GL is still in its early stages, the SA was conducted for diagnostic purposes, involving screening and ranking among different catchments. This was done independently of the optimization purpose. In future applications the dimensions of the parameter space should be reduced accordingly. The SA for example suggested a reduction of the parameter space (by 6-8 dimensions) in the different catchments of the study. The derived glacier parameter sets after optimization, are relatively well-spread in the parameter space. However, in the demonstration catchments the maximum melt factor tends to group at its defined lower bound. This indicates a potential reduction of the lower bound for an even better representation of glacier mass balance. Physically our chosen values could be reduced, however, as SWAT-GL internally makes a plausibility check between the estimated snow and ice degree-day factors, a further reduction might make internal corrections of the degree-day factor more likely. Besides, further reducing the lower bound of the parameter might exacerbate the strong underestimation of flow. The high values of the glacier melt temperature imply that the models seem to compensate for other temperature-dependent processes as the model seems to try to delay glacier melt. This indicates that glaciers are, despite the good SC representation, snow free and exposed to melting too early. A lag factor similar to the temperature lag factor of snow already present in SWAT could also give further control in the timing of glacier melt. However, our study has shown that the snow lag factor is not very sensitive, although its lower bound was chosen in a way to avoid abrupt and extreme snow melt events. We further propose that alternative solutions concerning the lag of ice and snow melt might be explored and evaluated in order to decouple them more clearly from the lag factor related to effective precipitation.

Overall, the standard deviation of the Elementary Effects indicate that glacier and snow processes behave strongly non-linear and exhibit potential interacting effects which we see as a further indication of SWAT-GL's suitability. The moderate interaction ability for SNOCOVMX is considered to be unusual, as it links snow and glacier processes which would suggest higher interaction and/or non-linearity. Future work might put attention on Time-Varying Sensitivity Analysis (TVSA), such as DYNIA or also using EEs, to obtain further insights in parameter dominance at different scales and periods over time (Chiogna et al., 2024). Especially in the context of climate impact assessment, insights of a potential loss in model skill due to a reduction in the dominance of (historically working) parameter sets in non-stationary systems become crucial (Wagener, 2022).

7.4.2 SWAT-GL's Performance in Representing Glaciated Catchments

The optimization procedure using NSGA-II for snow cover, glacier hypsometry and glacier mass balance worked well for the highly glaciated catchments with the exception of the SCG. For the SCG an abrupt change in the mass balance estimates in the middle of the simulation period causes implausible results. However, snow cover estimates were very good in all catchments and we highly recommend to use snow

cover as an objective function for an adequate representation of mountain hydrology when using SWAT-GL. Especially due to the fact that MODIS (or other) snow cover data is relatively easy to access and readily available, which is not the case for measurements of glacier mass balance estimates. In data-scarce regions, predominantly the typical setting of high-mountainous areas, snow cover in combination with downstream measured discharge might often be the only sources of data for calibration and validation.
While annual net mass balance was well represented, it was noticed that glacier melt tends to start too early leading to an extended overlapping period where snow and glacier melt contribute equally to runoff generation. The mass balance estimates are better represented for the bigger glaciers or catchments respectively. However, SWAT-GL was introduced to provide a simple but efficient approach to represent glacier dynamics on multiple scales. It is assumed that most application will generally be beyond the scale of the example of the small SCG, which might also be one reason for the relatively bad performance. Moreover, small-scale processes such as snow redistribution are equalized on larger scales (e.g., mesoscale) and thus less dominant.
As hypsometry measurements were available, they were used in the optimization process. In future work total glacier area might be a suitable alternative to the individual hypsometry time series. As the Δh-parameterization assumes upper parts of a glacier to be more stable we conclude that the approach might fail to represent the dynamics of the upper elevation sections at the SCG and LCG while it seems more appropriate for the WG and GG (see Fig. B.2). Using total area changes could therefore improve the representation of the overall MOO, as it would circumvent the attempt to reproduce a pattern SWAT-GL is structurally not able to. Similarly, if individual hypsometry time series are used one might consider to put less weight on the upper parts of the glacier.
We have shown that, by using discharge as a single-objective, as done for the WG, the performance could be substantially ameliorated (KGE $>$0.9). Using mass balance in the SOO (again WG) we have shown that the best solutions of the MOO were comparable to the ones resulting from the SOO with respect to the achieved NRMSE values. The statistical results of the mass balance estimates significantly dropped using discharge for SOO and could not compete with the MOO mass balance results. In contrast, the discharge performance as a result from the mass balance SOO was better than the discharge representation of the MOO. Unlike mass balance, discharge was not part of the MOO objectives and partly constraints the interpretability. Nevertheless, the difficulty to achieve model consistency in highly glaciated and mountainous catchments became particularly visible in the LC catchment. The studied glaciers generally have a high contribution to streamflow, as for example found in O'Neel et al. (2014) (for GG and WG). Since we consistently overestimate ablation for the LCG, it is initially contradictory that we obtain streamflow underestimations of up to about 50% (ablation in SWAT-GL mainly refers to glacier melt). Potential reasons can be manifold. A wrong representation of snow amounts and distribution, despite a good snow cover fit, or a simple underestimation of liquid precipitation (or a combination of the two) might be potential reasons. Moreover, similarly to what is described in O'Neel et al. (2014) the model could also underestimate summer ablation and winter accumulation which govern the mass balance, which would again be related to precipitation. However, it seems that precipitation input might be too low since the LCG meteorological data stems from a remote valley station and the model relies on optimized lapse rates. Therefore future modelling work could also try to not only use net mass balance but (if available) make further use of seasonal mass balance (winter accumulation & summer ablation) derivations in the calibration strategy (Schaefli and Huss, 2011). The winter mass balances could also be used to additionally validate the precipitation inputs from the stations and to adjust the precipitation laps rates.
In general, it became evident that SWAT-GL has great capabilities to be applied in glaciated catchments,

also for longer, non-stationary time scales. It is assumed that the simple degree-day approach integrated in the mass balance module alone could cause significant improvements in glaciated catchments. The simplicity of the approach also leads to high transferability with manageable effort.

7.4.3 Glacier Initialization

A further very sensitive factor that affects the simulation results, in particular under long simulation periods which are likely subject to persistent climate change effects, is the initialization of the glacier mass. Datasets such as from Farinotti et al. (2019) or Millan et al. (2022) provide valuable information on glacier thickness and thus mass initialization. But attention should be paid when simulation periods start decades ago before the considered thickness estimates. A comparison of the mass differences between the Farinotti estimates and our own calculations based on GlabTop-2 for the earliest possible DEM and outline (see Section 7.2.5) reveal substantial differences. The magnitudes are between -11% to +19%. The contradictory signs are produced as the Randolph Glacier Inventory outlines for the different glaciers stem from different acquisition years which are sometimes earlier than the outlines used for GlabTop-2. This emphasizes the importance of the initialization assumptions. In greater detail, temperature conditions back then might not be suitable to trigger glacier melt in an appropriate magnitude as lower glacier bounds are simply located too high. Basically, there would be a mismatch between the link of glacier elevation and runoff generation. This becomes evident when examining areal losses as fraction of initial area of each glacier over the simulation period (see Appendix B.3), where fractional area losses range from more than 10% (WG) to more than 40% (SCG).
The Δh approach implemented in SWAT-GL does not consider glacier flow and does allow for glacier area growth (not be confused with accumulation in ice water equivalent of a specific ES) in its current version, which is relevant especially for long simulation periods with phases of growth.

7.5 Conclusions

The recently extended version of SWAT, called SWAT-GL, was tested in representing the hydrology of four highly glaciated basins. The new SWAT-GL, which makes use of a physically-based glacier evolution routine has proven to provide robust hydrological simulations of catchments that are characterized by nivo-glacial processes. It thus serves its purpose and adds a valuable contribution to the hydrological modelling community, and in particular, the SWAT community.
We have also identified traditional model consistency issues prevalent in hydrological modeling and demonstrated their significance, even when multiple glacier and snow processes are included in the calibration procedure. Although SWAT-GL substantially improves model consistency, such problems should receive more attention. While we could show SWAT-GL's applicability even under long transient conditions, constraints remain and require further efforts to address. This is particularly true for climate impact studies, where simulation periods can exceed 100 years. In such studies, we advocate a minimum requirement that assesses the suitability of model components for climate impact statements to avoid flawed conclusions. Transient conditions, for example, could significantly affect degree-day factors, making initial choices inappropriate. A topic that is rarely addressed and discussed.
We identified parameter clustering at the edges of the initial parameter ranges, which indicate solutions that could impair physical plausibility. Moreover, contradictory patterns in the representation of snow and glacier processes (and discharge) were found. For example, a good representation of snow and glacier processes partly resulted in an unsatisfying representation of streamflow. We demonstrated that an ad-

equate to good snow cover simulation does not necessarily lead to an accurate representation of glacier components. These basic insights, although partly recognized, go beyond SWAT-GL applications and are of general importance for the modeling community. The sensitivity analysis of SWAT-GL emerged a strong temperature-dependence of the model. This underpinned the importance and role of lapse rate parameterizations, also as a major source of uncertainty, in high mountain catchments. In addition, it was shown that the parameter space of glacier and snow-related parameters could be significantly reduced across all basins, suggesting potential applicability to other study areas.

Even though SWAT-GL was tested in catchments that are unprecedented in terms of data availability, the authors see no restrictions in its transferability to areas with poor data. Global datasets of ice thickness estimates and glacier outline set a suitable baseline to apply SWAT-GL. Although relatively small glaciated catchments were employed, the approach can be scaled up without imposing any substantial additional computational demand or physical limitation on the approach.

In conclusion, the most significant merit we discovered with SWAT-GL was its ability to adequately represent glacier processes in contrasting catchments. This encourages its further use in modeling glaciated and high-mountain catchments. However, there are technical limitations, including the requirement for introducing supplementary concepts to enhance the model's flexibility, along with structural limitations.

Chapter 8

SWAT-GL: The Way Forward

8.1 Demonstration of SWAT-GL in the Upper Naryn

The following chapter aims to complement the SWAT-GL efforts described in the previous chapters and therefore revisits the Naryn Basin in which the expanded SWAT version is further tested and evaluated (Chapter 8.1). In addition, the test is followed by a discussion on potential future directions of SWAT-GL (Chapter 8.2).

8.1.1 Background & Methodology

In addition to the extensive exploration of SWAT-GL presented in Chapter 6 and Chapter 7, SWAT-GL was further tested in the rather data-scarce and large Upper Naryn. The focus of SWAT-GL's expanded evaluation was on the approximately 10,000 km^2 headwater part up to the gauge of Naryn City. The investigation is based on the results presented in Chapter 4, in which clear limitations in the headwater catchment were identified. In addition, a literature review is provided that covers and presents all available hydrological modeling results found in the literature for the Upper Naryn Basin. The results of the hydrological modeling refer to the performance criteria found for different gauges in the literature. The idea is to see whether SWAT-GL improves the representation of discharge in a large glaciated catchment without any glaciological information. Thus, the evaluation iss focusing on streamflow only. Methodologically, a SWAT-GL model for the Upper Naryn was set up according to Chapter 4 (or Schaffhauser et al. (2023)). For homogeneity reasons, the reference dataset from ISIMIP2 was used to force the model, namely temperature and precipitation from EWEMBI to cover the 1970s and 1980s with relatively good discharge data observations in the region. However, the expanded demonstration of SWAT-GL focuses solely on discharge to additionally highlight potential model consistency problems when using discharge as a single objective in glaciated catchments with SWAT-GL. The, study focuses on the three available headwater gauges Big Naryn, Small Naryn and Naryn City. It should be noted that the model is optimized for the outlet of the Upper Naryn, namely Naryn City. The Big and Small Naryn are not explicitly considered in the optimization procedure, an approach that is widespread in the literature, despite potential limitations. The results for the Big and Small Naryn are provided supplementary to demonstrate the implications of the approach and to emphasize differences to 4.

Calibration and validation periods are identical as before (Chapter 4), using daily observations from 1974 - 1981 and 1982 - 1987 for calibration and validation, respectively. The SWAT-GL model was initialized with data from publicly available sources to demonstrate a typical workflow in data-scarce

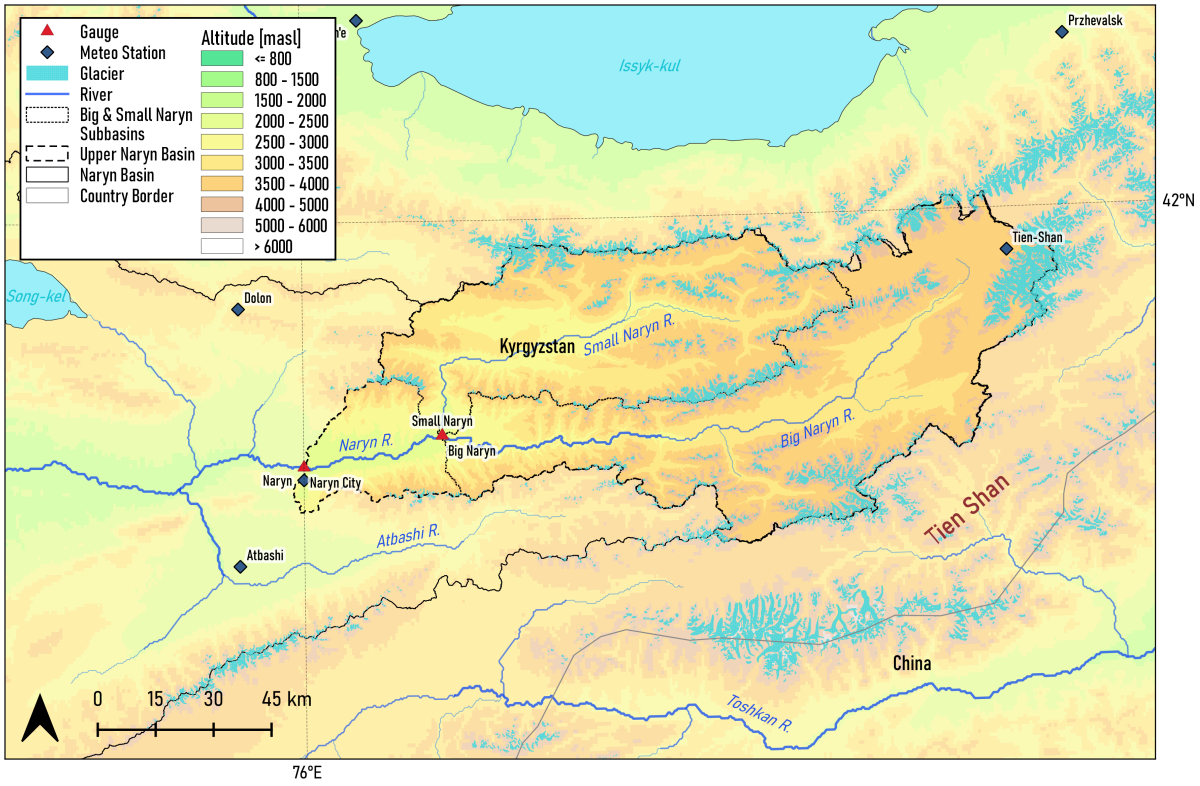

Figure 8.1: Illustration of the Upper Naryn. Note that the gauges of the Big and Small Naryn share nearly the same location whereby only one gauge symbol is shown.

catchments. Glacier outlines stem from the Randolph Glacier Inventory (RGI Consortium, 2017) and glacier thickness estimates from the of Farinotti et al. (2019). Four optimizers were tested and compared to identify potential differences in the final parameter sets and the representation of water balance components, in particular with respect to the glaciological component. The four optimizers used are three versions of Differential Evolution, namely DE/best/1, DE/rand/1 (Storn and Price, 1995) and JADE (Zhang and Sanderson, 2009), as well as an adapted Particle Swarm Optimization (PSO), called Pyramid PSO (Kennedy and Eberhart, 1995; Li et al., 2022). NSE was used as performance criteria during the optimization, but further performance metrics are presented alongside. Overall, a typical workflow and model application in a glaciated, data-scarce region is mimicked that solely relies on discharge and the basin outlet. It is aimed to illustrate the capabilities of SWAT-GL when used for such an application, however, to point to and discuss difficulties and limitations of such an approach in parallel to provide recommendations for future applications. The catchment is illustrated in Fig. 8.1. The glaciation in the Upper Naryn, according to data from the RGI, is around 7.5% and the minimum elevation of the catchment is around 2,000 m. The Upper Naryn covers the Big and Small Naryn catchments in which the two headwater streams originate. Annual precipitation ranges from 250 to 300 mm according to the 2 stations presented in Fig. 8.2. The rainy season shifts slightly from West to East with peaks in May, June in the Western basin and a more homogeneous peak in the East from May to August. Streamflow at the basin outlet (Naryn City) is subject to regime shifts, suggested by earlier onsets of the flood season, as well as increased peak flows during that period (Fig. 8.2 c)). The short investigation focuses on the three gauges Naryn City, Big Naryn and Small Naryn (see also Fig. 8.1).

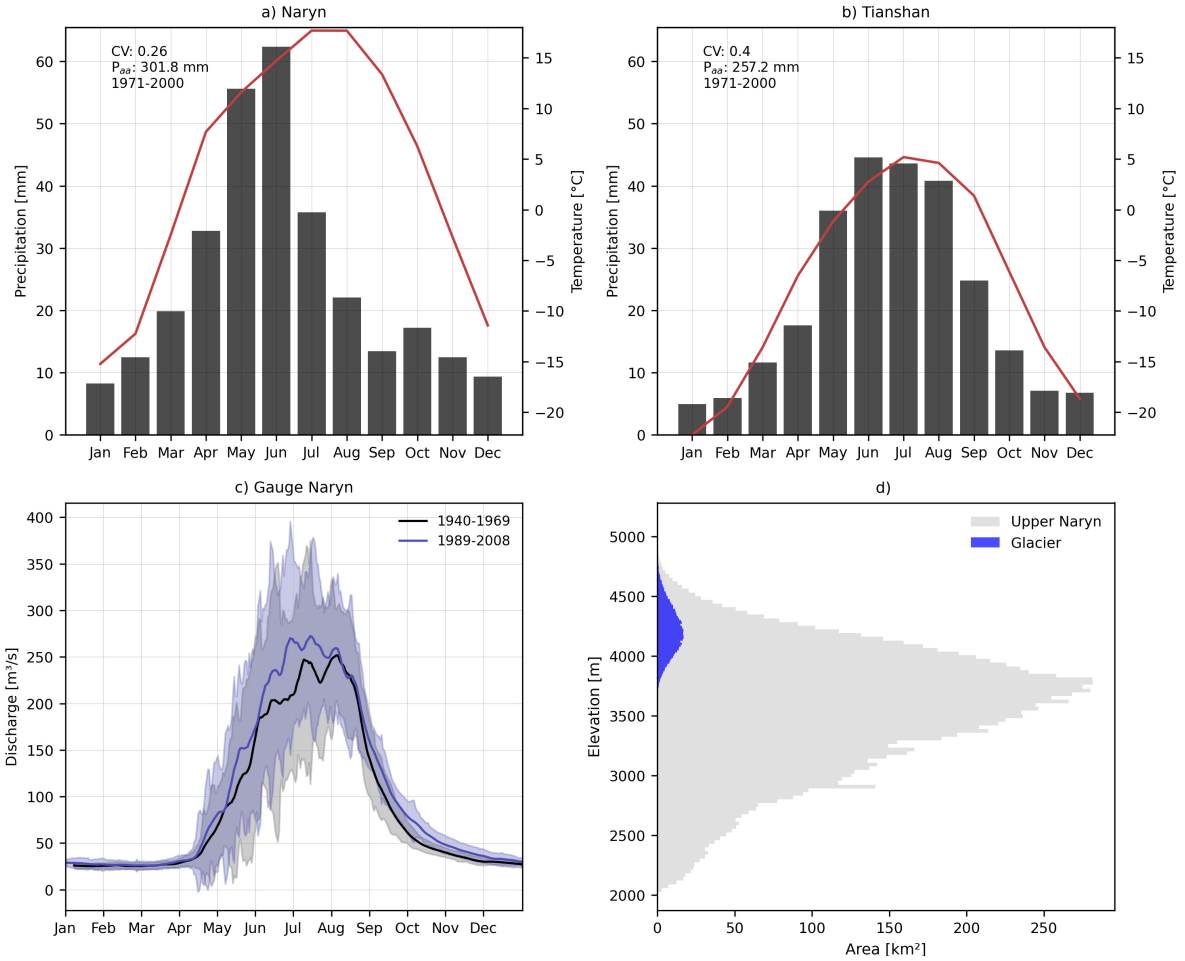

Figure 8.2: Overview of the climatology in the Upper Naryn for a) Naryn City and Tianshan b), the mean daily discharge for every day of the year during two different periods c) and the elevation distribution of the Upper Naryn (bin width around 31 m for the total basin elevation and around 16 m for the glacier elevation) and the glaciers inside the Upper Naryn d). CV and P_{aa} in the climatology plots refer to the Coefficient of Variation and the mean annual precipitation sum within the indicated period. The shaded area in c) indicates the range of discharge values for a specific day within the indicated periods.

Decomposition and Background of KGE & NSE

In the following, we will examine the evolution and current state of the two vastly used and discussed performance metrics NSE and KGE. Despite their great potential and meaningfulness, modifications to these metrics have not yet been widely adopted. The standard form of the NSE (Nash and Sutcliffe, 1970) and KGE (Gupta et al., 2009) are presented in Eq. 8.1 and 8.2. Eq. 8.1 also illustrates the relationship between MSE and NSE, as MSE-based metrics, such as NSE, are often referred to as MSE normalizations.

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}(O_i - S_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2} = 1 - \frac{\mathrm{MSE}}{\sigma_o^2} \tag{8.1}$$

$$\mathrm{KGE} = 1 - \sqrt{(r-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \tag{8.2}$$

where S is the simulated value, O is the observation, $\overline{O}$ represents the observed mean, MSE is the mean square error, r refers to the Pearson Correlation, α to the error in flow variability and β to the bias term

with:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(O_i - S_i)^2$$

$$r = \frac{\sum_{i=1}^{n}(O_i - \overline{O})(S_i - \overline{S})}{\sqrt{\sum_{i=1}^{n}(O_i - \overline{O})^2 \sum_{i=1}^{n}(S_i - \overline{S})^2}}$$

$$\alpha = \frac{\sigma_s}{\sigma_o}$$

$$\beta = \frac{\mu_s}{\mu_o}$$

where σ and μ refer to the standard deviation and mean, respectively.

General History of the NSE & KGE

The Kling-Gupta-Efficiency (Gupta et al., 2009), widely used in hydrological studies as one of the major criteria to assess model performance, was originally developed as an alternative metric to overcome limitations of the NSE and MSE. The latter two are based on squared error sums and the NSE uses the observation variance as baseline (that normalizes the MSE). Discussions on the limitations of the NSE have a long and enduring tradition and a variety of alternatives and improvements have developed over time. Concerns on the use of the NSE, refer for example to its overemphasis on high flows (due to its squared error type, also known as least square methods), the usage of the observation mean as benchmark (NSE = 0 corresponds to the observation mean) (Gupta et al., 2009; Gupta and Kling, 2011) or its sensitivity to outliers and time-offsets (McCuen et al., 2006). Problems also arise due to the normalization inherent in NSE, which makes it a relative metric, although NSE values are often treated (compared) in an absolute manner and therefore neglect individual hydrological conditions (Schaefli and Gupta, 2007). Accordingly, a direct comparison of NSE values is at least questionable and should be carried out with care. A strong seasonal cycle for example, is often associated with high NSE values and, conversely, a model must perform better in less dynamic catchments (Schaefli et al., 2005; Krause et al., 2005; McCuen et al., 2006). Additionally, Ehret and Zehe (2011) showed the sensitivity of least square measures to time shifts and Fowler et al. (2016, 2018) indicate constraints in the robustness of least square-derived parameter sets under changing climatic conditions (especially drier). The NSE formulation already dates back to 1970 (Nash and Sutcliffe, 1970) and the discussions on its limitations, interpretability and applications, which did not initially start with the development of the KGE but much earlier, still last until today (Garrick et al., 1978; Martinec and Rango, 1989; Węglarczyk, 1998; Legates and McCabe, 1999; Oudin et al., 2006; McCuen et al., 2006; Schaefli and Gupta, 2007; Jain and Sudheer, 2008; Ehret and Zehe, 2011; Willmott et al., 2011; Pushpalatha et al., 2012; Lin et al., 2017; Knoben et al., 2019; Schwemmle et al., 2021; Mathevet et al., 2024). NSE modifications or improvements try to resolve existing deficiencies of the original formulation and therefore usually focus on reducing effects from the squared error and those stemming from the baseline model. Besides, data transformations or penalty terms are often applied to overcome limitations. Lastly, multi-objective approaches are generally used to consider a diverse set of performance criteria and to provide more balanced hydrological simulations. An overview of NSE (and secondarily KGE) modifications are provided in Table 8.1.

Also the KGE, albeit less frequently, has been subject to modifications, which mainly focus on the α and β terms of Eq. 8.2. The most prominent one, is likely the modification of the variability ratio α proposed

by Kling et al. (2012), where α is replaced by γ:

$$\gamma = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} = \frac{\mathrm{CV_s}}{\mathrm{CV_o}}$$

where CV represents the coefficient of variation. The modification intends to remove cross-correlations between the bias and variability terms. It is worth to mention that KGE values derived from Eq. 8.3 provide the lower limit of the three individual terms β, r and γ (Kling et al., 2012).

Table 8.1: Overview of NSE and KGE modifications.

Performance Metric	Category	Description	Reference	Equation
Logarithmic NSE	Transformation	• Log-transformed variable • Emphasizes low flows	Oudin et al. (2006)	$\mathrm{logNSE} = 1 - \frac{\sum_{i=1}^{n}(ln(O_i)-ln(S_i))^2}{\sum_{i=1}^{n}(ln(O_i)-ln(\overline{O}))^2}$
Baseline-adjusted NSEs	Time-dependent benchmark models	• Using alternative benchmark model $\overline{O_i}$ (e.g. mean of calendar day, season, year etc.) • Reducing effects from default benchmark model ($\overline{O}$)	Murphy (1988); Martinec and Rango (1989); Schaefli and Gupta (2007)	$\mathrm{NSE'} = 1 - \frac{\sum_{i=1}^{n}(O_i-S_i)^2}{\sum_{i=1}^{n}(O_i-\overline{O_i})^2}$
Absolute NSE	Absolute errors	• Using absolute errors • Less emphasize on high flows	Legates and McCabe (1999)	$\mathrm{NSE_{abs}} = 1 - \frac{\sum_{i=1}^{n}\lvert O_i-S_i\rvert^j}{\sum_{i=1}^{n}\lvert O_i-\overline{O}\rvert^j}$
Baseline-adjusted absolute NSE	Absolute errors & time-dependent benchmark models	*See Absolute NSE & Baseline-adjusted NSEs*	Legates and McCabe (1999)	$\mathrm{NSE''} = 1 - \frac{\sum_{i=1}^{n}\lvert O_i-S_i\rvert}{\sum_{i=1}^{n}\lvert O_i-\overline{O_i}\rvert}$
Relative NSE	Transformation	• Uses relative variable values	Krause et al. (2005)	$\mathrm{NSE}_{rel} = 1 - \frac{\sum_{i=1}^{n}(\frac{O_i-S_i}{O_i})^2}{\sum_{i=1}^{n}(\frac{O_i-\overline{O}}{\overline{O}})^2}$
Split NSE	Subperiod-based computation	• Computes NSE for subperiod (e.g. each year) • Takes mean/median across all subperiods • More balanced weighting of dry and wet years	Fowler et al. (2018)	$\mathrm{sNSE} = \overline{\mathrm{NSE_t}}$
Square root NSE	Transformation	• Square root-transformed variable • Emphasizes low flows	Chiew and McMahon (1994); Oudin et al. (2006)	$\mathrm{NSE_{sqrt}} = 1 - \frac{\sum_{i=1}^{n}(\sqrt{O_i}-\sqrt{S_i})^2}{\sum_{i=1}^{n}(\sqrt{O_i}-\sqrt{\overline{O}})^2}$
5th-square root NSE	Transformation	• Fifth root-transformed variable • Emphasizes low flows	Chiew et al. (1993)	$\mathrm{NSE_{5sqrt}} = 1 - \frac{\sum_{i=1}^{n}(\sqrt[5]{O_i}-\sqrt[5]{S_i})^2}{\sum_{i=1}^{n}(\sqrt[5]{O_i}-\sqrt[5]{\overline{O}})^2}$
NSE w. Bias penalty	Penalized NSE	• Reduces high biases	Viney et al. (2009)	$\mathrm{NSE_{bias}} = \mathrm{NSE} - 5 \cdot \lvert ln(1+\mathrm{BIAS})\rvert^{2.5}$
Inverted NSE	Transformation	• Inverse-transformed variable • Emphasizes low flows	Le Moine (2008)	$\mathrm{NSE_{inv}} = 1 - \frac{\sum_{i=1}^{n}(\frac{1}{O_i}-\frac{1}{S_i})^2}{\sum_{i=1}^{n}(\frac{1}{O_i}-\frac{1}{\overline{O}})^2}$
Variability-adjusted KGE	α-replacement	• Used to avoid problems from mean values close to zero	Kling et al. (2012)	$\mathrm{KGE} = 1 - \sqrt{(r-1)^2+(\beta-1)^2+(\gamma-1)^2}$ with $\gamma = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} = \frac{\mathrm{CV_s}}{\mathrm{CV_o}}$
Bias-adjusted KGE	β-replacement	• Used to avoid cross-correlation of bias and variability	Tang et al. (2021)	$\mathrm{KGE} = 1 - \sqrt{(r-1)^2+(\alpha-1)^2+\beta^2}$ with $\beta = \frac{(\mu_s-\mu_o)}{\sigma_o}$
non-parametric KGE	β and α-replacement	• Avoids parametric assumptions on data	Pool et al. (2018)	$\mathrm{KGE} = 1 - \sqrt{(r-1)^2+(\alpha-1)^2+\beta^2}$ with $\rho = \frac{\sum_{i=1}^{n}(O_i-\overline{O})(S_i-\overline{S})}{\sqrt{\sum_{i=1}^{n}(O_i-\overline{O})^2}\sqrt{\sum_{i=1}^{n}(S_i-\overline{S})^2}}$ $\alpha_{NP} = 1 - \frac{1}{2}\sum_{k=1}^{n}\left(\frac{S(I(k))}{n\overline{S}} - \frac{O(J(k))}{n\overline{O}}\right)$
Split KGE	Sub-period-based computation	*See split NSE*	Fowler et al. (2018)	$\mathrm{sKGE} = \overline{\mathrm{KGE_t}}$

Decomposing the NSE & KGE

The development of the KGE in 2009 illustrates how the NSE can be decomposed in different ways, which served as the base for the construction of the related KGE. The different terms inherent in the KGE formulation are ideally suited for model diagnostic purposes and to identify potential deficiencies. Although the papers from Gupta et al. (2009) and Kling et al. (2012) are highly cited, the decomposition of the KGE into its constituents for model diagnostics is rarely done. This is also true for the second form of the KGE, originally provided in Gupta et al. (2009) and formulated as:

$$\mathrm{KGE} = 1 - \sqrt{s_r(r-1)^2 + s_\alpha(\alpha-1)^2 + s_\beta(\beta-1)^2} \tag{8.3}$$

in which s_r,s_α and s_β refer to weighting factors to put emphasis on specific components of the model. However, they are seldom applied in hydrological studies.
The original NSE decomposition (or MSE-based criteria, see Murphy (1988); Węglarczyk (1998)), from which the KGE was derived, divides the criterion in the Pearson Correlation, a conditional bias as well as an unconditional bias. Mathematically the decomposition can be written as:

$$\mathrm{NSE} = A - B - C \tag{8.4}$$

where:
$A = r^2$
$B = (r - (\frac{\sigma_s}{\sigma_o}))^2$
$C = \left(\frac{(\mu_s - \mu_o)}{\sigma_o}\right)^2$
with A measuring the linear relationship of simulation and observation, B represents the conditional bias and C the unconditional bias. An important fact to consider when using multiple evaluation criteria is their potential interdependence, which can cause flawed conclusions if disregarded, as shown by Węglarczyk (1998). From Eq. 8.4 for example, the relationship of the correlation coefficient and NSE becomes apparent (as the relationship between NSE and MSE was already explained). Węglarczyk (1998) illustrates this with respect to the integral square error (more widely known as the relative root mean square error) and the special correlation coefficient, which are both based on the MSE, however, using them in combination violates their independence. Gupta et al. (2009) then provided an alternative NSE decomposition:

$$\mathrm{NSE} = 2 \cdot \alpha \cdot r - \alpha^2 - {\beta_n}^2 \tag{8.5}$$

where:
$\alpha = \frac{\sigma_s}{\sigma_o}$
$\beta_n = \frac{(\mu_s - \mu_o)}{\sigma_o}$
with α representing a relative variability term (similar to Eq. 8.2) and β_n being a normalized bias term using the observed standard deviation (similar to the square root of C in Eq. 8.4). In detail, the terms focus on representing the temporal behavior of the observations and on reproducing the first two moments (mean and standard deviation) of the observed variable. It is desired to reach unity for r and α and zero for β. Accordingly, the NSE already encompasses different aspects when used for model optimization. Main concerns raised in the original work of Gupta et al. (2009) (similar to the literature provided above), were given due to the: 1) bias (β_n) normalization using the observed standard deviation, as well as 2) the double-accounting of α. The latter somewhat relates to the interdependence problem when using multiple objective functions as discussed in Węglarczyk (1998). α interacts with the linear relationship

r between observation and simulation (Eq. 8.5). It is demonstrated that the NSE favors solutions that underestimate simulated variability (by taking the first derivative of Eq. 8.5 with respect to α, resulting in $\alpha = r$ as optimal solution, and as r never reaches unity). The β_n normalization on the other hand, leads to an undervaluation of the bias term when observed variability is high. Using $\beta_n = 0$ and $\alpha = r$ in Eq. 8.5, results in NSE = R^2 as a potential maximum (given both other terms reach their optimum). The example given in Fig. 1 of Gupta et al. (2009) demonstrates that NSE values can be lower when $\alpha = 1$ (ideal) for a fixed value of r (assuming $\beta = 0$), compared to the case when $\alpha = r$ (optimal). For example, NSE would be 0.4 if $\alpha = 1$ and $r = 0.7$ (assuming $\beta = 0$), however, it would be 0.49 when $\alpha = r = 0.7$. The problems also apply to the MSE. Besides, various combinations of the aforementioned terms can result in similar NSE values and therefore constrain comparability even within the same model. It should be noted that the underestimation of flow variability and high flows using the NSE seems to contradict the general opinion that NSE optimization overemphasizes high flows. Gupta et al. (2009), however, underlines that these two characteristics are disconnected.
For the actual decomposition, Gupta et al. (2009) used the MSE (Eq. 8.6), which was derived by substituting Eq. 8.5 into Eq. 8.1:

$$\text{MSE} = 2 \cdot \sigma_s \cdot \sigma_o \cdot (1 - r) + (\sigma_s - \sigma_o)^2 + (\mu_s - \mu_o)^2 \tag{8.6}$$

The fractions of the individual terms can be computed as:

$$F_i = \frac{f_i}{\sum_{j=1}^{3} f_j} \tag{8.7}$$

and we can define the individual three contributing terms as:
$f_1 = f_r = 2 \cdot \sigma_s \cdot \sigma_o \cdot (1 - r)$
$f_2 = f_a = (\sigma_s - \sigma_o)^2$
$f_3 = f_b = (\mu_s - \mu_o)^2$
where the subscripts indicate indicate the involved metric, namely the variability, bias and temporal dynamic terms as defined before. The KGE intends to overcome aforementioned limitations while maintaining the multi-objective character of the NSE, as well as its individual components that focus on: temporal representation, variability and bias of the model. In fact, the KGE focuses on the bias term β, which is based on observed and simulated means so that all terms reach their optimum at unity, while it preserves the variability term α. As the KGE is based on an Euclidean Distance (as apparent in Eq. 8.2), best solutions have the shortest distance to the ideal point (unity for all three terms). In other words, the KGE optimization aims to find the closest point at the Pareto front to the ideal point. Again, r provides the maximum value of KGE that can be achieved provided the variability and bias terms reach unity. Gupta et al. (2009) also discussed limitations attached to the KGE, which applies for any objective function used. The general underestimation of simulated variability and thus for example runoff peaks associated with the NSE is not fully solved by introducing the KGE, albeit the impact is less pronounced. A characteristic that is also valid for extreme low flows, yet, the tendency of KGE (and NSE) is to overestimate these periods. Moreover, optimized models tend to be dominated by the correlation term and the temporal dynamics accordingly (also the case for the NSE).
The actual KGE decomposition used by Gupta et al. (2009) is based on the three terms of the Euclidean Distance and can be written as:

$$\text{EC} = \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \tag{8.8}$$

where the relative contributions of the three terms can expressed as:

$$G_i = \frac{g_i}{\sum_{j=1}^{3} g_j} \tag{8.9}$$

and the three terms as:
$g_1 = g_r = (r-1)^2$
$g_2 = g_a = (\alpha-1)^2$
$g_3 = g_b = (\beta-1)^2$
where the subscripts follow the same principle as in Eq. 8.7.
We use the KGE and NSE decomposition suggested in Eq. 8.7 and 8.9 for model diagnostic purposes in the Upper Naryn. In detail, the relative contributions of the individual G and F terms to the total sum of all three terms are examined. The idea is to identify potential shortcomings prevalent in SWAT and SWAT-GL. The experiment also investigates to which degree the correlation coefficient dominates the optimized models and how runoff peaks and flow variability are represented, given the assumption that optimizing for NSE typically emphasizes high flows. We follow the approach of Gupta et al. (2009) to highlight differences that become apparent from the decomposition.

8.1.2 Results & Discussion

General Model Performance

In Schaffhauser et al. (2023) the quality of representing the Upper Naryn (with the gauges Naryn City, Big Naryn, Small Naryn) was diverse and indicated limitations. SWAT-GL is expected to significantly improve the large PBIAS in the Big Naryn for example, which was mainly due to an underestimation of discharge in the Summer period that is associated with a significant glacier melt contribution. The results of the four different optimizers are overall comparable, however, a considerable drop between the calibration and validation is noticed for some performance criteria. Statistical results are presented in Table 8.2 and 8.3 for all three gauges. Although the four optimizers used NSE as objective functions, the remaining performance metrics are supplementary provided for completeness. The split NSE (sNSE) (Table 8.3) is the average of the individual NSE for each year. The sNSE gives a more balanced weight to dry and wet years.
Focusing on Naryn City, the DE/rand/1 optimizer outperforms the other algorithms, which is true not only for the NSE but for most statistical criteria. While in the calibration phase the optimizers are extremely close to each other, the superiority gets particularly visible in the validation period (all criteria). A small exception is the logarithmic NSE where the DE/best outperformed the other models, indicating a better baseflow representation here.Notable is the drop between calibration and validation in the PBIAS present in all optimized SWAT-GL models, except the DE/rand/1. For the PPSO, JADE and DE/best the difference between calibration and validation is >15% up to approximately 20%. In addition, all models experience a sign change in the PBIAS from a positive to a negative PBIAS, suggesting that the model is underestimating discharge in the calibration phase and then significantly overestimating flows during validation. This is interesting because the two other gauges (Big and Small Naryn) are strongly characterized by a lack of water in the model (PBIAS >20%) in the solutions of all optimizers. Besides, it is evident that for the outlet all optimizers produce very satisfying results with some reduced quality for the PPSO. However, surprisingly the global parameter sets of the Upper Naryn do not work to produce any baseflow in the Big Naryn subbasin (lNSE <-1 for all optimizers and all evaluation periods). A clear sign of the negative effects of a missing regionalization and the attached problems of an outlet-only

calibration, despite the very good results for the outlet. It is also interesting to see that the sNSE reflects the balanced high flow and low flow representation of the DE/rand/1 model, as it has high sNSE values in both the calibration and validation periods.

NSE & KGE Decomposition

In the following, results are presented for the NSE (MSE) and KGE decomposition of all models. The decomposition is used for model diagnostic purposes of the four optimized SWAT-GL models as well as the SWAT model from Schaffhauser et al. (2023). It is examined whether dominant terms within the performance metrics can be determined (with respect to the individual components) and to which extent the problems illustrated in Gupta et al. (2009) are present in the Upper Naryn models. Problems hereby refer to the points described in Section 8.1.1. Results are provided in Fig. 8.3 and 8.4 for Naryn City and the Big Naryn, respectively. Results are first discussed for Naryn City and then followed by those for the Big Naryn.

From Fig. 8.3 substantial differences in the contributions of the individual components of NSE and KGE

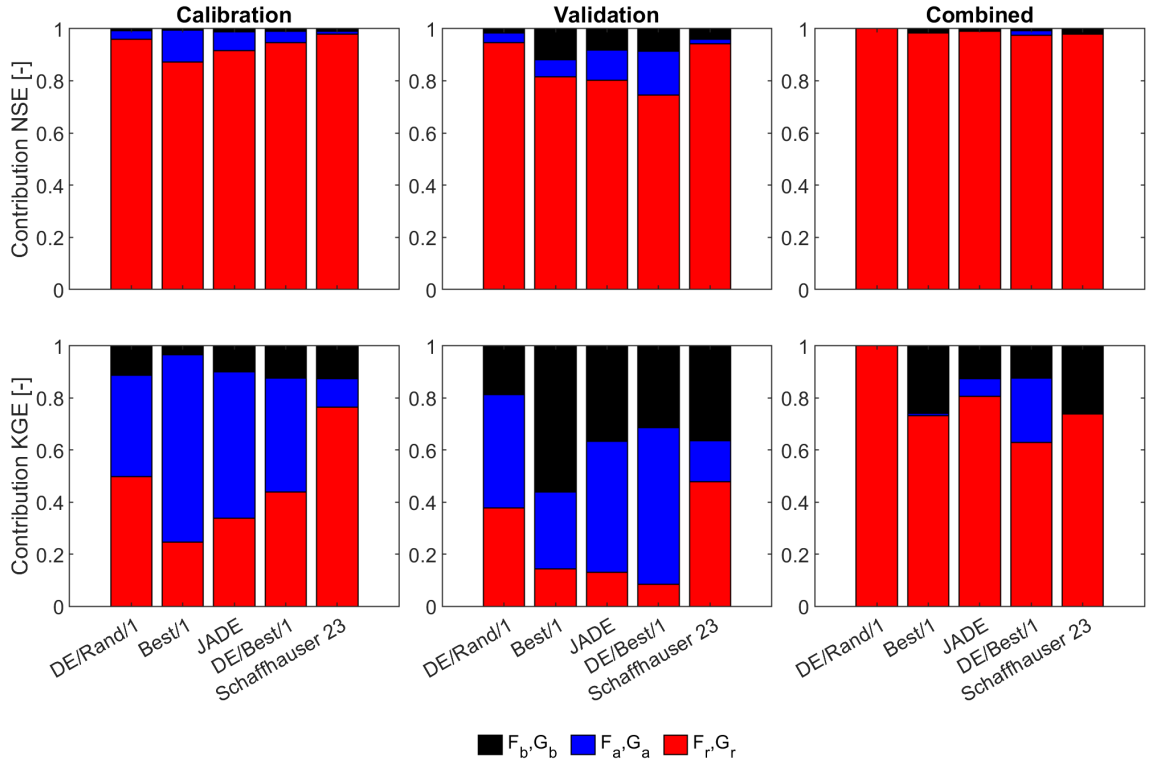

Figure 8.3: Relative contribution of the individual performance metric components to the overall NSE (MSE) and KGE at Naryn City. High contribution refers to strong negative impact on the performance. Results refer the calibration, validation and the total evaluation period (combined column that consists of the calibration and validation phase). The subscripts b, a and r indicate the source of the components, namely, bias, variability and temporal behavior. Terms denoted with F refer to the NSE and terms denoted with G to the KGE.

become apparent. A high contribution of a term indicates either that the component is far away from its optimum, it generally has a high weight or the performance criterion is sensitive to this component. Low contributions, on the other hand, can indicate values far from the optimum of the component.

Table 8.2: Calibration and validation results for the Upper Naryn based on NSE and KGE. The SWAT column refers to the simulations from Schaffhauser et al. (2023) and is based on SWAT. The remaining columns refer to the different optimizers used in the new analysis. For each column two values are provided (separated by a vertical bar |), the first value refers to the calibration (1974-1981) and the second to the validation period (1982-1987).

Gauge	NSE					KGE				
	DE/rand	DE/best	JADE	PPSO	SWAT	DE/rand	DE/best	JADE	PPSO	SWAT
Naryn City	0.84 | 0.83	0.84 | 0.76	0.84 | 0.75	0.84 | 0.75	0.74 | 0.72	0.89 | 0.88	0.83 | 0.77	0.86 | 0.76	0.88 | 0.74	0.84 | 0.82
Big Naryn	0.76 | 0.79	0.72 | 0.76	0.71 | 0.73	0.70 | 0.74	0.59 | 0.58	0.77 | 0.82	0.67 | 0.71	0.66 | 0.70	0.67 | 0.72	0.73 | 0.77
Small Naryn	0.74 | 0.72	0.70 | 0.70	0.70 | 0.70	0.73 | 0.72	0.78 | 0.68	0.62 | 0.65	0.56 | 0.62	0.56 | 0.62	0.60 | 0.66	0.84 | 0.81

Table 8.3: Same as Table 8.2 but for PBIAS, logNSE (lNSE) and split NSE (sNSE)

Gauge	PBIAS					lNSE					sNSE				
	DE/rand	DE/best	JADE	PPSO	SWAT	DE/rand	DE/best	JADE	PPSO	SWAT	DE/rand	DE/best	JADE	PPSO	SWAT
Naryn City	3.80 | -5.30	3.10 | -17.2	4.60 | -14.6	4.40 | -14.8	-5.5 | -10.8	0.84 | 0.83	0.91 | 0.86	0.85 | 0.81	0.78 | 0.73	0.70 | 0.70	0.84 | 0.82	0.84 | 0.76	0.83 | 0.74	0.83 | 0.75	0.69 | 0.69
Big Naryn	21.2 | 16.0	29.2 | 24.2	30.9 | 27.3	29.8 | 24.8	19.0 | 11.3	-1.1 | -0.43	-1.3 | -0.3	-3.0 | -1.5	-3.0 | -1.4	0.00 | 0.20	0.72 | 0.77	0.69 | 0.77	0.67 | 0.72	0.67 | 0.73	0.55 | 0.53
Small Naryn	26.4 | 22.0	29.8 | 23.2	29.5 | 24.1	29.0 | 24.0	5.40 | -7.6	0.74 | 0.73	0.73 | 0.73	0.75 | 0.71	0.61 | 0.54	0.61 | 0.49	0.75 | 0.74	0.71 | 0.71	0.71 | 0.72	0.73 | 0.73	0.77 | 0.54

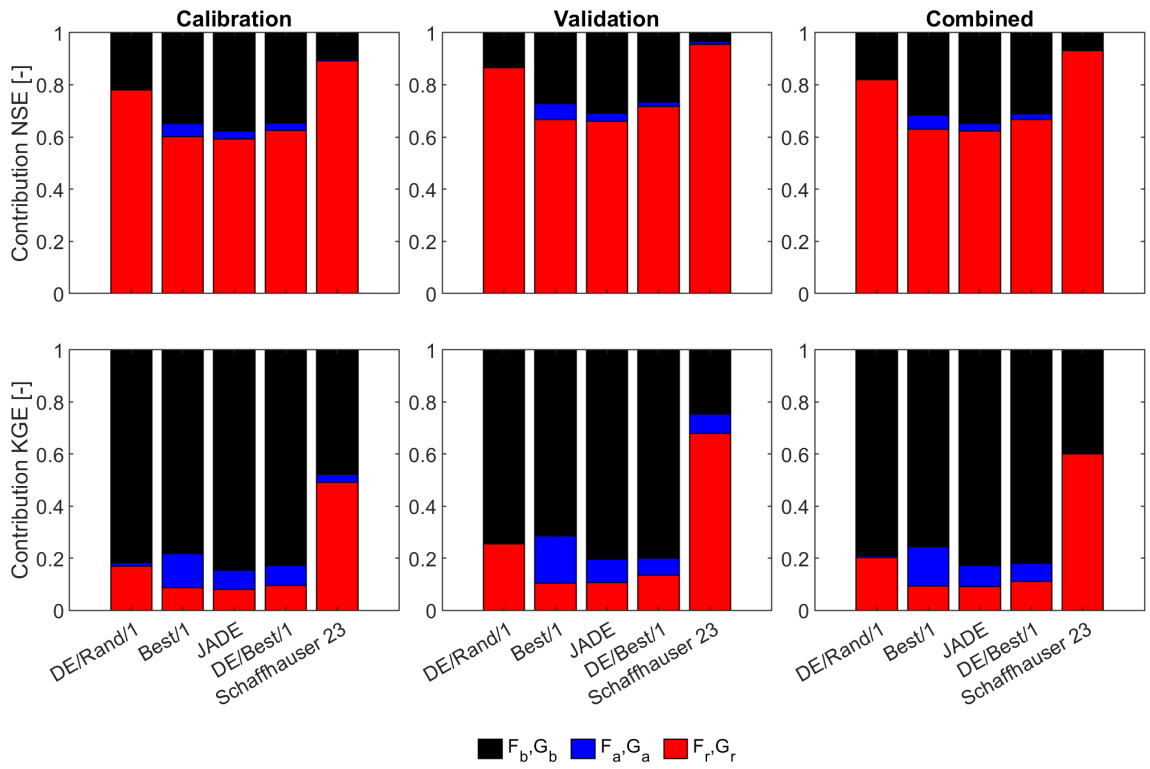

Figure 8.4: Same as Fig. 8.3 but for the Big Naryn.

For the NSE at Naryn City, it is evident that the temporal dynamics based on the Pearson correlation (expressed in F_r) seems to be the critical component mainly responsible for the gap to the optimal NSE value (true for calibration, validation and the combined period). Particularly in the calibration period, the results suggest that the variability (F_a) and bias terms (F_b) are relatively close to their optimums. In the validation period the influence of F_r drops slightly and the summed bias and variability can contribute up to 20%. In general, the study from 2023 and all optimized SWAT-GL models do not show significant differences in the individual contributions of the NSE. There is agreement that the models lack some temporal dynamics. For the KGE at Gauge Naryn, a much more balance contribution of the individual components is revealed (in particular for calibration and validation phase). Across all five models, G_r is still the most dominant term in the old SWAT model from (Schaffhauser et al., 2023). In contrast, the four optimized SWAT-GL models show variability deficiencies in the calibration and also bias shortcomings in the validation period. The KGE decomposition therefore contradicts the results from the NSE decomposition, where the bias and variability terms suggested a close to perfect representation. Vice versa, the information content for model diagnostics seems higher in the KGE. Looking at the total evaluation period for the KGE, the correlation-based term is becoming increasingly important. A pattern, which seems contra-intuitive at first, as one might expect a reflection that somewhat matches the average contributions of the calibration and validation period. However, this is not necessarily true, as we will show for the example of the DE/Rand/1-optimized SWAT-GL model in the following. While in the calibration and validation phase all three terms are of importance, the total evaluation period indicates G_b and G_a close to unity and thus an overwhelming contribution stemming from G_r. All individual components, as well as the standard deviations and means, are shown in Tab. 8.4. Results show that the validation period is wetter (w.r.t. observed mean discharge) than the calibration phase with a stable

Table 8.4: KGE Decomposition Example for SWAT-GL Optimization based on DE/Rand/1 at Naryn City (see also Fig. 8.3). Equations for β and α can be found in Eq. 8.2. μ is the mean discharge, σ the standard deviation and the subscripts s and o indicate simulation or observation, respectively.

Period	r (-)	α (-)	β (-)	σ_s (m^3/s)	σ_o (m^3/s)	μ_s (m^3/s)	μ_o (m^3/s)
Calibration	0.9194	0.9288	0.9616	81.8697	88.1421	80.2695	83.4769
Validation	0.9251	1.0807	1.0528	95.9739	88.8067	92.2135	87.5920
Total	0.9187	0.9993	1.0017	88.3791	88.4423	85.3877	85.2403

variability over all periods (w.r.t. observed standard deviation). However, simulated discharge tends to underestimate in the calibration and overestimate in the validation period, expressed as $\beta_{cal} < 1$ and $\beta_{val} > 1$. Similarly, $\alpha_{cal} < 1$ and $\alpha_{val} > 1$ of the model suggest lower dynamics in the model in the calibration and larger dynamics in the validation period. Again, the model first over- then underestimates the flow dynamics, which is reflected in the significant contribution of α and β in the KGE decomposition, as both terms are far from unity during calibration and validation. In contrast, taking the whole modeled period for evaluation purposes, the two signals average each other out and results for both terms are close to their ideal values. This points to several things, 1) it highlights the importance of a differentiated view of the evaluation periods as contrasting patterns can be present, 2) also the KGE decomposition can be subject to misleading conclusions and overestimated positive conclusions (w.r.t. α and β in our case), and 3) splitting the evaluation period in further parts can give valuable insights, as averaging out effects could hide potential deficiencies in the bias and variability terms. The last point underscores the value of sub-period-based performance metrics, such as the split KGE (Tab. 8.1). Not to mention the fact that contrasting combinations can results in similar values for α and β, and therefore different conclusions. If only the total evaluation period would be used for model diagnostics, there wouldn't be an indication that drier periods are "too" dry and wetter periods "too" wet in the model, but a well captured bias and variability would be inferred. Finally, we can see that in the NSE optimized model, α is very close to r in the calibration, confirming the discussion in Gupta et al. (2009), namely that the variability term is conditioned towards r rather than unity.

Results for the Big Naryn, shown in Fig. 8.4, are provided for comparability reasons. First, we see similar patterns as for Naryn City with a F_r dominance for NSE that is not present in the KGE decomposition (expressed as G_r). Furthermore, the KGE decomposition implies a relatively good representation of the temporal behavior of all five models, while there seems to be a significant bias (G_b) in all models and time periods, which contrasts the NSE results. However, in detail it becomes apparent that the bias is much stronger in the four SWAT-GL models, while the old model from (Schaffhauser et al., 2023) is characterized by a balanced contribution of G_r and G_b. Interestingly, the dominance of F_r that was revealed for Naryn City is weaker for the Big Naryn, where we can see a significant contribution of the bias term even in the NSE decomposition. Nevertheless, the major contributor remains F_r. Generally, the results highlight the notable differences, which can be present across different gauges even on a small scale. It also demonstrates the high information content that can be extracted by decomposing performance criteria in their constituents. However, the notable differences between KGE and NSE underline the caution that has to be paid when using NSE as performance metric and the decomposition for diagnostics. Besides, users should be aware that KGE and NSE are already representing multi-objective criteria aggregated in one metric.

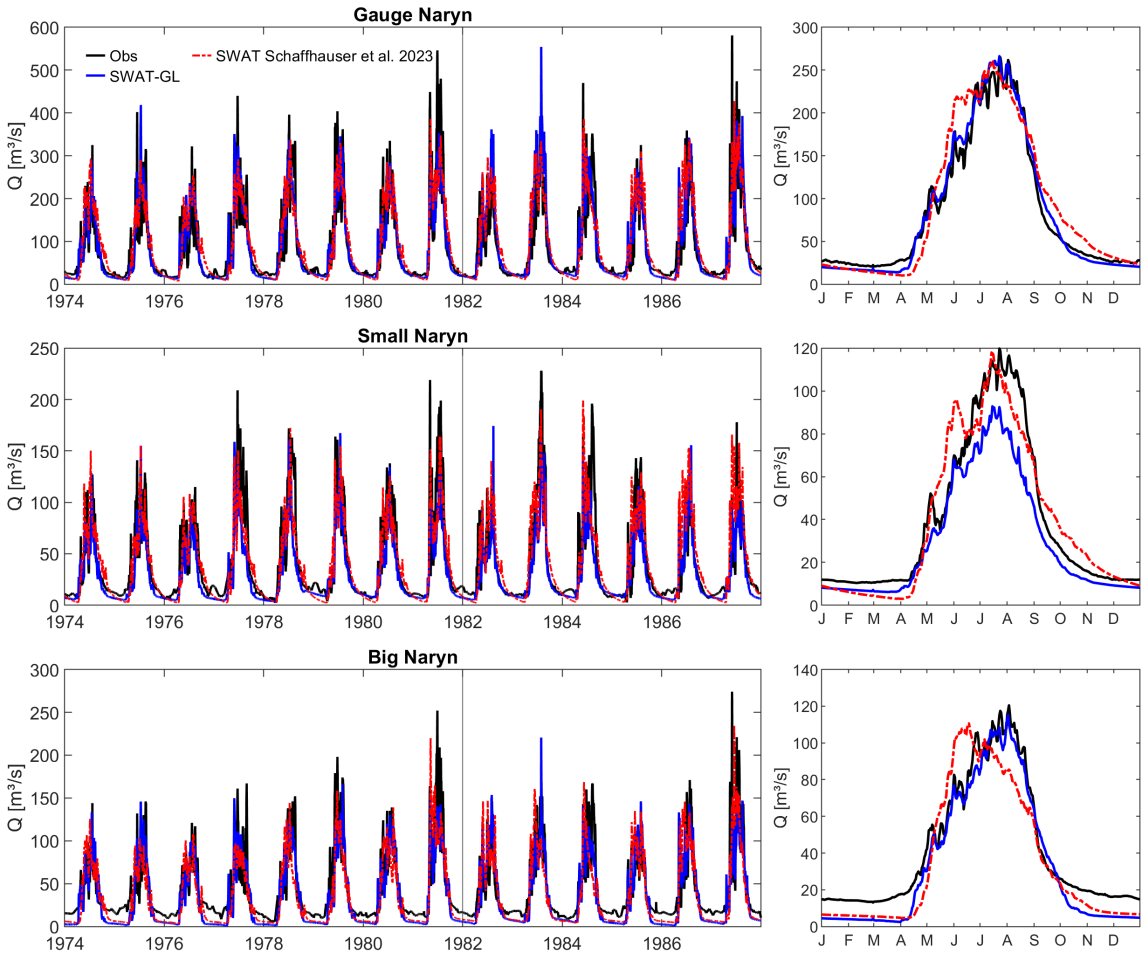

Figure 8.5: Hydrographs of observed flows (black), the old SWAT model (red) from Schaffhauser et al. (2023) and SWAT-GL (blue) using DE/rand/1 for all three gauges. Note: SWAT-GL optimization was performed only for Gauge Naryn, while the SWAT model from 2023 was regionalized. The right column shows the mean flows for every day of the year over the total calibration and validation period. The vertical black line in the left column of plots separates calibration and validation period.

SWAT-GL Optimization versus SWAT from Schaffhauser et al. (2023)

When comparing the SWAT-GL results of the four optimizers with the results of Schaffhauser et al. (2023) in Table 8.2 and 8.3 (indicated as *SWAT* column), especially the DE/rand algorithm is highly superior to the old model with respect to the NSE (both calibration and validation). The remaining algorithms also reveal better NSE's particularly in the calibration period. For the non-regionalized Small Naryn, the old SWAT model shows the highest NSE values in the calibration period, while in the validation period SWAT and all four optimizers are comparable. The good performance in the old SWAT model is likely attributable to the fact that high precipitation lapse rates compensated the missing glacier component. Interestingly, the KGE values of the old SWAT model at Naryn City are the second best in the validation period. However, in the calibration period all SWAT-GL optimized model version perform better, albeit all five model show KGE values >0.8. The poor representation of the Big Naryn in terms of NSE has significantly improved in all SWAT-GL models, although no regionalized optimization was performed. KGE values are surprisingly good for the Big Naryn and even better for the Small Naryn in the old SWAT model, despite a strong underestimation of flow at the Big Naryn. Particularly for the Small

Naryn, the old SWAT model shows superior results with respect to the KGE than the four SWAT-GL versions. This is likely associated with the lower PBIAS values compared to the SWAT-GL models that is somewhat reflected in the KGE through the β term, which is based on the simulated and observed mean flows. However, low flows, indicated by the lNSE are poorer in the old SWAT compared to the four optimizers at Naryn City as well as the Small Naryn. The overall more balanced representation of the SWAT-GL models becomes apparent in the sNSE values. The old SWAT model only reached sNSE values >0.7 in the calibration period of the Small Naryn, otherwise the model is consistently worse. Overall, the four SWAT-GL models show superior performance compared to the old SWAT model used in Schaffhauser et al. (2023), particularly for the optimized outlet, but in many cases also for the not-regionalized Small and Big Naryn. The promising performance of SWAT-GL is exemplified in Figure 8.5 in which we show the hydrographs for all three gauges from Schaffhauser et al. (2023), the observations and the best SWAT-GL model optimized with DE/rand/1. The hydrographs shows nicely how the whole year is well represented at Gauge Naryn with SWAT-GL. While in the old SWAT model recession started too early, was too slow and flow dynamics from July onward were not well represented the implemented glacier routine improves these components. It is also clearly visible that the Big Naryn pattern of the flood seasons is represented very well as well, especially compared to the old SWAT in which the flood season was initiated too early in the model and the glacier melt dominated period had a significant underestimation of flow. It is assumed that a regionalization of the Big Naryn with respect to baseflow would likely increase the already satisfying quality even further.
It should be noted that a discussion with local experts raised doubts about the quality of the Small Naryn discharge data. Accordingly, (future) optimization procedures should aim for trade-off solutions of the Big Naryn and Naryn City, which seem more reliable. Quality concerns can be further supported by the fact that realistic flow simulation at Naryn City coincide more with a realistic representation of either the Big Naryn or the Small Naryn, but less with a reasonable representation of both of them.

Glacier Representation & Impact on Performance Differences

From the section before the question arises of how much the glacier routine actually contributed to the improved performance or whether other factors played a more significant role. Therefore, we analyzed the split NSE and split PBIAS as functions of the observed mean annual flow and simulated glacier melt contribution to discharge. The investigation is presented for the DE/rand/1 optimized SWAT-GL model. Additionally, a close up of year 1974 is provided which represents a very dry year (in terms of discharge, ∼-20% below average flow in calibration and validation period).
Results are shown in Fig. 8.6. It becomes apparent that the old SWAT model is largely overestimating observed flows, while the SWAT-GL model represents the temporal dynamics and flow magnitudes quite well. The SWAT model also shows pronounced issues in the second half of the year (falling limb) with unrealistically high flows and slow recession. Besides, the glacier melt peak in July is only matched in the SWAT-GL model. The overestimation of dry years gets even more visible when looking at the split PBIAS versus observed annual mean flows, where a remarkable overestimation of the lowest annual flows is evident (<-20%). In contrast, the SWAT-GL counterpart has a relatively balanced representation of PBIAS values throughout the time series. Moreover, the higher mean annual flows are far better captured in the SWAT model than the dry years. This is also reflected in the similar plot, but for split NSE rather than split PBIAS. Here, we can a similar pattern where drier years are associate with much worse NSE values (partly <0.5). Again, SWAT-GL has an equally weighted performance between dry and wet years. The strength of SWAT-GL is further emphasized when the split NSE is compared with the simulated

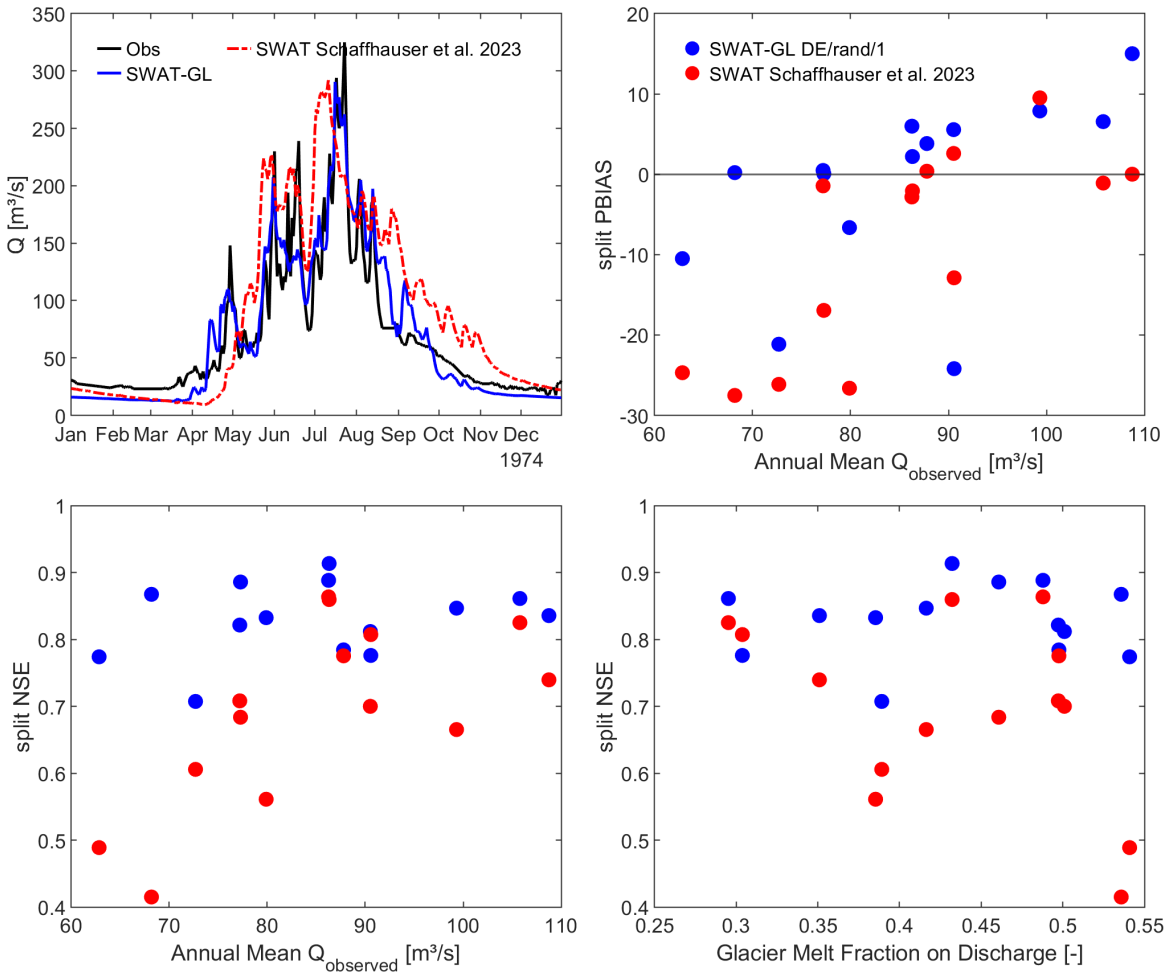

Figure 8.6: Zoom in of the very dry year 1974 (upper left), scatter plot of observed mean annual flow versus the individual PBIAS values of each year (split PBIAS) (upper right), scatter plot of observed mean annual flow versus the individual NSE values of each year (split NSE) (lower left) simulated annual glacier melt fraction on discharge versus split NSE (lower right). All plots refer to Naryn City (basin outlet).

glacier melt fraction of each year according to the DE/rand/1 optimized SWAT-GL model. Here it is indicated that years with a relatively high glacier melt contribution based on SWAT-GL, have generally lower NSE values in SWAT. In fact, it becomes clear that especially the dry years are those in which the glacier melt contribution is likely higher (in a relative way) also highlighting the importance of glaciers during water-scarce years in terms of precipitation. Moreover, the reason for the bad PBIAS in exactly those year in SWAT is likely the precipitation lapse rate that is probably set too high and is feeding discharge. In other words, streamflow receives too much rainfall in those years.

On the other hand, it is crucial to emphasize the clear limitations of the optimized models, which are based solely on discharge. First, the glacier melt fractions of the individual years might be on average too high, which is somewhat supported by the values provided by Saks et al. (2022) for the Big and Small Naryn. The generally high melt contribution is also reflected in the absolute glacier melt values, accompanied by a potential underestimation of accumulation, which leads to a strong negative net mass balance rate (see Fig. 8.7). In the figure it can be noticed that the elevation-dependent pattern of glacier

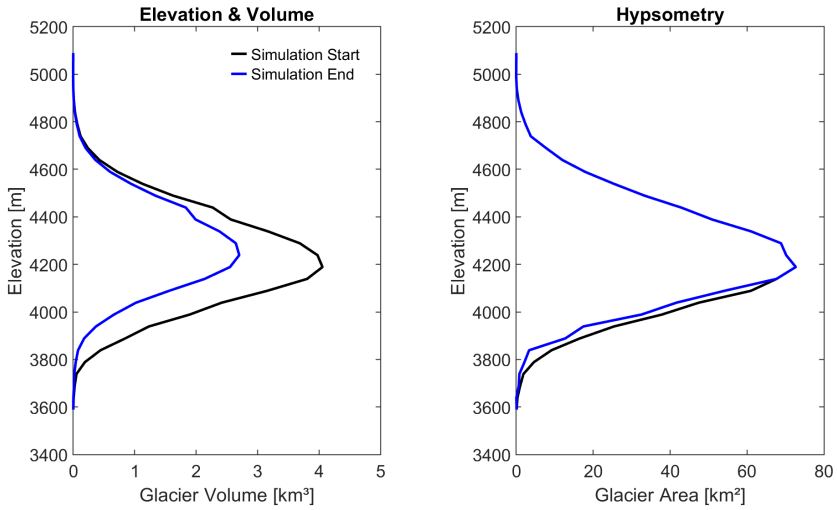

Figure 8.7: Glacier volume and area change over the simulation period per elevation section.

retreat looks realistic, yet, the glacier volume change is likely too high. The total simulated glacier volume in the basin is reduced by ∼45%, a value which is unrealistic for a period of only 14 years. A typical issue in hydro-glaciological modeling, when discharge is the only objective and relatively short time series are used. The chosen calibration and validation period demonstrated the need of relatively long evaluation periods in order to evaluate if the glacier response and balance is well captured. In fact, an evaluation period of 10-15 years seems to be of limited use in highly sensitive alpine catchments. In detail, the parameterization (of all optimizers) favors solutions that produce too much glacier melt and are probably characterized by underestimated precipitation in higher elevation zones. This ultimately leads to an undercatch of accumulation which would balance the net mass balance. In addition, the model would not be suited for climate impact assessment, regardless of its good discharge representation and thus again emphasizing the problem of single-objective optimization based on discharge only.

SWAT-GL Comparison With Other Studies

The Naryn Basin (and specifically the Upper Naryn) is regularly investigated and hydrologically modeled. The modeling results from this study and (Schaffhauser et al., 2023) are therefore put in the context of existing literature, to further evaluate how well SWAT-GL performs in comparison. The comparison serves for demonstration purposes and aims to indicate how accurately different gauges in the Naryn Basin are represented by various hydrological models. Results focus on Naryn City, which showed the best coverage among hydrological modeling studies.

Apart from comparing different model structures and calibration methods, the approach has the following limitations: first, different calibration and validation periods are used between the studies; second, different time steps are used among the studies and daily data was only used in Schaffhauser et al. (2023) and this study, while all other studies used monthly discharge data only; third, different performance criteria are used and provided restricting direct comparisons between the studies; fourth, different forcing datasets are used in the different studies to drive the models. In fact, since no other study used daily data, we decided to generate monthly flow time series from daily data and calculate all performance criteria additionally on a monthly base for comparison purposes. For example, the actual optimization at Naryn City of this this study and (Schaffhauser et al., 2023), was based on daily discharge and the

simulation results were upscaled to monthly time series in to calculate the performance metrics. However, own results are shown for both, monthly and daily simulations. Results are provided in Figure 8.8 for Naryn City with respect to the NSE and PBIAS that were available in most of the studies. The studies used for comparison are Shannon et al. (2023); Gan et al. (2015); Huang et al. (2022). Models used in these studies are SWAT (with an integrated Volume-Area Scaling for glacier mass balance estimates) (Gan et al., 2015), VIC (Huang et al., 2022) and DECIPHeR (Shannon et al., 2023). The results of all studies are in a comparable range, with slightly deteriorated results for the two daily models compared to the monthly models. However, the daily SWAT-GL models even outperformed some of the monthly models (with respect to NSE). The monthly SWAT-GL model in the calibration period also showed the overall best NSE among all compared models. With respect to PBIAS, especially the old SWAT models performed slightly worse than the other models. All in all, SWAT-GL is able to demonstrate a great potential in the Upper Naryn.

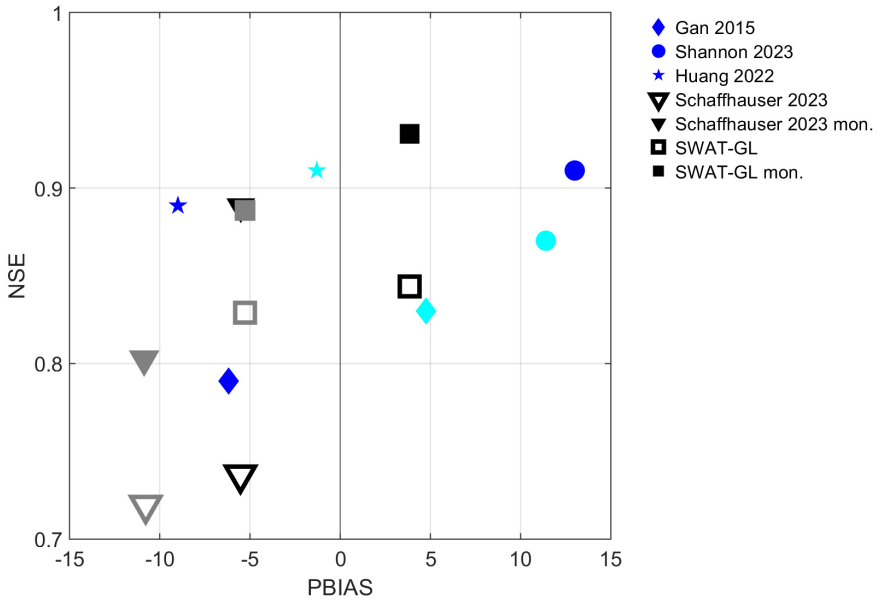

Figure 8.8: Comparison of SWAT-GL DE/rand/1 with other studies at Naryn City. Blue refers calibration results within studies conducted by other authors than the ones of this dissertation, cyan refers to the validation counterpart of this studies, Black and grey belong to the calibration and validation results from Schaffhauser et al. (2023) and SWAT-GL DE/rand/1, the non-filled results are the daily calibration and validation results from before, while the filled symbols represent monthly results. The SWAT-GL and SWAT2012 results were resampled to monthly values to calculate monthly performance criteria for comparability.

8.2 SWAT-GL Current Status & the Way Forward

A detailed overview of SWAT-GL can be obtained from the corresponding GitLab repository (`https://gitlab.com/lshm1/swat-g`), which is frequently updated in the course of changes. Although it was generally well-shown during this work (Chapters 4, 6, 7) that SWAT-GL provides great capabilities to simulate the hydrology of glaciated high mountain environments, future versions will hopefully contain further advances for the community. Besides, the goal is still to initiate the community to use SWAT-GL as a starting point and to incorporate own modifications and developments to foster a community-driven

development and exchange. The SWAT-GL publications presented in this dissertation have already generated a lively demand for applications within the community.

SWAT-GL modifications, at this day, can be assigned to either the snow or the glacier routine. An overview of all modifications is provided in Table 8.5. Many modifications focus on extending the existing snow routine of SWAT and to provide alternative approaches to the users. Among the implemented concepts are for example the widely acknowledged concepts, such as the temperature index approach from Hock (1999) or Pellicciotti et al. (2005) that integrate different radiation terms (and the corresponding radiation coefficients for calibration). Moreover, users can consider mixed precipitation to account for solid and liquid precipitation fractions simultaneously during precipitation events (Magnusson et al., 2014). With respect to rainfall, users also have the opportunity to consider rain on snow events, where they can define a precipitation threshold above which additional melt is allowed to occur due to an energy input from rainfall. If the threshold is reached, melt is increased based on an additional user-defined wet-day degree day factor. Magnusson et al. (2014) also proposed an exponential temperature index approach, which is introduced in SWAT-GL. However, while the glacier routine has been comprehensively evaluated as shown in this dissertation, the expansion of the snow concepts are not verified and validated yet and will need further time. Future steps involve several modifications for both, the snow and the

Table 8.5: Overview of concepts implemented in SWAT-GL. A detailed and regularly updated documentation can be accessed via the SWAT-GL GitLab page: `https://gitlab.com/lshm1/swat-g/-/blob/v3/swat_664_glacier_martell_clean/Documentation/SWAT-GL_Documentation.md?ref_type=heads`. The table contains novelties available in SWAT-GL and which are not available in the standard version of SWAT.

Concept	Description	Reference
Glacier		
Δh-parameterization	Glacier evolution based on Δh-parameterization	Huss et al. (2008, 2010)
Melt based on seasonal-varying degree day factor	Glacier melt module similar to SWAT's standard snow melt calculation	
Snow		
Rain on snow correction	Adjusts snow melt factor during rain on snow events to account for extra energy available for melt. User define precipitation threshold and wet degree day factor.	Gyawali and Bárdossy (2022)
Temperature index model with potential direct solar radiation	Alternative snow melt model extending temperature index approach for potential direct solar radiation. There is a snow and ice melt factor and a radiation factor. Approach is often referred to as HTI (Hock Temperature Index).	Hock (1999)
Enhanced temperature index model based on shortwave radiation	Alternative snow melt model extending temperature index approach for shortwave radiation. There is a snow and ice melt factor and a shortwave radiation factor. Approach is often referred to as ETI (Enhanced Temperature Index).	Pellicciotti et al. (2005)
Exponential temperature index model	Alternative snow melt model calculating a seasonally varying degree day factor as in standard SWAT, but snow melt is then further dependent on an exponential adjustment of a user-defined threshold temperature and smoothness factor.	Magnusson et al. (2014)
Mixed precipitation	Users can chose whether mixed (liquid/solid) precipitation can occur. Mixed precipitation occurs when temperature is between two user-defined thresholds. The solid fraction is calculated based on an exponential function that uses the temperature relative to the two user-defined temperature thresholds when precipitation is fully liquid or solid.	Magnusson et al. (2014)
Snow redistribution	In elevation bands snow towers are avoided by redistributing snow to lower elevated layers when a threshold value of snow water equivalent is reached.	Duethmann et al. (2015)

glacier component of SWAT-GL. With respect to the glacier representation, it is planned to additionally incorporate the widely-used V-A (Volume-Area) scaling (Bahr et al., 1997, 2015), which provides additional option for the evolution of the glacier. However, it has to be evaluated whether the current spatial representation of glaciers in SWAT-GL is suitable for an appropriate V-A modeling. One major aim for future developments is to stronger incorporate people from the glaciology field to receive valuable feedback, which considers latest developments in the glaciological community that could be directly translated into further improvements of SWAT-GL. Furthermore, it is planned to implement alternatives for sublimation and accumulation which are incorporated in a simple manner so far. Another component for example, which lacks an appropriate representation not only in SWAT-GL but is generally challenging in glacio-hydrological modeling is the adequate representation of debris cover and the two associated effects of insulation (decreases melt) and albedo decrease (increases melt). Especially the scale for which SWAT-GL is intended, will likely limit the appropriate representation of small scale elements. Another feature that is difficult to handle and likely beyond the scope of SWAT-GL, albeit often asked for, is the generation of glacier lakes when topography allows. One of the next steps will focus on the implementation of glacier advance, so that elevation sections could get glaciated again after they became ice-free. With respect to modification of SWAT-GL's snow routine, it is planned to first perform a comprehensive technical verification followed by a validation that ensures real-world alignment of the already implemented concepts presented before. Subsequently, features such as a full energy-balance routine might be considered. Besides, the currently implemented alternative approaches for snow melt computations for example, can only be parameterized globally (basin-scale) rather than spatially-distributed.

It should be noted that all advancements of SWAT-GL will consider the trade-off between complexity and the original usage and purpose of SWAT. It is not intended to add details that for example lead to a significant additional computational demand or do not match the semi-distributed, physically-based idea of SWAT. However, the most important and prioritized step is the transfer from SWAT-GL to SWAT-GL. Due to the structural changes of SWAT+, SWAT-GL might not be transferred in the same way as it is implemented so far. There are also other changes within SWAT+, such as only globally available lapse rates or surface runoff lags that are currently unsuitable for SWAT-GL and would lead to major changes in SWAT+, making an implementation even more difficult. Nevertheless, a SWAT+ version of SWAT-GL is needed and will be developed soon.

Chapter 9

Conclusions and Outlook

9.1 Concluding Remarks

The overall goal of the underlying dissertation is:

„Advancing the representation of high-mountain hydrology and demonstrating the implications of evolving climate models on climate impact assessment“

from which four research questions were derived and addressed:

i) *How does the evolution of climate model generations affect the impact assessment of the regional hydrological cycle?*

ii) *To which extent do generations of climate models contribute to the total uncertainty in hydrological projections?*

iii) *Does the introduction of a glacier-expanded version, called SWAT-GL, enhance the applicability and credibility of SWAT in high mountain environments*

iv) *How robust is SWAT-GL in contrasting, heavily glaciated basins, and what is the sensitivity of glacier parameters, including interactions with the existing snow routine?*

The presented work leads to the following conclusions:

Conclusion 1: Although the general direction may be consistent, climate sensitivity affects the magnitude of regional hydrological changes across climate model generations

The investigations demonstrated both, the concordance between hydrological impacts caused by different generations of climate models, as well as the substantial differences in their magnitudes that might exist. In dry alpine catchments, which are strongly affected by nivo-glacial processes, small temperature variations, as they are visible in the projections of different phases of climate models, can lead to amplified signals in the hydrological response. Temperature differences are due, in particular, to different climate sensitivities reflected in climate models as a result of advances across climate model generations. It has to be emphasized that despite changes in the extent of hydrological impacts, there is a lot of consistency in the directions of projected hydrological changes. Rather than focusing exclusively on distinctions, tremendous importance lies in emphasizing similarities projected among models within a generation

and across different generations of climate models. This approach facilitates the creation of robust and confident statements about regional implications of the hydrological cycle. The clear distinction of differences is particularly important in communication of uncertainties inherent to climate change studies. This was demonstrated for a large high mountain catchment using the two most recent phases of climate projections that are widely in use. The translation of different climate sensitivities into the catchment response is of high importance for a better understanding of the hydrological system and its sensitivity to climate variability.

Conclusion 2: The generation of climate models can significantly contribute to total uncertainty

The sources of uncertainty throughout the modeling chain up to the final impact statement are undeniably diverse and substantial. It was shown that the generation of climate models can be an essential source of uncertainty in hydrological projections. Particularly with regard to the influence it has on the interactions with different factors. In addition, the study revealed that the methodology offers insights into the development of climate models within the same family, but across different generations.
Nevertheless, other factors contribute equally or similarly to the overall uncertainty, and results can vary depending on the hydrological variable being investigated and the projected period. It is also noteworthy that further contributions to the projected impacts of climate change can easily arise from simple decisions, such as the choice of calibration dataset, which need to be explored at a minimum.

Conclusion 3: Advancements in process representation, like SWAT-GL, can significantly enhance the robustness and applicability of SWAT and other hydrological models

Model extensions and adaptations are essential to further advance hydrological modeling, even in the age of AI. These not only improve the general applicability, but also make an essential contribution to increasing the credibility of the modeling. For SWAT it was shown that the introduction of a glacier routine substantially affects SWAT's robustness and usability in glaciated catchments. Moreover, technical improvements such as SWAT-GL help to avoid misuse and to create a sound basis for climate impact studies. To enable a versatile and flexible applicability not only of SWAT, but of hydrological models in general, it is crucial to continue and foster research on process representation under different scales and conditions. Here, the importance of reproducibility and accessibility, to name just two of the pillars of the FAIR principles, is often underestimated. Open software code, as in the case of SWAT-GL, promotes a community-driven approach of model development and favors the comprehensibility of technical innovations. While SWAT-GL introduces a completely novel process and spatial unit, it is assumed that there is also large potential for already existing processes.

Conclusion 4: SWAT-GL offers great opportunities for diverse glaciated catchments and illustrates the sensitivity of process interactions within these environments

It is demonstrated that SWAT-GL not only successfully complements the model structure of SWAT technically, but also provides the necessary model improvement in glaciated catchments. SWAT-GL was able to reflect the glacio-hydrological complexities of four different glaciated catchments. The diversity

refers to the scale and topography of the catchments as well as to the prevailing hydro-meteorological conditions. Besides, the shown sensitivity of process interactions, such as those involving snow and glaciers, underlines the importance of representing them adequately. It's worth emphasizing that the extensive availability of multi-decadal data enabled to demonstrate SWAT-GL's capability in accurately depicting snow and glacier processes under transient climatic conditions. Conditions which often pose one of the greatest challenges in hydrological modeling. Conversely, this indicates the importance of having sufficiently long model evaluation periods under transient conditions to ensure an adequate process representation that supports robust climate impact studies. Nevertheless, it was also noted that there is still considerable scope to improve model consistency of SWAT-GL.

Conclusion 5: Novelties introduce new challenges

Despite any improvement in the general model consistency, it was shown that even in catchments driven exclusively by snow and glacier dynamics, there can be a considerable disagreement in the representation of streamflow. Both in terms of time and scope. This underscores once again the importance of adopting holistic views on the hydrological cycle. Although having a good representation of snow and glacier processes, achieving realistic discharge simulations is not necessarily the case. While a consistent representation of snow and glacier processes is likely a necessary condition, it is not sufficient on its own for adequately capturing the highly dynamic nature of (sub-)daily discharge. Problems are likely less pronounced on the monthly scale, however, capturing short-term (daily or sub-daily) dynamics of snow and glacier processes poses great challenges and would involve an enormous data requirement, rarely given in these environments. Vice versa, going the traditional way in which models are purely conditioned on discharge comes at the expense of model consistency and ultimately their broader applicability. Holistic approaches that take trade-offs into account are therefore essential for advancements in hydrological modeling.
Besides, every technical novelty reveals further weaknesses and missing elements. In the case of SWAT-GL, the need for seasonally varying lapse rates or radiation-based melt approaches represent only two of such examples.

9.2 Final Discussion & Outlook

In the following outlook, the two main elements of this work, namely climate impact assessment and hydrological modelling as a whole, will be revisited in order to conclude with possible future directions. The term *potential directions* should be emphasized in particular, as both topics are neither black nor white and recommendations as well as critical remarks are based on own observations and experiences gained in the course of developing this work. However, critical examinations of the state of the art and current practices is considered essential to further advance the field.
First, different aspects of climate impact assessment are discussed, followed by a discussion on broader hydrological modelling. Overall, the following topics will be covered:

I) Future Directions in Climate Impact Assessment

   i) The impact statement & its communication

   ii) Hydrosphere feedbacks

   iii) Catchment behavior under past and future climatic conditions

   iv) Providing a global context

v) Attributing trends

II) Hydrological Modeling - A Glimpse in the Future

i) Performance evaluation

ii) Performance criteria

iii) Underpinning the understanding of the process

iv) Holistic, interdisciplinary & human-centered hydrology

9.2.1 Future Directions in Climate Impact Assessment

The Impact Statement & its Communication

It was demonstrated that advances in climate modeling can significantly influence the impact statement and ultimately conclusions for a specific region. The uncertainty inherent in climate model generations should be taken into account, if possible. An easy yet efficient way for studies is to clearly indicate the extent to which the impact statement differs from previous ones, highlighting potential advancements in climate modeling. However, it might be even more relevant to stress similarities found in the evaluation of impacts of climate change that are based on different climate model generations. It is precisely these similarities that enhance trustworthiness and robustness of the impact analysis. This can be further seen in the communication of uncertainties undeniably attached to climate change studies. However, an adequate communication of these raises awareness and benefits the credibility of the study and field. The IPCC language, regardless of its criticism and (justified) suggestions for improvement (Janzwood, 2020), with its distinction between three uncertainty scales at least sets the right direction here. The scale of *agreement* between impact statements is likely the most easy way to orient oneself, but at the same time represents a fundamental necessity. Even if a classification of one's own work in the existing literature should be the standard, the reality in the field of climate impact research is often different. However, especially in times of exponentially increasing publication numbers where the *"publish or perish"* principle is booming (Fernandez-Cano, 2021), this is an important albeit time-consuming work. The concept of using a common terminology, such as but not necessarily from the IPCC, could even be applied within climate impact studies. In its most basic way, a hydrology model that reflects different physical variables shows differences in the quality exactly those are represented. Quality differences occur, either for reasons of model structure or due to calibration schemes that usually focus on a limited number of variables. However, impact statements often cover a wide range of variables without any qualitative differentiation. Once again, an easy approach could be to use the aforementioned but common (and partly comparable) language of uncertainty, where the *evidence* scale would offer a quality-based robustness distinction of simulated impacts of climate change. In its simplest form, a model which was not calibrated (nor validated) for evapotranspiration, but only for discharge, is less likely to provide evapotranspiration projections as robust as those for discharge. The latter should therefore be classified as more reliable, and reflected as such in communication. Conversely, if model results are treated democratically regardless of their background, undifferentiated statements about the effects of climate change can fuel climate skepticism, especially in today's world.

In any case, it would be advisable to conduct at least basic investigations into the robustness of the impact statements, as demonstrated here using the example of parameter effectiveness under different forcing datasets. A further example is to investigate the effects of the baseline period that can have considerable implications on the impact statement, such as a sign reversal, as shown in Liersch et al. (2020).

Hydrosphere Feedbacks

A common aim among climate impact studies in the field of hydrology is the identification of potential shifts in different water balance components as a response to a change in climate. The significance of these studies is unquestionable, and the increasing trends observed in climate impact literature will hopefully continue to grow in the future. However, far fewer studies focus on mechanistic analyses, a potential understudied topic that could therefore set the direction for future studies. Mechanistic should hereby represent a more thorough exploration of driver and response mechanisms that involves the evaluation of process dominance under potential future changes. A catchment, which is subject to persistent climate change might face substantial modifications of the water and energy balance, and process importance could change substantially. While, for instance, understanding the magnitude and direction of changes in evapotranspiration is indeed crucial, there remains ample room for additional explorations in the climate impact community. An example is to shift focus on the role of vegetation functioning under climate change, which is directly associated with the partitioning of evapotranspiration (into transpiration and soil evaporation). Although it is often neglected, regardless whether the underlying model structure does not allow its inclusion at all or the full capabilities of a model are not exploited (as for example in Schaffhauser et al. (2023)), the role of CO_2 on the hydrological response is well known. It is advisable for the community to pay more attention to these feedbacks, such as the effect of elevated CO_2 concentrations on plant activity and ultimately on the interaction of evapotranspiration and soil moisture. Good examples exist, such as in Villani et al. (2024), but in the author's opinion are, they are not yet mainstream. The range of potential mechanistic investigation worthwhile to pay more attention is long. High-mountain environments, as covered in this work, offer a large pool of possibilities. Examples can range from elevation-dependent assessments of hydrological mechanisms to the role of groundwater in the cryosphere. The latter, in particular glacier-groundwater connectivity, seems to remain an understudied topic as just recently shown in van Tiel et al. (2024). Besides, hydrological models are much less frequently used to investigate processes on a sub-daily scale, nevertheless future precipitation partitioning significantly relies on *how* and *when* it falls, which in turn governs infiltration rates. Aggregated daily and sub-daily projections might differ and cause diversions in groundwater storage projections. With respect to SWAT, the Green & Ampt (Heber Green and Ampt, 1911) approach offers the corresponding capabilities, however, is rarely used in the SWAT-based climate change studies (and SWAT studies at all). A more mechanism-centered orientation within the impact community might lead to interesting insights. E.g., the aforementioned study from Villani et al. (2024) highlighted the importance of model representations of evapotranspiration that are valid for CO_2 concentrations higher than 660 ppm. Or to formulate it in a similar way as Vicente-Serrano et al. (2022) in the context of Earth System Models, reliable representations of transpiration and CO_2 effects, which improve climate projections, are one of the most urgent and important research fields that need attention. Moreover, as a side effect more process-oriented impact assessments could promote further model improvements in physically-based models such as SWAT.

Catchment Behavior Under Past and Future Climatic Conditions

In our work we highlighted the importance of robustness assessments as proposed for example in Krysanova et al. (2018, 2020), serving still as one of the most essential steps and requirements in impact assessment. An intrinsic assumption, or hope, is that models that are able to represent the catchment behavior well under past climatic conditions will also perform equally well under comparable future climate conditions. Chosen calibration and validation periods are often relatively short. However, Stephens et al. (2020)

have shown that even equal past and future climatic conditions can cause divergent catchment responses. It was shown to be particularly true if short historical periods are used for model evaluation. Again, vegetation dynamics (including increased CO_2 concentrations) are shown to be one of the key drivers for this effect. In detail, the varying response under persistent climate change. For example, if specific climate conditions sustain for a specific amount of time, vegetation adapts and alters transpiration compared to the historical phase, regardless of equal temperature and precipitation conditions. Similarly, groundwater storages under decades of dry and hot conditions are barely comparable to those of a five years period of similar dry and hot conditions. A legacy effect of catchments that is of particular importance when simulating long (transient) periods. Well acknowledged, but no less challenging for climate impact modelers, is that in most cases the model is extrapolated to conditions largely deviating from those in the historical period. Something which is especially true under the high-emission scenarios. These modeling challenges are also addressed by Duethmann et al. (2020) in the case of Austria. But this is exactly where the opportunity lies for the community. Despite the increasing trend of climate impact studies there are more than enough challenges to be addressed. The potential of exploration is huge. Based on the aforementioned points, studies could explore more structural uncertainties, e.g. how different process-representations behave under future conditions or hydrological models could be improved to represent longer periods subject to transient conditions. A further and specific example that addresses model structure-related uncertainties might focus on the widely used Hargreaves method Hargreaves and Samani (1985) to estimate reference evapotranspiration. The method, based on temperature and extraterrestrial radiation only, builds on lysimeter data from the 1980s for Alta Fescue grass. 8 years of data was used to create the empirical relationship. As in Schaffhauser et al. (2023), the method is commonly used to extrapolate evapotranspiration until 2100. However, does the empirical relationship between temperature, extraterrestrial radiation and the calibrated value of 0.0023 for K_{ET} hold under extremely shifted climatic conditions along with varied plant feedbacks?

Providing a Global Context

As mentioned in the beginning of this work, 2023 was an impressive year of a frightening nature. This was just recently confirmed in the report *"State of the Climate in 2023"* to which more than 590 scientists contributed (Blunden and Boyer, 2024). The Paris Agreement defined the limits of 1.5°C and 2°C, two distinct values barely found in climate impact studies. Although the Paris Agreement does not establish a specific time frame, which may slow the implementation of measures in some areas, greater emphasis might be placed on these figures in climate impact assessments. One could argue that, for this reason, studies should be more closely aligned with global limits. Also regional climate impact studies could aim to incorporate a global perspective rather than overlook it. In detail, climate impact studies could shed light on how a warming of 1.5°C, 2°C or 3°C translates to local hydrological changes, rather than focusing solely on projections for near or far future periods. Back in 2017, James et al. (2017) provided valuable suggestions how the global context could be considered in studies. In public discourse, the question often arises whether half a degree matters or not. These regional variations in the impacts caused by global climate change could give important insights to decision makers and may serve as the tipping point to initiate meaningful action. Besides, they provide a valuable link to the future direction of climate impact-related research. As demonstrated in this work, it's not only the generation of climate models that matters, but significant differences can also exist between models within the same generation. Therefore, it may be important not only to focus on the period when an ensemble mean (or median) reaches global target values, but also to examine the trajectories of the individual climate models (James

et al., 2017). The question, which is proposed can be stated as *"Does the pathway how a global threshold temperature is reached affect regional impacts on hydrology?"*.

Attributing Trends

Trend tests are widely applied in hydrological studies, including climate impact studies. It is also common practice to assign present trends, mostly in the 20th Century, in hydrological variables to anthropogenic climate change, at least as an indicator of it. Even if the evidence is overwhelming or even incontestable that humans are responsible for the lion's share of climate change, the outreach of determining the actual contribution may be much greater. Attributing trends to man-made climate change helps to underline its presence, to indicate its impact and to raise awareness on the consequences of our actions. Comparative studies that show how past impacts might have been dampened, or even vanished, in absence of anthropogenic climate change can be a promising way in to engage stakeholders. If we know to which degree trends (or events) might be caused by human activities, the communication between science and policy makers or various stakeholders could be facilitated. Since extreme events such as floods or droughts in particular, are generally associated with monetary losses, information on the probability of these events intensifying or increasing in frequency might have a decisive influence on action. Projecting impacts over several decades offers important insights, but encountering those effects directly on our doorstep yesterday presents a more immediate and tangible reality. While there is a lot of literature on climate attribution (usually extreme event attribution), the field of attributing climate change (or specific events) to its causes, there is less literature on climate impact attribution that addresses how observed impacts can be attributed to climate change. Recent and promising examples, in which it was investigated whether and how hydrological changes can be attributed to climate change are for example provided in Javed et al. (2023); Didovets et al. (2024). A key aspect of the two examples is that both are based on ATTRICI (ATTRIbuting Climate Impacts) data (Mengel et al., 2021), which provides counterfactual meteorological data, a dataset that does not include long-term climate trends. The data is ready to use and easy to access and hence serves as a fast and efficient approach to perform impact attribution studies within the hydrological community. The benefits and potential of these (rapidly emerging) datasets are far from being fully exploited. Although the data is only statistically detrended and not derived from climate model runs that exclude human-induced emissions, it can provide valuable insights.

9.2.2 Hydrological Modeling - A Glimpse in the Future

Finally, we will take a last look at various aspects of hydrological modeling and its potential future to close the circle started in the Introduction.

Taking again a brief Scopus analysis as a basis, hydrological modeling has gained increasing popularity over the past few decades. Since 2000, the number of documents retrieved using the combined search term *"hydrological model"* or *"hydrology model"* (as a proxy for hydrological model applications) has increased roughly tenfold, rising from 146 documents in 2000 versus 1,450 documents in 2023. Nevertheless, despite substantial progress in some respects, there remain challenges that the community needs to address. Some of these issues are linked more to poor habits and practices rather than to fundamental process-based hydrological problems or gaps in understanding, as illustrated by the well-known work of Blöschl et al. (2019). As a result, many of these improvements can be easily implemented.

Performance Evaluation

A key aspect in nearly all our hydrological model applications, regardless of the absolute model task, is evaluation of the model performance. It is important to place achieved model skills during the calibration and validation phases into context, providing the reader not only a numeric value but also with an estimate whether this value can be considered good or not. A commonly used approach, as done for example in Schaffhauser et al. (2023), is to categorize the chosen performance metrics, which is often based on the widely recognized (and useful) classification system of Moriasi et al. (2007). The popularity and implementation of the proposed classification scheme can be indirectly corroborated by a brief Scopus review, which reveals that the number of citations of the referenced work has grown linearly with the increase in studies related to hydrological modeling, rising from 4 citations in 2007 to 1,125 citations in 2023. The small example is illustrated in Fig. 9.1. It is great that the community works towards making model performances comparable. However, it often appears that modelers rely too heavily on the

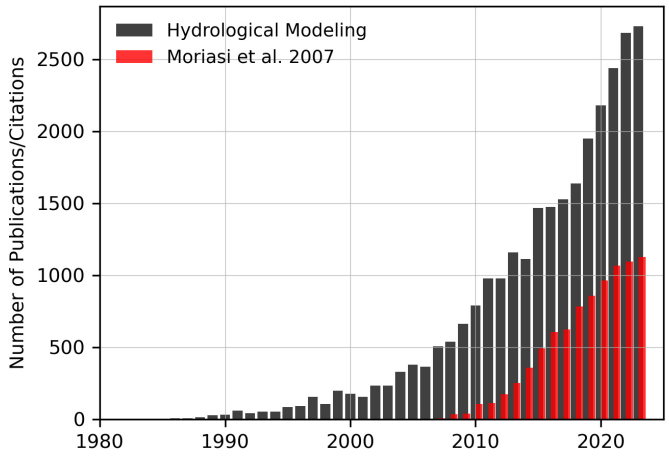

Figure 9.1: Results of the Scopus based literature review. Black bars indicate how the number of studies dealing with hydrological modeling have evolved, red bars illustrate how in parallel the citation of Moriasi et al. (2007) evolved that provides guidance on model ratings.

classification scheme and results that may not be particularly robust or satisfactory, even if classified as such. Especially results classified as "satisfactory", for example, are sometimes presented in a very positive manner, suggesting that the model quality is adequate for the intended task. In reality, "satisfactory" is the second lowest category out of four, with values between 0.5 and 0.65 for the NSE. Relying solely on these predefined classes can be risky, as it may lead modelers to commit too early even when substantially better results can be achieved. This can result in insufficient effort into an accurate representation of hydrological processes, which can impact the study's conclusions. For instance, it might cause a lack of cross-validation of simulated processes, particularly when fully automatic optimizers are employed. This can result in a poor understanding of the basin's processes or a lack of insight into model limitations. Addressing questions like "What doesn't work well in the model and why?" is crucial for advancing hydrological modeling. Not to mention how equifinality affects the representation of variables under similar model performances. Although a thorough literature review of previous modeling activities within the study area is one of the initial tasks that every modeler should ideally undertake, the specific results of model performances are rarely visualized or summarized in tables in hydrological studies. Studies often tend to cover this (incompletely) in the introduction, where it is noted whether previous work was

done and what major conclusions were drawn. However, a more detailed and quantitative presentation of actual model performance values of past studies could substantially increase comparability across studies. This would aid the reader's ability to evaluate the results. Moreover, solutions could make use of existing and fair approaches that deal with this topic. The benchmark test proposed by Seibert et al. (2018b), for example, provides a good way to fairly assess model performance. However, benchmarking is not yet common practice and still remains the exception. Ideally, solutions should include both a clear illustration of other groups modeling efforts and the integration of a benchmarking test. The latter is especially important if a model is subject to a strong seasonal component, as for example in the tropics or snow-dominated basins. As discussed in this work, good performance criteria might be achieved relatively easily in these regions and potentially good performance values could be easily outperformed and thus lowers the actual model skill.
Shifts towards a more open and self-demanding model evaluation would be desirable and positively impact the quality of modeling studies. In turn, this could enhance the overall perspective on model evaluation, which is often heavily focused on individual objective function values. Complete water balance results, which are not yet commonly included in hydrological modeling studies, should become mandatory to give readers a comprehensive view of the impact that a specific model evaluation strategy has had. Consistent implementations would further enhance comparability between studies.

Performance Criteria

Somewhat attached to the aspects made concerning performance evaluation, is the performance criteria itself. Widely acknowledged and followed nowadays, is the use of an ensemble of different performance criteria. Although optimizers that only use one objective are, however, equally widespread. Despite great efforts within the community to introduce novel and superior criteria, these are rarely adopted in the practice of hydrological applications. Examples of developments refer to the relatively old and more frequent used Index of Agreement that was continuously enhanced (Willmott, 1981; Willmott et al., 1985, 2011), or deal with more balanced representations of specific inter-annual conditions reflected in annually separated versions of the NSE or KGE (called split NSE or KGE) (Fowler et al., 2018), as well as the usage of KMoments-based approaches as suggested in Pizarro and Jorquera (2024). A more comprehensive overview and discussion was provided in this work. Recommendations frequently emphasize the importance of considering transformation of the target variable prior to the calculation of standard performance criteria. This is particularly true when least square approaches based on squared errors are used. Among the most commonly used evaluation metrics are the NSE and KGE, both of which minimize squared errors. Despite the discussion on their shortcomings that has a long history dating already back to 1978 in Garrick et al. (1978), the metrics are relatively unrivaled in the community. However, a welcomed tendency observed in the community is that more attempts being made to consider low flow-focused metrics, such as the logNSE, to account for parts of the deficiencies inherent in metrics such as the NSE. Nevertheless, especially under changing climatic conditions a careful evaluation of the choice of the objective function is advisable. Improper selections might constrain the robustness of interpretations or statements with respect to diverse aspects of the target variable. The choice should suit the final goal of the study. Preliminary investigations of the current climate and, ideally, projected future climate conditions can aid the choice as it allows to see if significant drying might justify moving from squared-error approaches to methods based on absolute errors.

Underpinning the Understanding of the Process

Hydrology is linked to a wide range of disciplines, and modelers can find themselves moving into unfamiliar, though related, areas as part of their work. Modeling therefore can rapidly become relevant to or intersect with fields such as agriculture, forestry, ecology, meteorology (and climatology), glaciology, limnology, or geology. Hence, it might be all the more important to continuously expand and improve our process understanding. In fact, hydrology is highly interdisciplinary and increasingly involves advanced technical disciplines, such as remote sensing and artificial intelligence (AI), which influence the methods used to investigate and represent the hydrological system. In the author's view, programming skills might have generally improved, although perhaps with great fluctuations, or technical innovations have at least made programming considerably easier. However, interdisciplinarity and creating advanced skills in these topics can occupy a significant amount of time. Time, a rare commodity, especially in the scientific environment, which in the worst case has a negative impact on the actual scientific work and is currently being discussed a lot, as in a recently published Nature Editorial (Nature, 2024). This contrasts with the expectations that scientists should have a detailed understanding of modern techniques across various fields, including not only hydrology but also remote sensing, handling diverse reanalysis and remote sensing datasets, and AI, where such skills appear almost essential. It is easy to see that this can happen at the expense of the understanding process hydrology. Thus, multifaceted opinions exist on how much expertise should young scientists have already internalized in the subject itself at the beginning of their career? Answers certainly depends strongly on where one draws the line in the interface discipline of hydrology mentioned at the beginning, which in turn is very subjective. Therefore, a practical example using the Priestley-Taylor (Priestley and Taylor, 1972) method for calculating potential evapotranspiration should be provided. While there is likely agreement that a hydrological modeler should be familiar with the concept, there may be no consensus on whether a hydrologist should know all background on the Priestley-Taylor coefficient α_{pet}, also known as dryness coefficient. There might be even more disagreement regarding whether the scientist should know the default value used in SWAT ($\alpha_{pet} = 1.28$), which for example differs from HEC-HMS ($\alpha_{pet} = 1.26$). The divergence could become even more pronounced when discussing the reasons behind these values and the conditions they represent.
It is thus important that our understanding of process hydrology is internalized, maintained and expanded. This is fundamental to further improve our hydrological models. Many of the concepts we rely on have been used for quite some time, underscoring how previous generations had to deal deeply with these processes, potentially allowing them to build greater experience and expertise in process hydrology. It might be dangerous if modelers in the scientific community view the core of their applications, the model, merely as a tool or means to an end, thereby treating it as a black box. On the other hand, don't acknowledging the inherent uncertainties and blindly following a pure model-believing paradigm is equally dangerous. It should be emphasized that we do not necessarily have less process-oriented knowledge than in the past; instead, the focus might have shifted, and priorities are now aligned differently to meet the demands of modern, fast-paced, and output-driven research. The latter is closely linked to the issue of time and an additional ongoing topic of public debate. A recently published Editorial titled *"Should we publish fewer papers?"* from Jin (2024) offers valuable insights into modern academic life. This suggests that gaining a thorough hydrological understanding of the processes might be anything but easy.
Additionally, there is still a significant need for research that focuses on process-based hydrology. Just recently Tarasova et al. (2024) published a comprehensive review, where they identified knowledge gaps with respect to catchment characterization. The work demonstrates for instance that dry environments are heavily understudied, as only 16% of the studies focus on these environments. However, those en-

vironments cover 42% of our land surface. Although the results are not directly linked to hydrological modeling studies, it is likely that similar conclusions could be drawn within the broader field of modeling studies. A specific example of future directions for process-based modeling studies may lay in evaluating and potentially improving runoff generation processes under contrasting conditions, such as dry and humid environments. Dry environments are typically characterized by infiltration excess, while humid environments are usually driven by saturation excess. This raises the question of whether SWAT is equally suitable across different catchments and runoff generation characteristics, especially given that most studies use the empirical SCS-CN method, regardless of its environmental setup. An interesting synthesis dealing with interflow, recently published by McGuire et al. (2024), provides valuable process-related ideas that can be translated into the modeling domain. Interflow, a runoff process that might often be overlooked, albeit its dominance in catchments with steeper slopes and underlying soils of limited permeability. SWAT, which does not differentiate between lateral and vertical hydraulic conductivity, can face challenges in accurately representing lateral flow, as such parameters end up controlling multiple processes. Besides, if impermeable layers are implemented alongside crack flow, the latter outweighs the impermeable layer. Studies that assess different catchment characteristics while evaluating process representations (model structure) could be highly beneficial for the modeling community. However, categorizing catchments based on specific characteristics (climatic, topographic etc.), such as dry or humid, might be too simplistic and incomplete as a feature, given the extensive variability in hydrological features across different scales within dry regions alone. This highlights the importance of process-oriented hydrology and underscores the need for a thorough understanding of the studied catchment. Hydrological signatures, therefore, play a crucial role in effectively utilizing and improving our hydrological models.

Holistic, Interdisciplinary & Human-centered Hydrology

The above-mentioned role of hydrology as an interface discipline holds great potential. Even though our projects have always been interdisciplinary, it is quite likely that interdisciplinarity will become even more important in the future. The human fingerprint is already large, be it due to (anthropogenic) climate change or socio-economic drivers that lead to human interventions and change our land and water management, making unaffected hydrological systems rare. It is especially these areas in which human intervention is already visible or will be influenced to a much greater extent in the future, where a holistic view of water resources is essential to protect and manage it sustainably. Hydrological modeling plays an important role, but on its own, does not constitute a holistic approach. Model coupling and integration, yet not new, may become even more critical in the future to bridge disciplinary boundaries. In particular, it is the feedback between different components that must be adequately represented - something a hydrological model alone can hardly achieve. Feedbacks exist everywhere, whether between water availability and demand, land cover and atmosphere, or the anthroposphere, which impacts all aspects. It needs solutions, in particular in impact assessment, that try to integrate a human-centered component. Challenging. But a continued focus on pure future hydrology without considering potential new water infrastructure or regulations and policy changes that affect our hydrology, we can't expect big future steps in the field. As one now might think of digital twins of the Earth as advanced tool that integrates many of these different components, that is right - however, the reflections published by Saltelli et al. (2024) and titled *"Bring digital twins back to Earth"* contains interesting thoughts on the strong focus on high resolution physical variables along with a lacking societal component. More studies focusing on interactions between economy and hydrology may be indicative and helpful here, as many human interventions are monetary driven. There is certainly still great potential to better integrate and

utilize the knowledge of related disciplines such as agriculture or forestry. This is particularly important in order to develop holistic modeling systems that go beyond piecemeal, where modeling groups from different disciplines pursue individual tasks independently of each other and merge results in the end as a post-processing task. Future holistic approaches may (and should) explore further ways. Something that also applies for the interplay between hydrology and the rapidly emerged field of AI, where groups often focus either on the AI approach or traditional process-based modeling but synergies and collaboration are not yet fully exploited. The considerable progress and highly promising results achieved by AI models in hydrology led to controversial discussions and mixed feelings in the community. An example from recent years is the article from Nearing et al. (2021), which tackles the question what role hydrology in the age of machine learning plays. A title that was probably deliberately chosen to challenge the community. The Google group just published a global AI-based flood forecasting system with very promising results (Nearing et al., 2024). In fact, hydrology should take greater advantage of the strengths offered by novel technologies and remain open to alternatives, whether they complement or eventually replace established approaches. New or improved understanding of the processes may emerge, which could ultimately be reflected in the models we currently use.

Many of the points mentioned, along with this work itself, align with Aristotle's famous phrase:

"The whole is greater than the sum of its parts."

Appendices

Appendix A

Appendix Chapter 4

Table A.1: Overview of the model parameters used in the calibration of the hydrological models driven by the reference datasets of the two ISIMIP families, namely EWEMBI and W5E5. Parameter ranges were individually set and fixed for each calibrated subbasin/gauge.

Category	Parameter	File	Description
Water Balance	PLAPS	.sub	Precipitation Lapse Rate [mm/km]
	TLAPS	.sub	Temperature Lapse Rate [°C/km]
Snow	TIMP	.sno	Snowpack Temperature Lag Factor [-]
	SMFMX	.sno	Melt Factor for Snow on 21st June [mm H_20/(°C*d)]
	SMFMN	.sno	Melt Factor for Snow on 21st December [mm H_20/(°C*d)]
	SMTMP	.sno	Snowmelt Temperature [°C]
	SFTMP	.sno	Snowfall Temperature [°C]
Groundwater	ALPHA_BF	.gw	Baseflow Alpha Factor [days]
	GWQMN	.gw	Shallow Aquifer Threshold for Baseflow to Occur [mm]
	GW_DELAY	.gw	Groundwater Delay [days]
	RCHRG_DP	.gw	Fraction of Percolation for Deep Aquifer [-]
	REVAPMN	.gw	Shallow Aquifer Threshold for Revap to occur [mm]
ETP & Runoff	EPCO	.hru	Plant Uptake Compensation Factor [-]
	ESCO	.hru	Soil Evaporation Compensation Factor [-]
	SURLAG	.hru	Surface Runoff Lag Time [days]
	CN2	.mgt	Curve Number for Moisture Condition II [-]
Soil	SOL_AWC	.sol	Available Water Capacity for the Soil Layer [-]
	SOL_k	.sol	Available Water Capacity for the Soil Layer [mm/h]
	SOL_BD	.sol	Soil Bulk Density [mg/m^3]

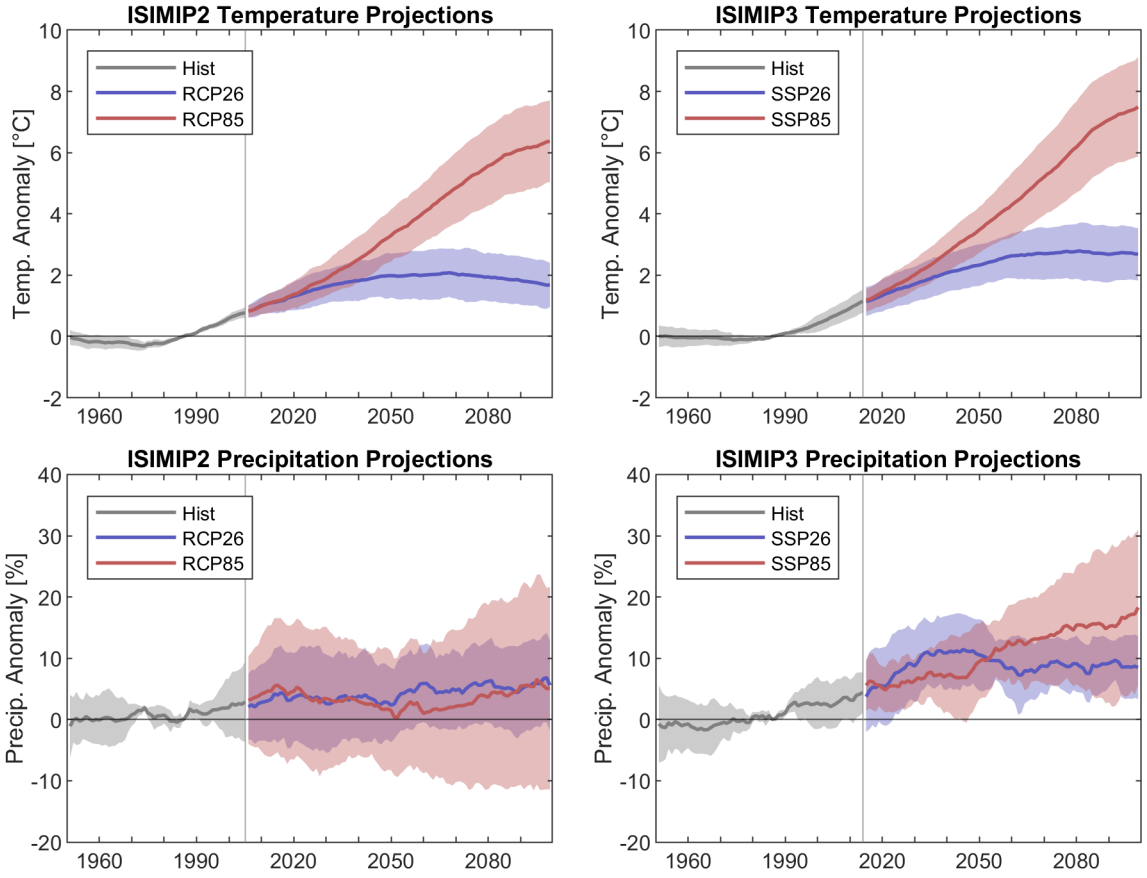

Figure A.1: Temperature and precipitation trajectories at the basin scale for the high and low emission scenarios of ISIMIP2 and ISIMIP3 until 2099. The shaded area refers to the standard deviation of the inter-model range. The straight lines represent the MMM. The historical period is indicated as grey.

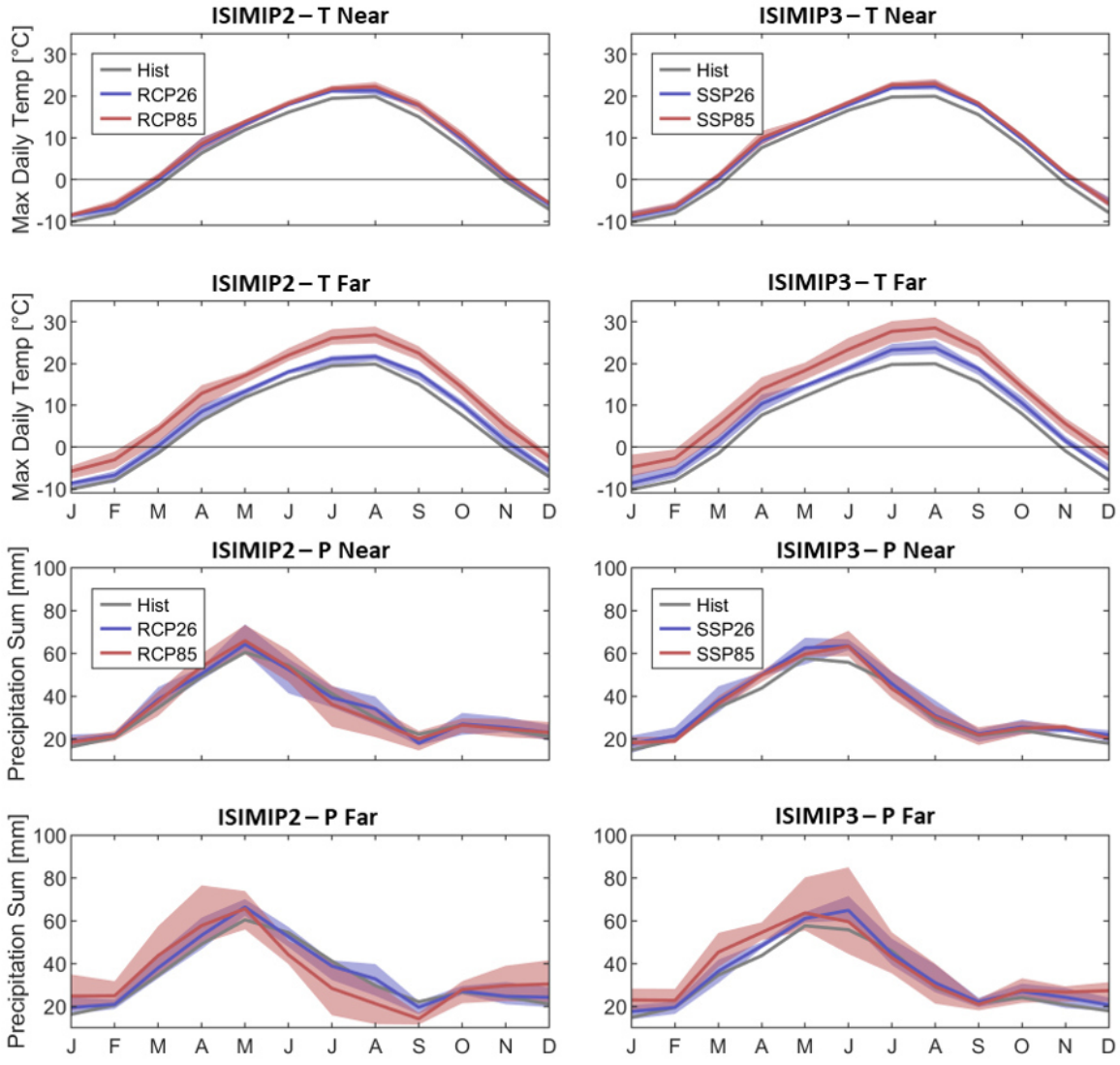

Figure A.2: Seasonal temperature and precipitation trajectories at the basin scale for the high and low emission scenarios of ISIMIP2 and ISIMIP3 for the near and far future period. The shaded area refers to the standard deviation of the inter-model range. The straight lines represent the MMM. The historical period is indicated as grey. P refers to precipitation and T refers to temperature.

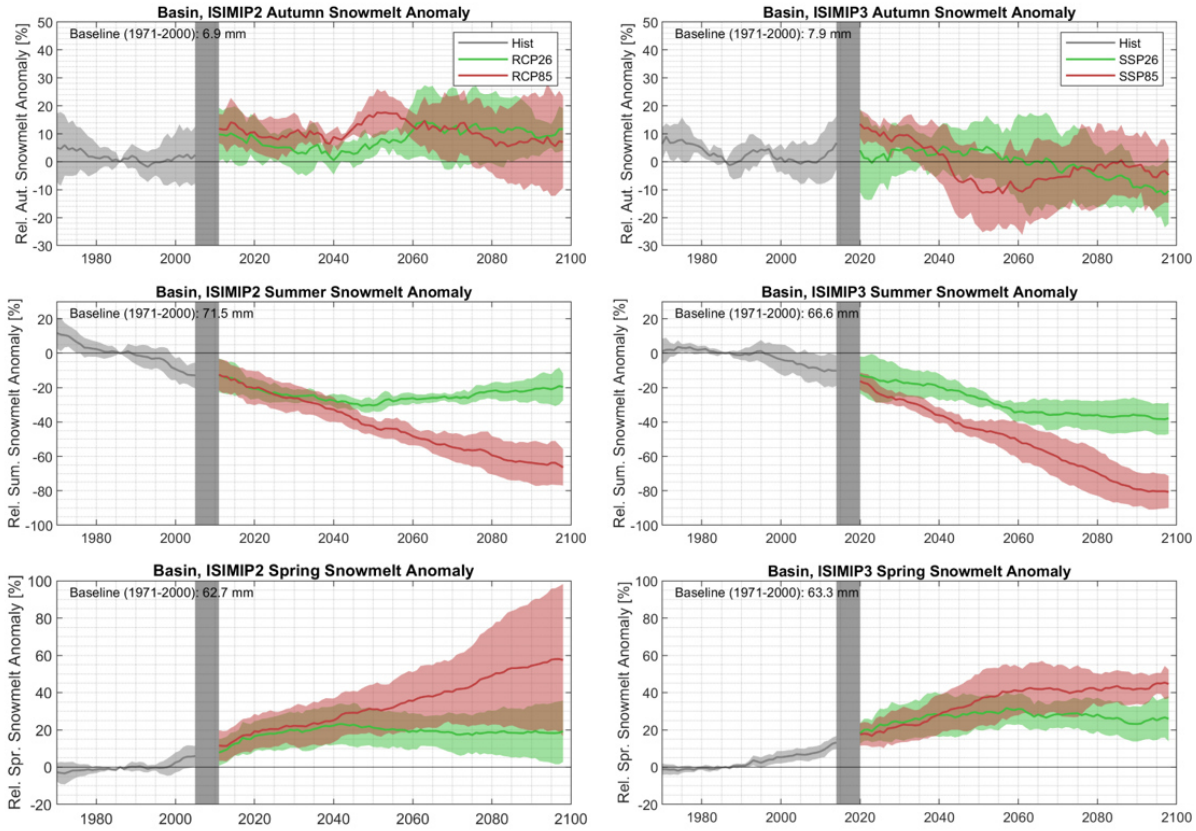

Figure A.3: Seasonal snowmelt projections under ISIMIP2 and ISIMIP3. The meteorological seasons are averages of: Spring - March, April, May; Summer - JJA; Autumn - SON. The vertical grey bar indicates the warm-up season required in the hydrological model, where no data is available.

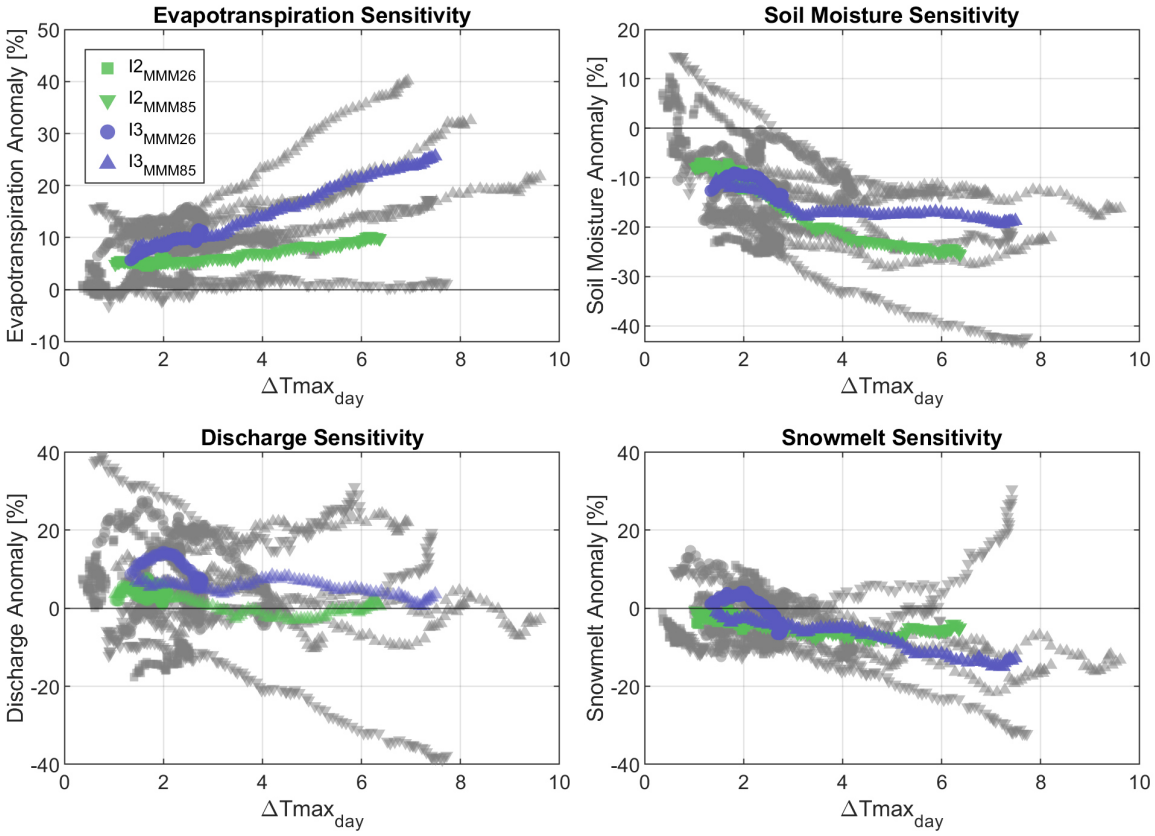

Figure A.4: Sensitivity of evapotranspiration, soil moisture, discharge and snowmelt to temperature variations under ISIMIP2 and ISIMIP3. Sensitivities are assessed for projected annual anomalies of each variables and the respective projected temperature change. I2 refers to ISIMIP2 and I3 to ISIMIP3. Grey markers correspond to the individual GCM results, color symbols represent the multi-model mean (MMM). Results consider all projected values.

Appendix B

Appendix Chapter 7

Table B.1: Parameter ranges and median values of all glaciers for the last generation of the optimization. Note: only the glacier parameters are shown here, which is the main novelty and purpose of the article.

Glacier	GLMLTMP			GLMFMX			GLMFMN			f_frze			f_accu		
	Min	Max	Median	Min	Max	Median	Min	Max	Median	Min	Max	Median	Min	Max	Median
GG	0	4.00	0.70	3.50	6.51	4.05	2.50	8.00	5.76	0.001	0.01	0.0092	0.01	0.70	0.09
LCG	1.56	4.62	3.75	3.50	12.00	3.70	2.50	8.00	2.95	0.001	0.01	0.0033	0.01	0.70	0.45
SCG	0	4.00	3.92	3.50	12.00	4.40	2.50	8.00	2.74	0.001	0.01	0.0059	0.01	0.70	0.01
Wol	1.88	5.00	4.49	3.50	9.53	7.13	2.50	8.00	3.30	0.001	0.01	0.001	0.01	0.70	0.03

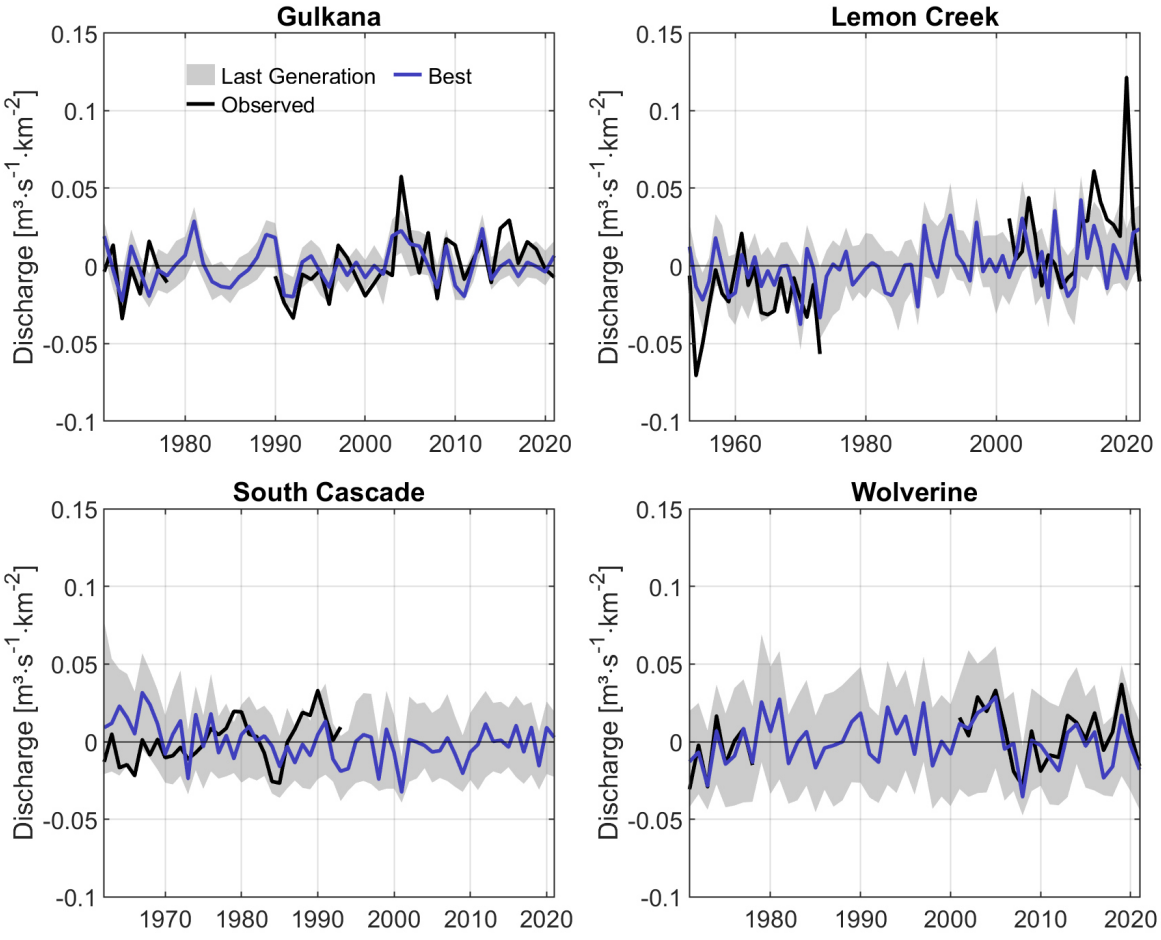

Figure B.1: Simulation results for centered mean annual discharge.

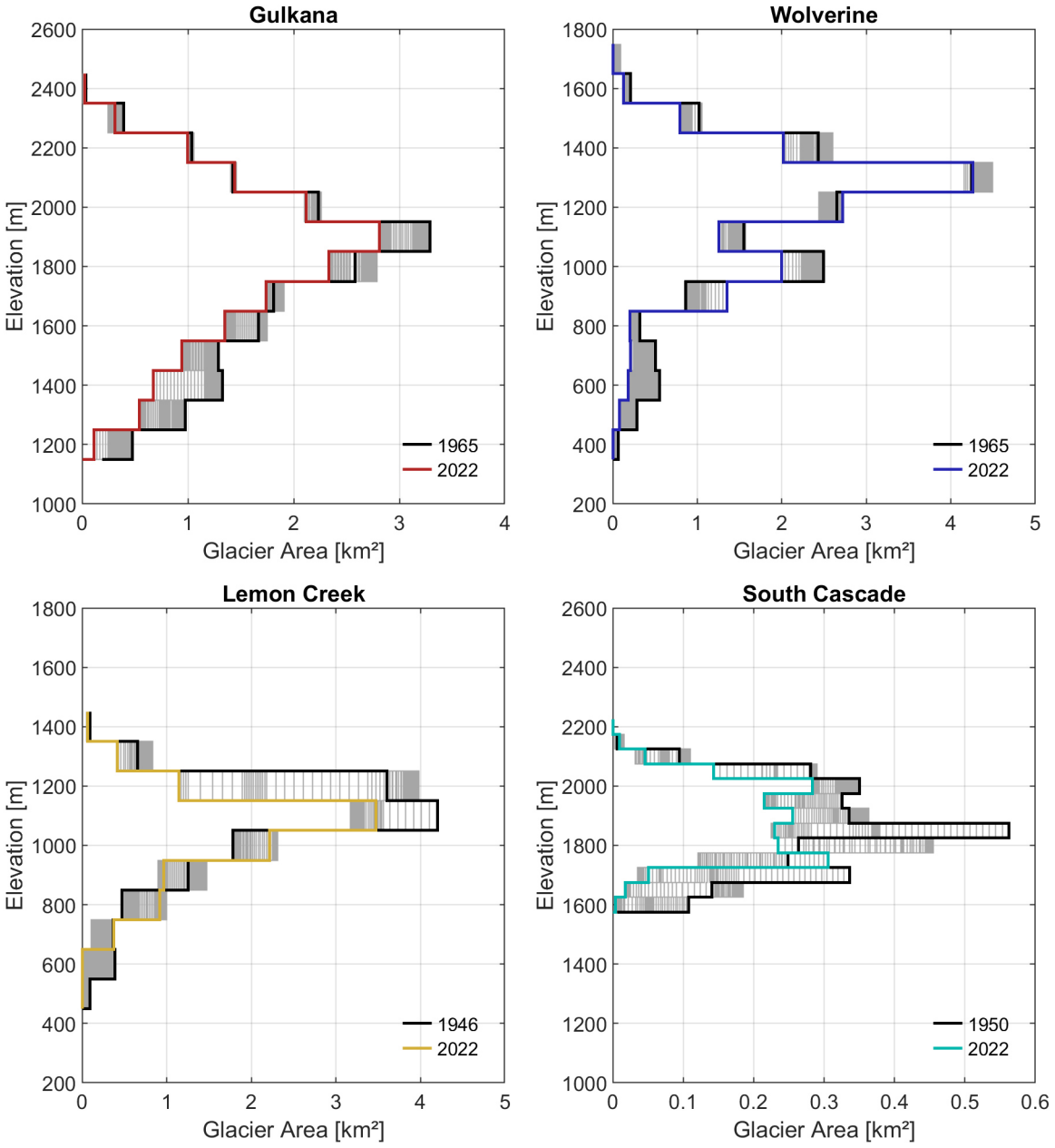

Figure B.2: Annual glacier hypsometry observations for all four glaciers. Grey indicates each year where data was available while black represents the first available year and the individual colors the last available year of measurements.

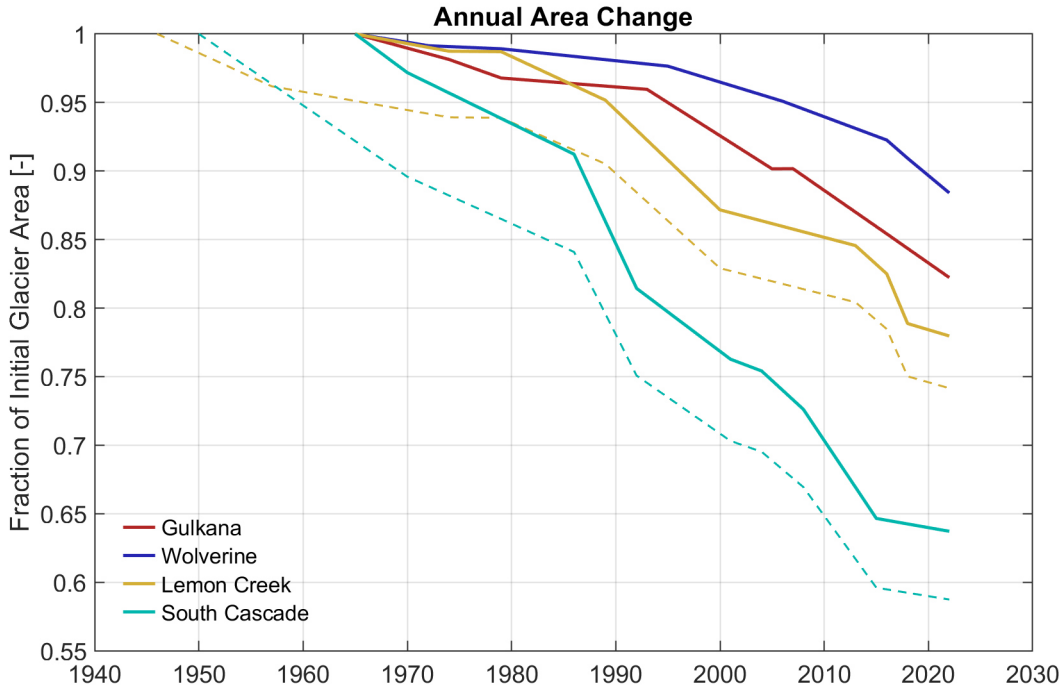

Figure B.3: Observed annual glacier area observations of all glaciers expressed as fraction of initial area.

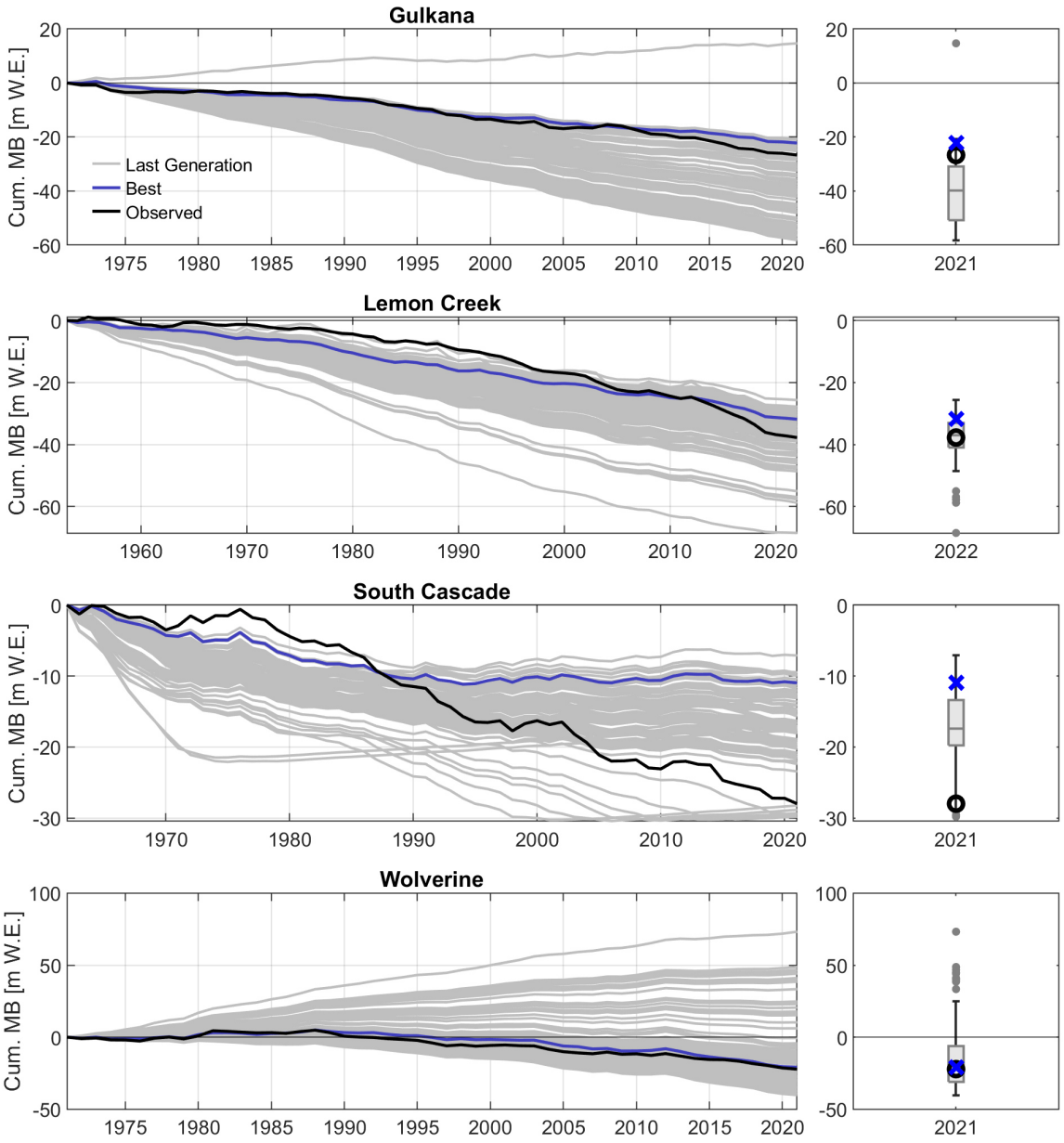

Figure B.4: Similar to cumulative mass balance plot of Fig. 7.7 but providing all individual solutions of the final generation instead of the range. The boxplots provide an estimate of the distribution of the cumulative mass balance at the end of simulation period. In detail, for example at the GG the boxplot consists of the 100 individual cumulative mass balance estimates of year 2021. Blue crosses indicate the results of the best annual mass balance representation in the optimization and black circle indicate the final observed value.

Bibliography

Abrahim, B.N., Cullen, N.J., Conway, J.P., Sirguey, P., 2023. Applying a distributed mass-balance model to identify uncertainties in glaciological mass balance on brewster glacier, new zealand. Journal of Glaciology , 1–17URL: https://doi.org/10.1017/jog.2022.123, doi:10.1017/jog.2022.123.

Adnan, M., Kang, S., Zhang, G., Saifullah, M., Anjum, M.N., Ali, A.F., 2019. Simulation and analysis of the water balance of the nam co lake using SWAT model. Water 11, 1383. URL: https://doi.org/10.3390/w11071383, doi:10.3390/w11071383.

Aizen, V.B., Aizen, E.M., Melack, J.M., 1995. CLIMATE, SNOW COVER, GLACIERS, AND RUNOFF IN THE TIEN SHAN, CENTRAL ASIA. Journal of the American Water Resources Association 31, 1113–1129. URL: https://doi.org/10.1111/j.1752-1688.1995.tb03426.x, doi:10.1111/j.1752-1688.1995.tb03426.x.

Ali, S.H., Bano, I., Kayastha, R.B., Shrestha, A., 2017. COMPARATIVE ASSESSMENT OF RUNOFF AND ITS COMPONENTS IN TWO CATCHMENTS OF UPPER INDUS BASIN BY USING a SEMI DISTRIBUTED GLACIO-HYDROLOGICAL MODEL. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2/W7, 1487–1494. URL: https://doi.org/10.5194/isprs-archives-xlii-2-w7-1487-2017, doi:10.5194/isprs-archives-xlii-2-w7-1487-2017.

Andrianaki, M., Shrestha, J., Kobierska, F., Nikolaidis, N.P., Bernasconi, S.M., 2019. Assessment of SWAT spatial and temporal transferability for a high-altitude glacierized catchment. Hydrology and Earth System Sciences 23, 3219–3232. URL: https://doi.org/10.5194/hess-23-3219-2019, doi:10.5194/hess-23-3219-2019.

Arnold, J.G., Fohrer, N., 2005. SWAT2000: current capabilities and research opportunities in applied watershed modelling. Hydrological Processes 19, 563–572. URL: https://doi.org/10.1002/hyp.5611, doi:10.1002/hyp.5611.

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. LARGE AREA HYDROLOGIC MODELING AND ASSESSMENT PART i: MODEL DEVELOPMENT. Journal of the American Water Resources Association 34, 73–89. URL: https://doi.org/10.1111/j.1752-1688.1998.tb05961.x, doi:10.1111/j.1752-1688.1998.tb05961.x.

Arrhenius, S., 1896. Xxxi.on the influence of carbonic acid in the air upon the temperature of the ground. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 41, 237–276. URL: http://dx.doi.org/10.1080/14786449608620846, doi:10.1080/14786449608620846.

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, 20355–20362. URL: http://dx.doi.org/10.1029/97JB01696, 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, 95–140. URL: https://doi.org/10.1002/2014rg000470, doi:10.1002/2014rg000470.

Baker, E.H., Mcneil, C.J., Sass, L., Peitzsch, E.H., Whorton, E.N., Florentine, C.E., Clark, A.M., Miller, Z.S., Fagre, D.B., O'Neel, S.R., 2018. Usgs benchmark glacier mass balance and project data. URL: https://www2.usgs.gov/landresources/lcs/glacierstudies/mbbmark.asp, doi:10.5066/F7BG2N8R.

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. URL: https://doi.org/10.1016/j.wasec.2020.100072, doi:10.1016/j.wasec.2020.100072.

Barandun, M., Pohl, E., Naegeli, K., McNabb, R., Huss, M., Berthier, E., Saks, T., Hoelzle, M., 2021. Hot spots of glacier mass balance variability in central asia. Geophysical Research Letters 48. URL: https://doi.org/10.1029/2020gl092084, doi:10.1029/2020gl092084.

Berg, A., Lintner, B.R., Findell, K., Seneviratne, S.I., van den Hurk, B., Ducharne, A., Chéruy, F., Hagemann, S., Lawrence, D.M., Malyshev, S., Meier, A., Gentine, P., 2015. Interannual coupling between summertime surface temperature and precipitation over land: Processes and implications for climate change. Journal of Climate 28, 1308–1328. URL: https://doi.org/10.1175/jcli-d-14-00324.1, doi:10.1175/jcli-d-14-00324.1.

Berg, A., Sheffield, J., 2018. Soil moisture-evapotranspiration coupling in CMIP5 models: Relationship with simulated climate and projections. Journal of Climate 31, 4865–4878. URL: https://doi.org/10.1175/jcli-d-17-0757.1, doi:10.1175/jcli-d-17-0757.1.

Berg, A., Sheffield, J., 2019. Historic and projected changes in coupling between soil moisture and evapotranspiration (ET) in CMIP5 models confounded by the role of different ET components. Journal of Geophysical Research: Atmospheres 124, 5791–5806. URL: https://doi.org/10.1029/2018jd029807, doi:10.1029/2018jd029807.

Bernauer, T., Siegfried, T., 2012. Climate change and international water conflict in central asia. Journal of Peace Research 49, 227–239. URL: https://doi.org/10.1177/0022343311425843, doi:10.1177/0022343311425843.

Berndtsson, R., Tussupova, K., 2020. The future of water management in central asia. Water 12, 2241. URL: https://doi.org/10.3390/w12082241, doi:10.3390/w12082241.

Betrie, G.D., Mohamed, Y.A., van Griensven, A., Srinivasan, R., 2011. Sediment management modelling in the blue nile basin using SWAT model. Hydrology and Earth System Sciences 15, 807–818. URL: https://doi.org/10.5194/hess-15-807-2011, doi:10.5194/hess-15-807-2011.

Bieger, K., Arnold, J.G., Rathjens, H., White, M.J., Bosch, D.D., Allen, P.M., Volk, M., Srinivasan, R., 2017. Introduction to SWAT+, a completely restructured version of the soil and water assessment tool. JAWRA Journal of the American Water Resources Association 53, 115–130. URL: https://online

library.wiley.com/doi/abs/10.1111/1752-1688.12482, doi:https://doi.org/10.1111/1752-1688.12482, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/1752-1688.12482.

Blöschl, G., Bierkens, M.F., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J.W., McDonnell, J.J., Savenije, H.H., Sivapalan, M., Stumpp, C., Toth, E., Volpi, E., Carr, G., Lupton, C., Salinas, J., Széles, B., Viglione, A., Aksoy, H., Allen, S.T., Amin, A., Andréassian, V., Arheimer, B., Aryal, S.K., Baker, V., Bardsley, E., Barendrecht, M.H., Bartosova, A., Batelaan, O., Berghuijs, W.R., Beven, K., Blume, T., Bogaard, T., Borges de Amorim, P., Böttcher, M.E., Boulet, G., Breinl, K., Brilly, M., Brocca, L., Buytaert, W., Castellarin, A., Castelletti, A., Chen, X., Chen, Y., Chen, Y., Chifflard, P., Claps, P., Clark, M.P., Collins, A.L., Croke, B., Dathe, A., David, P.C., de Barros, F.P.J., de Rooij, G., Di Baldassarre, G., Driscoll, J.M., Duethmann, D., Dwivedi, R., Eris, E., Farmer, W.H., Feiccabrino, J., Ferguson, G., Ferrari, E., Ferraris, S., Fersch, B., Finger, D., Foglia, L., Fowler, K., Gartsman, B., Gascoin, S., Gaume, E., Gelfan, A., Geris, J., Gharari, S., Gleeson, T., Glendell, M., Gonzalez Bevacqua, A., González-Dugo, M.P., Grimaldi, S., Gupta, A.B., Guse, B., Han, D., Hannah, D., Harpold, A., Haun, S., Heal, K., Helfricht, K., Herrnegger, M., Hipsey, M., Hlaváčiková, H., Hohmann, C., Holko, L., Hopkinson, C., Hrachowitz, M., Illangasekare, T.H., Inam, A., Innocente, C., Istanbulluoglu, E., Jarihani, B., Kalantari, Z., Kalvans, A., Khanal, S., Khatami, S., Kiesel, J., Kirkby, M., Knoben, W., Kochanek, K., Kohnová, S., Kolechkina, A., Krause, S., Kreamer, D., Kreibich, H., Kunstmann, H., Lange, H., Liberato, M.L.R., Lindquist, E., Link, T., Liu, J., Loucks, D.P., Luce, C., Mahé, G., Makarieva, O., Malard, J., Mashtayeva, S., Maskey, S., Mas-Pla, J., Mavrova-Guirguinova, M., Mazzoleni, M., Mernild, S., Misstear, B.D., Montanari, A., Müller-Thomy, H., Nabizadeh, A., Nardi, F., Neale, C., Nesterova, N., Nurtaev, B., Odongo, V.O., Panda, S., Pande, S., Pang, Z., Papacharalampous, G., Perrin, C., Pfister, L., Pimentel, R., Polo, M.J., Post, D., Prieto Sierra, C., Ramos, M.H., Renner, M., Reynolds, J.E., Ridolfi, E., Rigon, R., Riva, M., Robertson, D.E., Rosso, R., Roy, T., Sá, J.H., Salvadori, G., Sandells, M., Schaefli, B., Schumann, A., Scolobig, A., Seibert, J., Servat, E., Shafiei, M., Sharma, A., Sidibe, M., Sidle, R.C., Skaugen, T., Smith, H., Spiessl, S.M., Stein, L., Steinsland, I., Strasser, U., Su, B., Szolgay, J., Tarboton, D., Tauro, F., Thirel, G., Tian, F., Tong, R., Tussupova, K., Tyralis, H., Uijlenhoet, R., van Beek, R., van der Ent, R.J., van der Ploeg, M., Van Loon, A.F., van Meerveld, I., van Nooijen, R., van Oel, P.R., Vidal, J.P., von Freyberg, J., Vorogushyn, S., Wachniew, P., Wade, A.J., Ward, P., Westerberg, I.K., White, C., Wood, E.F., Woods, R., Xu, Z., Yilmaz, K.K., Zhang, Y., 2019. Twenty-three unsolved problems in hydrology (uph) – a community perspective. Hydrological Sciences Journal 64, 1141–1158. URL: http://dx.doi.org/10.1080/02626667.2019.1620507, doi:10.1080/02626667.2019.1620507.

Blunden, J., Boyer, T., 2024. State of the climate in 2023. Bulletin of the American Meteorological Society 105, S1–S484. URL: http://dx.doi.org/10.1175/2024BAMSStateoftheClimate.1, doi:10.1175/2024bamsstateoftheclimate.1.

Bonekamp, P.N.J., de Kok, R.J., Collier, E., Immerzeel, W.W., 2019. Contrasting meteorological drivers of the glacier mass balance between the karakoram and central himalaya. Frontiers in Earth Science 7. URL: https://doi.org/10.3389/feart.2019.00107, doi:10.3389/feart.2019.00107.

Bosshard, T., Carambia, M., Goergen, K., Kotlarski, S., Krahe, P., Zappa, M., Schär, C., 2013. Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections. Water Resources Research 49, 1523–1536. URL: https://doi.org/10.1029/2011wr011533, doi:10.1029/2011wr011533.

Campolongo, F., Saltelli, A., Cariboni, J., 2011. From screening to quantitative sensitivity analysis. a

unified approach. Computer Physics Communications 182, 978–988. URL: https://doi.org/10.1016/j.cpc.2010.12.039, doi:10.1016/j.cpc.2010.12.039.

Chaemiso, S.E., Abebe, A., Pingale, S.M., 2016. Assessment of the impact of climate change on surface hydrological processes using swat: a case study of omo-gibe river basin, ethiopia. Modeling Earth Systems and Environment 2, 1–15. URL: http://dx.doi.org/10.1007/s40808-016-0257-9, doi:10.1007/s40808-016-0257-9.

Changkun, M., Sun, L., Liu, S., Shao, M., Luo, Y., 2015. Impact of climate change on the streamflow in the glacierized chu river basin, central asia. Journal of Arid Land 7, 501–513. URL: https://doi.org/10.1007/s40333-015-0041-0, doi:10.1007/s40333-015-0041-0.

Chen, Y., Li, W., Fang, G., Li, Z., 2017a. Review article: Hydrological modeling in glacierized catchments of central asia – status and challenges. Hydrology and Earth System Sciences 21, 669–684. URL: https://hess.copernicus.org/articles/21/669/2017/, doi:10.5194/hess-21-669-2017.

Chen, Y., Li, W., Fang, G., Li, Z., 2017b. Review article: Hydrological modeling in glacierized catchments of central asia – status and challenges. Hydrology and Earth System Sciences 21, 669–684. URL: https://doi.org/10.5194/hess-21-669-2017, doi:10.5194/hess-21-669-2017.

Chiew, F., McMahon, T., 1994. Application of the daily rainfall-runoff model modhydrolog to 28 australian catchments. Journal of Hydrology 153, 383–416. URL: http://dx.doi.org/10.1016/0022-1694(94)90200-3, doi:10.1016/0022-1694(94)90200-3.

Chiew, F., Stewardson, M., McMahon, T., 1993. Comparison of six rainfall-runoff modelling approaches. Journal of Hydrology 147, 1–36. URL: http://dx.doi.org/10.1016/0022-1694(93)90073-I, doi:10.1016/0022-1694(93)90073-i.

Chiogna, G., Marcolini, G., Engel, M., Wohlmuth, B., 2024. Sensitivity analysis in the wavelet domain: a comparison study. Stochastic Environmental Research and Risk Assessment URL: http://dx.doi.org/10.1007/s00477-023-02654-3, doi:10.1007/s00477-023-02654-3.

Copernicus Climate Change Service, 2019. Era5-land hourly data from 2001 to present. URL: https://cds.climate.copernicus.eu/doi/10.24381/cds.e2161bac, doi:10.24381/CDS.E2161BAC.

Cucchi, M., Weedon, G.P., Amici, A., Bellouin, N., Lange, S., Schmied, H.M., Hersbach, H., Buontempo, C., 2020. WFDE5: bias-adjusted ERA5 reanalysis data for impact studies. Earth System Science Data 12, 2097–2120. URL: https://doi.org/10.5194/essd-12-2097-2020, doi:10.5194/essd-12-2097-2020.

Dawar, D., Ludwig, S., 2014. Differential evolution with dither and annealed scale factor, in: 2014 IEEE Symposium on Differential Evolution (SDE), IEEE. URL: http://dx.doi.org/10.1109/SDE.2014.7031528, doi:10.1109/sde.2014.7031528.

Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation 6, 182–197. URL: https://doi.org/10.1109/4235.996017, doi:10.1109/4235.996017.

Dembélé, M., Hrachowitz, M., Savenije, H.H.G., Mariéthoz, G., Schaefli, B., 2020. Improving the predictive skill of a distributed hydrological model by calibration on spatial patterns with multiple satellite data sets 56. URL: https://doi.org/10.1029/2019wr026085, doi:10.1029/2019wr026085.

Deng, H., Chen, Y., 2017. Influences of recent climate change and human activities on water storage variations in central asia. Journal of Hydrology 544, 46–57. URL: https://doi.org/10.1016/j.jhydrol.2016.11.006, doi:10.1016/j.jhydrol.2016.11.006.

Dessu, S.B., Melesse, A.M., 2012. Impact and uncertainties of climate change on the hydrology of the mara river basin, kenya/tanzania. Hydrological Processes 27, 2973–2986. URL: http://dx.doi.org/10.1002/hyp.9434, doi:10.1002/hyp.9434.

Didovets, I., Krysanova, V., Nurbatsina, A., Fallah, B., Krylova, V., Saparova, A., Niyazov, J., Kalashnikova, O., Hattermann, F.F., 2024. Attribution of current trends in streamflow to climate change for 12 central asian catchments. Climatic Change 177. URL: http://dx.doi.org/10.1007/s10584-023-03673-3, doi:10.1007/s10584-023-03673-3.

Didovets, I., Lobanova, A., Krysanova, V., Menz, C., Babagalieva, Z., Nurbatsina, A., Gavrilenko, N., Khamidov, V., Umirbekov, A., Qodirov, S., Muhyyew, D., Hattermann, F.F., 2021. Central asian rivers under climate change: Impacts assessment in eight representative catchments. Journal of Hydrology: Regional Studies 34, 100779. URL: https://doi.org/10.1016/j.ejrh.2021.100779, doi:10.1016/j.ejrh.2021.100779.

Dirmeyer, P.A., Jin, Y., Singh, B., Yan, X., 2013. Trends in land-atmosphere interactions from CMIP5 simulations. Journal of Hydrometeorology 14, 829–849. URL: https://doi.org/10.1175/jhm-d-12-0107.1, doi:10.1175/jhm-d-12-0107.1.

Döll, P., Fiedler, K., Zhang, J., 2009. Global-scale analysis of river flow alterations due to water withdrawals and reservoirs. Hydrology and Earth System Sciences 13, 2413–2432. URL: https://doi.org/10.5194/hess-13-2413-2009, doi:10.5194/hess-13-2413-2009.

Dozier, J., 1989. Spectral signature of alpine snow cover from the landsat thematic mapper. Remote Sensing of Environment 28, 9–22. URL: https://www.sciencedirect.com/science/article/pii/0034425789901016, doi:https://doi.org/10.1016/0034-4257(89)90101-6.

Du, X., Silwal, G., Faramarzi, M., 2022. Investigating the impacts of glacier melt on stream temperature in a cold-region watershed: Coupling a glacier melt model with a hydrological model. Journal of Hydrology 605, 127303. URL: https://doi.org/10.1016/j.jhydrol.2021.127303, doi:10.1016/j.jhydrol.2021.127303.

Duethmann, D., Blöschl, G., Parajka, J., 2020. Why does a conceptual hydrological model fail to correctly predict discharge changes in response to climate change? Hydrology and Earth System Sciences 24, 3493–3511. URL: http://dx.doi.org/10.5194/hess-24-3493-2020, doi:10.5194/hess-24-3493-2020.

Duethmann, D., Bolch, T., Farinotti, D., Kriegel, D., Vorogushyn, S., Merz, B., Pieczonka, T., Jiang, T., Su, B., Güntner, A., 2015. Attribution of streamflow trends in snow and glacier melt-dominated catchments of the tarim river, central asia. Water Resources Research 51, 4727–4750. URL: https://doi.org/10.1002/2014wr016716, doi:10.1002/2014wr016716.

Edwards, P.N., 2011. History of climate modeling. WIREs Climate Change 2, 128–139. URL: https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcc.95, doi:https://doi.org/10.1002/wcc.95, arXiv:https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcc.95.

Ehret, U., Zehe, E., 2011. Series distance – an intuitive metric to quantify hydrograph similarity in terms of occurrence, amplitude and timing of hydrological events. Hydrology and Earth System Sciences 15, 877–896. URL: http://dx.doi.org/10.5194/hess-15-877-2011, doi:10.5194/hess-15-877-2011.

ESA, 2017. Land Cover CCI Product User Guide Version 2. Tech. Rep. (2017). Available at maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf.

FAO, IIASA, ISRIC, ISSCAS, JRC, 2012. Harmonized World Soil Database (version 1.2). AO, Rome, Italy and IIASA, Laxenburg, Austria.

Farinotti, D., Huss, M., Fürst, J.J., Landmann, J., Machguth, H., Maussion, F., Pandit, A., 2019. A consensus estimate for the ice thickness distribution of all glaciers on earth. Nature Geoscience 12, 168–173. URL: https://doi.org/10.1038/s41561-019-0300-3, doi:10.1038/s41561-019-0300-3.

Fernandez-Cano, A., 2021. Letter to the editor: publish, publish ... cursed! Scientometrics 126, 3673–3682. URL: http://dx.doi.org/10.1007/s11192-020-03833-7, doi:10.1007/s11192-020-03833-7.

Fowler, K., Peel, M., Western, A., Zhang, L., 2018. Improved rainfall-runoff calibration for drying climate: Choice of objective function. Water Resources Research 54, 3392–3408. URL: http://dx.doi.org/10.1029/2017WR022466, doi:10.1029/2017wr022466.

Fowler, K.J.A., Peel, M.C., Western, A.W., Zhang, L., Peterson, T.J., 2016. Simulating runoff under changing climatic conditions: Revisiting an apparent deficiency of conceptual rainfall-runoff models. Water Resources Research 52, 1820–1846. URL: http://dx.doi.org/10.1002/2015WR018068, doi:10.1002/2015wr018068.

Frey, H., Machguth, H., Huss, M., Huggel, C., Bajracharya, S., Bolch, T., Kulkarni, A., Linsbauer, A., Salzmann, N., Stoffel, M., 2014. Estimating the volume of glaciers in the himalayan–karakoram region using different methods. The Cryosphere 8, 2313–2333. URL: https://doi.org/10.5194/tc-8-2313-2014, doi:10.5194/tc-8-2313-2014.

Frieler, K., Lange, S., Piontek, F., Reyer, C.P.O., Schewe, J., Warszawski, L., Zhao, F., Chini, L., Denvil, S., Emanuel, K., Geiger, T., Halladay, K., Hurtt, G., Mengel, M., Murakami, D., Ostberg, S., Popp, A., Riva, R., Stevanovic, M., Suzuki, T., Volkholz, J., Burke, E., Ciais, P., Ebi, K., Eddy, T.D., Elliott, J., Galbraith, E., Gosling, S.N., Hattermann, F., Hickler, T., Hinkel, J., Hof, C., Huber, V., Jägermeyr, J., Krysanova, V., Marcé, R., Schmied, H.M., Mouratiadou, I., Pierson, D., Tittensor, D.P., Vautard, R., van Vliet, M., Biber, M.F., Betts, R.A., Bodirsky, B.L., Deryng, D., Frolking, S., Jones, C.D., Lotze, H.K., Lotze-Campen, H., Sahajpal, R., Thonicke, K., Tian, H., Yamagata, Y., 2017. Assessing the impacts of 1.5 °C global warming - simulation protocol of the inter-sectoral impact model intercomparison project (ISIMIP2b). Geoscientific Model Development 10, 4321–4345. URL: https://doi.org/10.5194/gmd-10-4321-2017, doi:10.5194/gmd-10-4321-2017.

Galos, S.P., Klug, C., Maussion, F., Covi, F., Nicholson, L., Rieg, L., Gurgiser, W., Mölg, T., 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, 1417–1439. URL: https://tc.copernicus.org/articles/11/1417/2017/, doi:10.5194/tc-11-1417-2017.

Gan, R., Luo, Y., Zuo, Q., Sun, L., 2015. Effects of projected climate change on the glacier and runoff generation in the naryn river basin, central asia. Journal of Hydrology 523, 240–251. URL: https://doi.org/10.1016/j.jhydrol.2015.01.057, doi:10.1016/j.jhydrol.2015.01.057.

Garcia Sanchez, D., Lacarrière, B., Musy, M., Bourges, B., 2014. Application of sensitivity analysis in building energy simulations: Combining first- and second-order elementary effects methods. Energy and Buildings 68, 741–750. URL: http://dx.doi.org/10.1016/j.enbuild.2012.08.048, doi:10.1016/j.enbuild.2012.08.048.

Garrick, M., Cunnane, C., Nash, J.E., 1978. A criterion of efficiency for rainfall-runoff models. Journal of Hydrology 36, 375–381. URL: https://api.semanticscholar.org/CorpusID:128762270.

Gelfan, A., Kalugin, A., Krylenko, I., Nasonova, O., Gusev, Y., Kovalev, E., 2020. Does a successful comprehensive evaluation increase confidence in a hydrological model intended for climate impact assessment? Climatic Change 163, 1165–1185. URL: https://doi.org/10.1007/s10584-020-02930-z, doi:10.1007/s10584-020-02930-z.

Gessner, U., Naeimi, V., Klein, I., Kuenzer, C., Klein, D., Dech, S., 2013. The relationship between precipitation anomalies and satellite-derived vegetation activity in central asia. Global and Planetary Change 110, 74–87. URL: https://doi.org/10.1016/j.gloplacha.2012.09.007, doi:10.1016/j.gloplacha.2012.09.007.

Grusson, Y., Sun, X., Gascoin, S., Sauvage, S., Raghavan, S., Anctil, F., Sáchez-Pérez, J.M., 2015. Assessing the capability of the swat model to simulate snow, snow melt and streamflow dynamics over an alpine watershed. Journal of Hydrology 531, 574–588. URL: http://dx.doi.org/10.1016/j.jhydrol.2015.10.070, doi:10.1016/j.jhydrol.2015.10.070.

Guo, H., He, S., Li, M., Bao, A., Chen, T., Zheng, G., Maeyer, P.D., 2021. Future changes of drought characteristics in coupled model intercomparison project phase 6 shared socioeconomic pathway scenarios over central asia. International Journal of Climatology URL: https://doi.org/10.1002/joc.7450, doi:10.1002/joc.7450.

Gupta, H.V., Kling, H., 2011. On typical range, sensitivity, and normalization of mean squared error and nash-sutcliffe efficiency type metrics. Water Resources Research 47. URL: http://dx.doi.org/10.1029/2011WR010962, doi:10.1029/2011wr010962.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and nse performance criteria: Implications for improving hydrological modelling. Journal of Hydrology 377, 80–91. URL: http://dx.doi.org/10.1016/j.jhydrol.2009.08.003, doi:10.1016/j.jhydrol.2009.08.003.

Gyawali, D.R., Bárdossy, A., 2022. Development and parameter estimation of snowmelt models using spatial snow-cover observations from modis. Hydrology and Earth System Sciences 26, 3055–3077. URL: http://dx.doi.org/10.5194/hess-26-3055-2022, doi:10.5194/hess-26-3055-2022.

Hagg, W., Braun, L., Kuhn, M., Nesgaard, T., 2007. Modelling of hydrological response to climate change in glacierized central asian catchments. Journal of Hydrology 332, 40–53. URL: https://doi.org/10.1016/j.jhydrol.2006.06.021, doi:10.1016/j.jhydrol.2006.06.021.

Hagg, W., Hoelzle, M., Wagner, S., Mayr, E., Klose, Z., 2013. Glacier and runoff changes in the rukhk catchment, upper amu-darya basin until 2050. Global and Planetary Change 110, 62–73. URL: https://doi.org/10.1016/j.gloplacha.2013.05.005, doi:10.1016/j.gloplacha.2013.05.005.

Hamed, K.H., Ramachandra Rao, A., 1998. A modified mann-kendall trend test for autocorrelated data. Journal of Hydrology 204, 182–196. URL: http://dx.doi.org/10.1016/S0022-1694(97)00125-X, doi:10.1016/s0022-1694(97)00125-x.

Hamidov, A., Helming, K., Balla, D., 2016. Impact of agricultural land use in central asia: a review. Agronomy for Sustainable Development 36. URL: https://doi.org/10.1007/s13593-015-0337-7, doi:10.1007/s13593-015-0337-7.

Hargreaves, G.H., Samani, Z.A., 1985. Reference crop evapotranspiration from temperature. Applied Engineering in Agriculture 1, 96–99. URL: http://dx.doi.org/10.13031/2013.26773, doi:10.13031/2013.26773.

Hassan, J., qing Chen, X., Kayastha, R.B., Nie, Y., 2021. Multi-model assessment of glacio-hydrological changes in central karakoram, pakistan. Journal of Mountain Science 18, 1995–2011. URL: https://doi.org/10.1007/s11629-021-6748-9, doi:10.1007/s11629-021-6748-9.

Hattermann, F.F., Vetter, T., Breuer, L., Su, B., Daggupati, P., Donnelly, C., Fekete, B., Flörke, F., Gosling, S.N., Hoffmann, P., Liersch, S., Masaki, Y., Motovilov, Y., Müller, C., Samaniego, L., Stacke, T., Wada, Y., Yang, T., Krysnaova, V., 2018. Sources of uncertainty in hydrological climate impact assessment: a cross-scale study. Environmental Research Letters 13, 015006. URL: https://dx.doi.org/10.1088/1748-9326/aa9938, doi:10.1088/1748-9326/aa9938.

Hausfather, Z., Drake, H.F., Abbott, T., Schmidt, G.A., 2020. Evaluating the performance of past climate model projections. Geophysical Research Letters 47. URL: http://dx.doi.org/10.1029/2019GL085378, doi:10.1029/2019gl085378.

Heber Green, W., Ampt, G.A., 1911. Studies on soil phyics. The Journal of Agricultural Science 4, 1–24. URL: http://dx.doi.org/10.1017/S0021859600001441, doi:10.1017/s0021859600001441.

Helfricht, K., Huss, M., Fischer, A., Otto, J.C., 2019. Calibrated ice thickness estimate for all glaciers in austria. Frontiers in Earth Science 7. URL: https://doi.org/10.3389/feart.2019.00068, doi:10.3389/feart.2019.00068.

Hill, A., Minbaeva, C., Wilson, A., Satylkanov, R., 2017. Hydrologic controls and water vulnerabilities in the naryn river basin, kyrgyzstan: A socio-hydro case study of water stressors in central asia. Water 9, 325. URL: https://doi.org/10.3390/w9050325, doi:10.3390/w9050325.

Himanshu, S.K., Pandey, A., Yadav, B., Gupta, A., 2019. Evaluation of best management practices for sediment and nutrient loss control using SWAT model. Soil and Tillage Research 192, 42–58. URL: https://doi.org/10.1016/j.still.2019.04.016, doi:10.1016/j.still.2019.04.016.

Hock, R., 1999. A distributed temperature-index ice- and snowmelt model including potential direct solar radiation. Journal of Glaciology 45, 101–111. URL: http://dx.doi.org/10.3189/S0022143000003087, 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. URL: https://doi.org/10.1016/j.hydroa.2022.100123, doi:10.1016/j.hydroa.2022.100123.

Horlings, A., 2016. A Numerical Modeling Investigation on Calving and the Recession of South Cascade Glacier. Technical Report. URL: https://doi.org/10.15760/honors.307, doi:10.15760/honors.307.

Hu, Z., Chen, X., Chen, D., Li, J., Wang, S., Zhou, Q., Yin, G., Guo, M., 2018. "dry gets drier, wet gets wetter": A case study over the arid regions of central asia. International Journal of Climatology 39, 1072–1091. URL: https://doi.org/10.1002/joc.5863, doi:10.1002/joc.5863.

Hu, Z., Zhang, C., Hu, Q., Tian, H., 2014. Temperature changes in central asia from 1979 to 2011 based on multiple datasets. Journal of Climate 27, 1143–1167. URL: https://doi.org/10.1175/jcli-d-13-00064.1, doi:10.1175/jcli-d-13-00064.1.

Huang, A., Zhou, Y., Zhang, Y., Huang, D., Zhao, Y., Wu, H., 2014. Changes of the annual precipitation over central asia in the twenty-first century projected by multimodels of CMIP5. Journal of Climate 27, 6627–6646. URL: https://doi.org/10.1175/jcli-d-14-00070.1, doi:10.1175/jcli-d-14-00070.1.

Huang, J., Su, F., Yao, T., Sun, H., 2022. Runoff regime, change, and attribution in the upper syr darya and amu darya, central asia. Journal of Hydrometeorology 23, 1563–1585. URL: http://dx.doi.org/10.1175/JHM-D-22-0036.1, doi:10.1175/jhm-d-22-0036.1.

Huang, J., Zhang, G., Zhang, Y., Guan, X., Wei, Y., Guo, R., 2020a. Global desertification vulnerability to climate change and human activities. Land Degradation & Development 31, 1380–1391. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ldr.3556, doi:https://doi.org/10.1002/ldr.3556, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/ldr.3556.

Huang, S., Shah, H., Naz, B.S., Shrestha, N., Mishra, V., Daggupati, P., Ghimire, U., Vetter, T., 2020b. Impacts of hydrological model calibration on projected hydrological changes under climate change—a multi-model assessment in three large river basins. Climatic Change 163, 1143–1164. URL: https://doi.org/10.1007/s10584-020-02872-6, doi:10.1007/s10584-020-02872-6.

Huss, M., Farinotti, D., Bauder, A., Funk, M., 2008. Modelling runoff from highly glacierized alpine drainage basins in a changing climate. Hydrological Processes 22, 3888–3902. URL: https://doi.org/10.1002/hyp.7055, doi:10.1002/hyp.7055.

Huss, M., Hock, R., 2015. A new model for global glacier change and sea-level rise. Frontiers in Earth Science 3. URL: https://doi.org/10.3389/feart.2015.00054, doi:10.3389/feart.2015.00054.

Huss, M., Hock, R., 2018. Global-scale hydrological response to future glacier mass loss. Nature Climate Change 8, 135–140. URL: https://doi.org/10.1038/s41558-017-0049-x, doi:10.1038/s41558-017-0049-x.

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, 815–829. URL: https://doi.org/10.5194/hess-14-815-2010, doi:10.5194/hess-14-815-2010.

Ibrahim, Y., Balzter, H., Kaduk, J., Tucker, C., 2015. Land degradation assessment using residual trend analysis of GIMMS NDVI3g, soil moisture and rainfall in sub-saharan west africa from 1982 to 2012. Remote Sensing 7, 5471–5494. URL: https://doi.org/10.3390/rs70505471, doi:10.3390/rs70505471.

Immerzeel, W.W., Pellicciotti, F., Bierkens, M.F.P., 2013. Rising river flows throughout the twenty-first century in two himalayan glacierized watersheds. Nature Geoscience 6, 742–745. URL: https://doi.org/10.1038/ngeo1896, doi:10.1038/ngeo1896.

IPCC, 2023a. IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland. Cambridge University Press. URL: http://dx.doi.org/10.59327/IPCC/AR6-9789291691647, doi:10.59327/ipcc/ar6-9789291691647.

IPCC, 2023b. IPCC, 2023: Climate Change 2023: Synthesis Report, Summary for Policymakers. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland. Cambridge University Press. URL: http://dx.doi.org/10.59327/IPCC/AR6-9789291691647.001, doi:10.59327/ipcc/ar6-9789291691647.001.

Ismail, M.F., Naz, B.S., Wortmann, M., Disse, M., Bowling, L.C., Bogacki, W., 2020. Comparison of two model calibration approaches and their influence on future projections under climate change in the upper indus basin 163, 1227–1246. URL: https://doi.org/10.1007/s10584-020-02902-3, doi:10.1007/s10584-020-02902-3.

Jain, S.K., Sudheer, K.P., 2008. Fitting of hydrologic models: A close look at the nash–sutcliffe index. Journal of Hydrologic Engineering 13, 981–986. URL: http://dx.doi.org/10.1061/(ASCE)1084-0699(2008)13:10(981), doi:10.1061/(asce)1084-0699(2008)13:10(981).

James, R., Washington, R., Schleussner, C.F., Rogelj, J., Conway, D., 2017. Characterizing half-a-degree difference: a review of methods for identifying regional climate responses to global warming targets. WIREs Climate Change 8. URL: https://doi.org/10.1002/wcc.457, doi:10.1002/wcc.457.

Janzwood, S., 2020. Confident, likely, or both? the implementation of the uncertainty language framework in ipcc special reports. Climatic Change 162, 1655–1675. URL: http://dx.doi.org/10.1007/s10584-020-02746-x, doi:10.1007/s10584-020-02746-x.

Javed, M., Didovets, I., Böhner, J., Hasson, S.u., 2023. Attributing historical streamflow changes in the jhelum river basin to climate change. Climatic Change 176. URL: http://dx.doi.org/10.1007/s10584-023-03628-8, doi:10.1007/s10584-023-03628-8.

Ji, H., Fang, G., Yang, J., Chen, Y., 2019. Multi-objective calibration of a distributed hydrological model in a highly glacierized watershed in central asia. Water 11, 554. URL: https://doi.org/10.3390/w11030554, doi:10.3390/w11030554.

Jiang, J., Zhou, T., Chen, X., Zhang, L., 2020. Future changes in precipitation over central asia based on CMIP6 projections. Environmental Research Letters 15, 054009. URL: https://doi.org/10.1088/1748-9326/ab7d03, doi:10.1088/1748-9326/ab7d03.

Jiang, L., Bao, A., Jiapaer, G., Liu, R., Yuan, Y., Yu, T., 2022. Monitoring land degradation and assessing its drivers to support sustainable development goal 15.3 in central asia. Science of The Total Environment 807, 150868. URL: https://doi.org/10.1016/j.scitotenv.2021.150868, doi:10.1016/j.scitotenv.2021.150868.

Jiang, L., Guli·Jiapaer, Bao, A., Guo, H., Ndayisaba, F., 2017. Vegetation dynamics and responses to climate change and human activities in central asia. Science of The Total Environment 599-600, 967–980. URL: https://doi.org/10.1016/j.scitotenv.2017.05.012, doi:10.1016/j.scitotenv.2017.05.012.

Jiang, X., Wang, N., He, J., Wu, X., Song, G., 2010. A distributed surface energy and mass balance model and its application to a mountain glacier in china. Chinese Science Bulletin 55, 2079–2087. URL: https://doi.org/10.1007/s11434-010-3068-9, doi:10.1007/s11434-010-3068-9.

Jin, S., 2024. Should we publish fewer papers? ACS Energy Letters 9, 4196–4198. URL: http://dx.doi.org/10.1021/acsenergylett.4c01991, doi:10.1021/acsenergylett.4c01991.

Katzav, J., Parker, W.S., 2015. The future of climate modeling. Climatic Change 132, 475–487. URL: http://dx.doi.org/10.1007/s10584-015-1435-x, doi:10.1007/s10584-015-1435-x.

Kay, A.L., Davies, H.N., Bell, V.A., Jones, R.G., 2008. Comparison of uncertainty sources for climate change impacts: flood frequency in england. Climatic Change 92, 41–63. URL: https://doi.org/10.1007/s10584-008-9471-4, doi:10.1007/s10584-008-9471-4.

Kennedy, J., Eberhart, R., 1995. Particle swarm optimization, p. 1942 – 1948. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-0029535737&partnerID=40&md5=e6bf04ae50f3268ae545d88ed91d1fc5. cited by: 55921.

Khan, A., Koch, M., 2018. Correction and informed regionalization of precipitation data in a high mountainous region (upper indus basin) and its effect on SWAT-modelled discharge. Water 10, 1557. URL: https://doi.org/10.3390/w10111557, doi:10.3390/w10111557.

Kling, H., Fuchs, M., Paulin, M., 2012. Runoff conditions in the upper danube basin under an ensemble of climate change scenarios. Journal of Hydrology 424–425, 264–277. URL: http://dx.doi.org/10.1016/j.jhydrol.2012.01.011, doi:10.1016/j.jhydrol.2012.01.011.

Knoben, W.J.M., Freer, J.E., Woods, R.A., 2019. Technical note: Inherent benchmark or not? comparing nash–sutcliffe and kling–gupta efficiency scores. Hydrology and Earth System Sciences 23, 4323–4331. URL: http://dx.doi.org/10.5194/hess-23-4323-2019, doi:10.5194/hess-23-4323-2019.

Krause, P., Boyle, D.P., Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. Advances in Geosciences 5, 89–97. URL: http://dx.doi.org/10.5194/adgeo-5-89-2005, doi:10.5194/adgeo-5-89-2005.

Krysanova, V., Donnelly, C., Gelfan, A., Gerten, D., Arheimer, B., Hattermann, F., Kundzewicz, Z.W., 2018. How the performance of hydrological models relates to credibility of projections under climate change. Hydrological Sciences Journal 63, 696–720. URL: https://www.tandfonline.com/doi/full/10.1080/02626667.2018.1446214, doi:10.1080/02626667.2018.1446214. number: 5.

Krysanova, V., Hattermann, F.F., Kundzewicz, Z.W., 2020. How evaluation of hydrological models influences results of climate impact assessment—an editorial 163, 1121–1141. URL: https://doi.org/10.1007/s10584-020-02927-8, doi:10.1007/s10584-020-02927-8.

Kundzewicz, Z., Krysanova, V., Benestad, R., Hov, Ø., Piniewski, M., Otto, I., 2018. Uncertainty in climate change impacts on water resources. Environmental Science & Policy 79, 1–8. URL: https://doi.org/10.1016/j.envsci.2017.10.008, doi:10.1016/j.envsci.2017.10.008.

Kutuzov, S., Shahgedanova, M., 2009. Glacier retreat and climatic variability in the eastern terskey–alatoo, inner tien shan between the middle of the 19th century and beginning of the 21st century. Global and Planetary Change 69, 59–70. URL: https://doi.org/10.1016/j.gloplacha.2009.07.001, doi:10.1016/j.gloplacha.2009.07.001.

Lalande, M., Ménégoz, M., Krinner, G., Naegeli, K., Wunderle, S., 2021. Climate change in the high mountain asia in CMIP6. Earth System Dynamics 12, 1061–1098. URL: https://doi.org/10.5194/esd-12-1061-2021, doi:10.5194/esd-12-1061-2021.

Lange, S., 2018. Bias correction of surface downwelling longwave and shortwave radiation for the EWEMBI dataset. Earth System Dynamics 9, 627–645. URL: https://doi.org/10.5194/esd-9-627-2018, doi:10.5194/esd-9-627-2018.

Lange, S., 2019. Trend-preserving bias adjustment and statistical downscaling with ISIMIP3basd (v1.0). Geoscientific Model Development 12, 3055–3070. URL: https://doi.org/10.5194/gmd-12-3055-2019, doi:10.5194/gmd-12-3055-2019.

Lange, S., Menz, C., Gleixner, S., Cucchi, M., Weedon, G.P., Amici, A., Bellouin, N., Müller Schmied, H., Hersbach, H., Buontempo, C., Cagnazzo, C., 2021. Wfde5 over land merged with era5 over the ocean (w5e5 v2.0). URL: https://data.isimip.org/10.48364/ISIMIP.342217, doi:10.48364/ISIMIP.342217.

Le Moine, N., 2008. Le bassin versant de surface vu par le souterrain: une voie d'amélioration des performances et du réalisme des modèles pluie-débit? Ph.D. thesis. Doctorat Géosciences et Ressources Naturelles, Université Pierre et Marie ....

Legates, D.R., McCabe, G.J., 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resources Research 35, 233–241. URL: http://dx.doi.org/10.1029/1998WR900018, doi:10.1029/1998wr900018.

Li, T., Shi, J., Deng, W., Hu, Z., 2022. Pyramid particle swarm optimization with novel strategies of competition and cooperation. Applied Soft Computing 121, 108731. URL: http://dx.doi.org/10.1016/j.asoc.2022.108731, doi:10.1016/j.asoc.2022.108731.

Li, Z., Chen, Y., Li, W., Deng, H., Fang, G., 2015. Potential impacts of climate change on vegetation dynamics in central asia. Journal of Geophysical Research: Atmospheres 120, 12345–12356. URL: https://doi.org/10.1002/2015jd023618, doi:10.1002/2015jd023618.

Liersch, S., Drews, M., Pilz, T., Salack, S., Sietz, D., Aich, V., A D Larsen, M., Gädeke, A., Halsnæ s, K., Thiery, W., Huang, S., Lobanova, A., Koch, H., Hattermann, F.F., 2020. One simulation, different conclusions—the baseline period makes the difference! Environmental Research Letters 15, 104014. URL: http://dx.doi.org/10.1088/1748-9326/aba3d7, doi:10.1088/1748-9326/aba3d7.

Lin, F., Chen, X., Yao, H., 2017. Evaluating the use of nash-sutcliffe efficiency coefficient in goodness-of-fit measures for daily runoff simulation with swat. Journal of Hydrologic Engineering 22. URL: http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0001580, doi:10.1061/(asce)he.1943-5584.0001580.

Linsbauer, A., Paul, F., Haeberli, W., 2012. Modeling glacier thickness distribution and bed topography over entire mountain ranges with GlabTop: Application of a fast and robust approach. Journal of

Geophysical Research: Earth Surface 117, n/a–n/a. URL: https://doi.org/10.1029/2011jf002313, doi:10.1029/2011jf002313.

Linsbauer, A., Paul, F., Hoelzle, M., Frey, H., Haeberli, W., 2009. The swiss alps without glaciers ? a gis-based modelling approach for reconstruction of glacier beds, in: Purves, R.S., et al (Eds.), Proceedings of Geomorphometry 2009. Department of Geography, University of Zurich, Zurich, CH, pp. 243–247. URL: https://doi.org/10.5167/uzh-27834. conference name: Geomorphometry 2009 International Conference, 31 August - 2 September 2009, Zurich, Switzerland.

Luo, M., Liu, T., Meng, F., Duan, Y., Bao, A., Frankl, A., Maeyer, P.D., 2018. Spatiotemporal characteristics of future changes in precipitation and temperature in central asia. International Journal of Climatology 39, 1571–1588. URL: https://doi.org/10.1002/joc.5901, doi:10.1002/joc.5901.

Luo, Y., Arnold, J., Liu, S., Wang, X., Chen, X., 2013. Inclusion of glacier processes for distributed hydrological modeling at basin scale with application to a watershed in tianshan mountains, northwest china. Journal of Hydrology 477, 72–85. URL: https://doi.org/10.1016/j.jhydrol.2012.11.005, doi:10.1016/j.jhydrol.2012.11.005.

Lutz, A.F., Immerzeel, W.W., Shrestha, A.B., Bierkens, M.F.P., 2014. Consistent increase in high asia's runoff due to increasing glacier melt and precipitation. Nature Climate Change 4, 587–592. URL: https://doi.org/10.1038/nclimate2237, doi:10.1038/nclimate2237.

Ma, H., Zhong, L., Fu, Y., Cheng, M., Wang, X., Cheng, M., Chang, Y., 2023. A study on hydrological responses of the fuhe river basin to combined effects of land use and climate change. Journal of Hydrology: Regional Studies 48, 101476. URL: http://dx.doi.org/10.1016/j.ejrh.2023.101476, doi:10.1016/j.ejrh.2023.101476.

Ma, X., Zhu, J., Yan, W., Zhao, C., 2021. Projections of desertification trends in central asia under global warming scenarios. Science of The Total Environment 781, 146777. URL: https://doi.org/10.1016/j.scitotenv.2021.146777, doi:10.1016/j.scitotenv.2021.146777.

Magnusson, J., Gustafsson, D., Hüsler, F., Jonas, T., 2014. Assimilation of point swe data into a distributed snow cover model comparing two contrasting methods. Water Resources Research 50, 7816–7835. URL: http://dx.doi.org/10.1002/2014WR015302, doi:10.1002/2014wr015302.

Manabe, S., Bryan, K., 1969. Climate calculations with a combined ocean-atmosphere model. Journal of the Atmospheric Sciences 26, 786–789. URL: http://dx.doi.org/10.1175/1520-0469(1969)026<0786:CCWACO>2.0.CO;2, doi:10.1175/1520-0469(1969)026<0786:ccwaco>2.0.co;2.

Mango, L.M., Melesse, A.M., McClain, M.E., Gann, D., Setegn, S.G., 2011. Land use and climate change impacts on the hydrology of the upper mara river basin, kenya: results of a modeling study to support better resource management. Hydrology and Earth System Sciences 15, 2245–2258. URL: http://dx.doi.org/10.5194/hess-15-2245-2011, doi:10.5194/hess-15-2245-2011.

Marahatta, S., Devkota, L.P., Aryal, D., 2021. Application of SWAT in hydrological simulation of complex mountainous river basin (part i: Model development). Water 13, 1546. URL: https://doi.org/10.3390/w13111546, doi:10.3390/w13111546.

Marhaento, H., Booij, M.J., Hoekstra, A.Y., 2018. Hydrological response to future land-use change and climate change in a tropical catchment. Hydrological Sciences Journal 63, 1368–1385. URL: http://dx.doi.org/10.1080/02626667.2018.1511054, doi:10.1080/02626667.2018.1511054.

Martens, B., Miralles, D.G., Lievens, H., van der Schalie, R., de Jeu, R.A.M., Fernández-Prieto, D., Beck, H.E., Dorigo, W.A., Verhoest, N.E.C., 2017. GLEAM v3: satellite-based land evaporation and root-zone soil moisture. Geoscientific Model Development 10, 1903–1925. URL: https://doi.org/10.5194/gmd-10-1903-2017, doi:10.5194/gmd-10-1903-2017.

Martinec, J., Rango, A., 1989. Merits of statistical criteria for the performance of hydrological models1. JAWRA Journal of the American Water Resources Association 25, 421–432. URL: http://dx.doi.org/10.1111/j.1752-1688.1989.tb03079.x, doi:10.1111/j.1752-1688.1989.tb03079.x.

Mathevet, T., Le Moine, N., Andréassian, V., Gupta, H., Oudin, L., 2024. Multi-objective assessment of hydrological model performances using nash–sutcliffe and kling–gupta efficiencies on a worldwide large sample of watersheds. Comptes Rendus. Géoscience 355, 117–141. URL: http://dx.doi.org/10.5802/crgeos.189, doi:10.5802/crgeos.189.

McCuen, R.H., Knight, Z., Cutter, A.G., 2006. Evaluation of the nash–sutcliffe efficiency index. Journal of Hydrologic Engineering 11, 597–602. URL: http://dx.doi.org/10.1061/(ASCE)1084-0699(2006)11:6(597), doi:10.1061/(asce)1084-0699(2006)11:6(597).

McGuire, K.J., Klaus, J., Jackson, C.R., 2024. Interflow, subsurface stormflow and throughflow: A synthesis of field work and modelling. Hydrological Processes 38. URL: http://dx.doi.org/10.1002/hyp.15263, doi:10.1002/hyp.15263.

McNeil, C.J., Sass, L., Florentine, C., Baker, E.H., Peitzsch, E.H., Whorton, E.N., Miller, Z., Fagre, D.B., Clark, A.M., O'Neel, S.R., 2016. Glacier-wide mass balance and compiled data inputs: USGS benchmark glaciers. URL: https://alaska.usgs.gov/products/data.php?dataid=79, doi:10.5066/F7HD7SRF.

Meehl, G.A., Senior, C.A., Eyring, V., Flato, G., Lamarque, J.F., Stouffer, R.J., Taylor, K.E., Schlund, M., 2020. Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 earth system models. Science Advances 6. URL: https://doi.org/10.1126/sciadv.aba1981, doi:10.1126/sciadv.aba1981.

Meinshausen, M., Nicholls, Z.R.J., Lewis, J., Gidden, M.J., Vogel, E., Freund, M., Beyerle, U., Gessner, C., Nauels, A., Bauer, N., Canadell, J.G., Daniel, J.S., John, A., Krummel, P.B., Luderer, G., Meinshausen, N., Montzka, S.A., Rayner, P.J., Reimann, S., Smith, S.J., van den Berg, M., Velders, G.J.M., Vollmer, M.K., Wang, R.H.J., 2020. The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500. Geoscientific Model Development 13, 3571–3605. URL: https://doi.org/10.5194/gmd-13-3571-2020, doi:10.5194/gmd-13-3571-2020.

Mengel, M., Treu, S., Lange, S., Frieler, K., 2021. Attrici v1.1 – counterfactual climate for impact attribution. Geoscientific Model Development 14, 5269–5284. URL: http://dx.doi.org/10.5194/GMD-14-5269-2021, doi:10.5194/gmd-14-5269-2021.

Merchán-Rivera, P., Geist, A., Disse, M., Huang, J., Chiogna, G., 2022. A bayesian framework to assess and create risk maps of groundwater flooding. Journal of Hydrology 610, 127797. URL: http://dx.doi.org/10.1016/j.jhydrol.2022.127797, doi:10.1016/j.jhydrol.2022.127797.

Millan, R., Mouginot, J., Rabatel, A., Morlighem, M., 2022. Ice velocity and thickness of the world's glaciers. Nature Geoscience 15, 124–129. URL: https://doi.org/10.1038/s41561-021-00885-z, doi:10.1038/s41561-021-00885-z.

Miralles, D.G., Holmes, T.R.H., Jeu, R.A.M.D., Gash, J.H., Meesters, A.G.C.A., Dolman, A.J., 2011. Global land-surface evaporation estimated from satellite-based observations. Hydrology and Earth System Sciences 15, 453–469. URL: https://doi.org/10.5194/hess-15-453-2011, doi:10.5194/hess-15-453-2011.

Mishra, V., Shah, H., López, M.R.R., Lobanova, A., Krysanova, V., 2020. Does comprehensive evaluation of hydrological models influence projected changes of mean and high flows in the godavari river basin? Climatic Change 163, 1187–1205. URL: https://doi.org/10.1007/s10584-020-02847-7, doi:10.1007/s10584-020-02847-7.

Moriasi, D.N., Arnold, J.G., Liew, M.W.V., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE 50, 885–900. URL: http://dx.doi.org/10.13031/2013.23153, doi:10.13031/2013.23153.

Morris, M.D., 1991. Factorial sampling plans for preliminary computational experiments. Technometrics 33, 161–174. URL: https://doi.org/10.1080/00401706.1991.10484804, doi:10.1080/00401706.1991.10484804.

Muhammad, S., Tian, L., Ali, S., Latif, Y., Wazir, M.A., Goheer, M.A., Saifullah, M., Hussain, I., Shiyin, L., 2020. Thin debris layers do not enhance melting of the karakoram glaciers. Science of The Total Environment 746, 141119. URL: https://doi.org/10.1016/j.scitotenv.2020.141119, doi:10.1016/j.scitotenv.2020.141119.

Munro, D.S., 2005. A revised canadian perspective: progress in glacier hydrology. Hydrological Processes 19, 231–245. URL: https://doi.org/10.1002/hyp.5769, doi:10.1002/hyp.5769.

Murphy, A.H., 1988. Skill scores based on the mean square error and their relationships to the correlation coefficient. Monthly Weather Review 116, 2417–2424. URL: http://dx.doi.org/10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2, doi:10.1175/1520-0493(1988)116<2417:ssbotm>2.0.co;2.

NASA JPL, 2013a. Nasa shuttle radar topography mission global 1 arc second. URL: https://lpdaac.usgs.gov/products/srtmgl1v003/, doi:10.5067/MEASURES/SRTM/SRTMGL1.003.

NASA JPL, 2013b. Nasa shuttle radar topography mission global 3 arc second. URL: https://lpdaac.usgs.gov/products/srtmgl3v003/, doi:10.5067/MEASURES/SRTM/SRTMGL3.003.

Nash, J., Sutcliffe, J., 1970. River flow forecasting through conceptual models part i — a discussion of principles. Journal of Hydrology 10, 282–290. URL: http://dx.doi.org/10.1016/0022-1694(70)90255-6, doi:10.1016/0022-1694(70)90255-6.

Nature, 2024. Science must protect thinking time in a world of instant communication. Nature 631, 709–709. URL: http://dx.doi.org/10.1038/d41586-024-02381-x, doi:10.1038/d41586-024-02381-x.

Naz, B.S., Frans, C.D., Clarke, G.K.C., Burns, P., Lettenmaier, D.P., 2014. Modeling the effect of glacier recession on streamflow response using a coupled glacio-hydrological model. Hydrology and Earth System Sciences 18, 787–802. URL: https://doi.org/10.5194/hess-18-787-2014, doi:10.5194/hess-18-787-2014.

Nazari-Sharabian, M., Taheriyoun, M., Ahmad, S., Karakouzian, M., Ahmadi, A., 2019. Water quality modeling of mahabad dam watershed–reservoir system under climate change conditions, using SWAT and system dynamics. Water 11, 394. URL: https://doi.org/10.3390/w11020394, doi:10.3390/w11020394.

Nearing, G., Cohen, D., Dube, V., Gauch, M., Gilon, O., Harrigan, S., Hassidim, A., Klotz, D., Kratzert, F., Metzger, A., Nevo, S., Pappenberger, F., Prudhomme, C., Shalev, G., Shenzis, S., Tekalign, T.Y., Weitzner, D., Matias, Y., 2024. Global prediction of extreme floods in ungauged watersheds. Nature 627, 559–563. URL: http://dx.doi.org/10.1038/s41586-024-07145-1, doi:10.1038/s41586-024-07145-1.

Nearing, G.S., Kratzert, F., Sampson, A.K., Pelissier, C.S., Klotz, D., Frame, J.M., Prieto, C., Gupta, H.V., 2021. What role does hydrological science play in the age of machine learning? Water Resources Research 57. URL: http://dx.doi.org/10.1029/2020WR028091, doi:10.1029/2020wr028091.

NOAA National Centers for Environmental Information, 2024. Monthly global climate report for annual 2023. Accessed Feb 5, 2024. https://www.ncei.noaa.gov/access/monitoring/monthly-report/global/202313.

de Oliveira Serrão, E.A., Silva, M.T., Ferreira, T.R., de Ataide, L.C.P., dos Santos, C.A., de Lima, A.M.M., de Paulo Rodrigues da Silva, V., de Assis Salviano de Sousa, F., Gomes, D.J.C., 2022. Impacts of land use and land cover changes on hydrological processes and sediment yield determined using the SWAT model. International Journal of Sediment Research 37, 54–69. URL: https://doi.org/10.1016/j.ijsrc.2021.04.002, doi:10.1016/j.ijsrc.2021.04.002.

Olsson, J., Arheimer, B., Borris, M., Donnelly, C., Foster, K., Nikulin, G., Persson, M., Perttu, A.M., Uvo, C., Viklander, M., Yang, W., 2016. Hydrological climate change impact assessment at small and large scales: Key messages from recent progress in sweden. Climate 4, 39. URL: https://doi.org/10.3390/cli4030039, doi:10.3390/cli4030039.

Omani, N., Srinivasan, R., Karthikeyan, R., Smith, P., 2017. Hydrological modeling of highly glacierized basins (andes, alps, and central asia). Water 9, 111. URL: https://doi.org/10.3390/w9020111, doi:10.3390/w9020111.

O'Neel, S., Hood, E., Arendt, A., Sass, L., 2014. Assessing streamflow sensitivity to variations in glacier mass balance. Climatic Change 123, 329–341. URL: https://doi.org/10.1007/s10584-013-1042-7, doi:10.1007/s10584-013-1042-7.

O'Neel, S., McNeil, C., Sass, L.C., Florentine, C., Baker, E.H., Peitzsch, E., McGrath, D., Fountain, A.G., Fagre, D., 2019. Reanalysis of the US geological survey benchmark glaciers: long-term insight into climate forcing of glacier mass balance. Journal of Glaciology 65, 850–866. URL: https://doi.org/10.1017/jog.2019.66, doi:10.1017/jog.2019.66.

Oo, H.T., Zin, W.W., Thin Kyi, C.C., 2020. Analysis of streamflow response to changing climate conditions using swat model. Civil Engineering Journal 6, 194–209. URL: http://dx.doi.org/10.28991/cej-2020-03091464, doi:10.28991/cej-2020-03091464.

Oudin, L., Andréassian, V., Mathevet, T., Perrin, C., Michel, C., 2006. Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations. Water Resources Research 42. URL: http://dx.doi.org/10.1029/2005WR004636, doi:10.1029/2005wr004636.

Ozturk, T., Turp, M.T., Türkeş, M., Kurnaz, M.L., 2017. Projected changes in temperature and precipitation climatology of central asia CORDEX region 8 by using RegCM4.3.5. Atmospheric Research 183, 296–307. URL: https://doi.org/10.1016/j.atmosres.2016.09.008, doi:10.1016/j.atmosres.2016.09.008.

Pellicciotti, F., Brock, B., Strasser, U., Burlando, P., Funk, M., Corripio, J., 2005. An enhanced temperature-index glacier melt model including the shortwave radiation balance: development and testing for haut glacier d'arolla, switzerland. Journal of Glaciology 51, 573–587. URL: http://dx.doi.org/10.3189/172756505781829124, doi:10.3189/172756505781829124.

Peña-Ramos, J.A., Bagus, P., Fursova, D., 2021. Water conflicts in central asia: Some recommendations on the non-conflictual use of water. Sustainability 13, 3479. URL: https://doi.org/10.3390/su13063479, doi:10.3390/su13063479.

Peng, Q., Wang, R., Jiang, Y., Li, C., Guo, W., 2021. The change of hydrological variables and its effects on vegetation in central asia. Theoretical and Applied Climatology 146, 741–753. URL: https://doi.org/10.1007/s00704-021-03730-w, doi:10.1007/s00704-021-03730-w.

Pettitt, A.N., 1979. A non-parametric approach to the change-point problem. Applied Statistics 28, 126. URL: http://dx.doi.org/10.2307/2346729, doi:10.2307/2346729.

Phillips, N.A., 1956. The general circulation of the atmosphere: A numerical experiment. Q. J. R. Meteorol. Soc. 82, 123–164.

Pianosi, F., Beven, K., Freer, J., Hall, J.W., Rougier, J., Stephenson, D.B., Wagener, T., 2016. Sensitivity analysis of environmental models: A systematic review with practical workflow. Environmental Modelling &amp Software 79, 214–232. URL: https://doi.org/10.1016/j.envsoft.2016.02.008, doi:10.1016/j.envsoft.2016.02.008.

Pizarro, A., Jorquera, J., 2024. Advancing objective functions in hydrological modelling: Integrating knowable moments for improved simulation accuracy. Journal of Hydrology 634, 131071. URL: http://dx.doi.org/10.1016/j.jhydrol.2024.131071, doi:10.1016/j.jhydrol.2024.131071.

Poméon, T., Diekkrüger, B., Springer, A., Kusche, J., Eicker, A., 2018. Multi-objective validation of swat for sparsely-gauged west african river basins—a remote sensing approach. Water 10. URL: https://www.mdpi.com/2073-4441/10/4/451, doi:10.3390/w10040451.

Pool, S., Vis, M., Seibert, J., 2018. Evaluating model performance: towards a non-parametric variant of the kling-gupta efficiency. Hydrological Sciences Journal 63, 1941–1953. URL: http://dx.doi.org/10.1080/02626667.2018.1552002, doi:10.1080/02626667.2018.1552002.

Pradhananga, N.S., Kayastha, R.B., Bhattarai, B.C., Adhikari, T.R., Pradhan, S.C., Devkota, L.P., Shrestha, A.B., Mool, P.K., 2014. Estimation of discharge from langtang river basin, rasuwa, nepal, using a glacio-hydrological model. Annals of Glaciology 55, 223–230. URL: https://doi.org/10.3189/2014aog66a123, doi:10.3189/2014aog66a123.

Priestley, C.H.B., Taylor, R.J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Monthly Weather Review 100, 81–92. URL: http://dx.doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2, doi:10.1175/1520-0493(1972)100<0081:otaosh>2.3.co;2.

Pushpalatha, R., Perrin, C., Moine, N.L., Andréassian, V., 2012. A review of efficiency criteria suitable for evaluating low-flow simulations. Journal of Hydrology 420–421, 171–182. URL: http://dx.doi.org/10.1016/j.jhydrol.2011.11.055, doi:10.1016/j.jhydrol.2011.11.055.

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. URL: https://doi.org/10.1016/j.scitotenv.2020.137488, doi:10.1016/j.scitotenv.2020.137488.

Qi, J., Zhang, X., Yang, Q., Srinivasan, R., Arnold, J.G., Li, J., Waldholf, S.T., Cole, J., 2020. SWAT ungauged: Water quality modeling in the upper mississippi river basin. Journal of Hydrology 584, 124601. URL: https://doi.org/10.1016/j.jhydrol.2020.124601, doi:10.1016/j.jhydrol.2020.124601.

Rahman, K., Maringanti, C., Beniston, M., Widmer, F., Abbaspour, K., Lehmann, A., 2012. Streamflow modeling in a highly managed mountainous glacier watershed using SWAT: The upper rhone river watershed case in switzerland. Water Resources Management 27, 323–339. URL: https://doi.org/10.1007/s11269-012-0188-9, doi:10.1007/s11269-012-0188-9.

Rajib, M.A., Merwade, V., Yu, Z., 2016. Multi-objective calibration of a hydrologic model using spatially distributed remotely sensed/in-situ soil moisture 536, 192–207. URL: https://doi.org/10.1016/j.jhydrol.2016.02.037, doi:10.1016/j.jhydrol.2016.02.037.

Rasouli, K., Pomeroy, J.W., Whitfield, P.H., 2019a. Are the effects of vegetation and soil changes as important as climate change impacts on hydrological processes? Hydrology and Earth System Sciences 23, 4933–4954. URL: https://doi.org/10.5194/hess-23-4933-2019, doi:10.5194/hess-23-4933-2019.

Rasouli, K., Pomeroy, J.W., Whitfield, P.H., 2019b. Hydrological responses of headwater basins to monthly perturbed climate in the north american cordillera. Journal of Hydrometeorology 20, 863–882. URL: https://doi.org/10.1175/jhm-d-18-0166.1, doi:10.1175/jhm-d-18-0166.1.

Reyer, C.P., Otto, I.M., Adams, S., Albrecht, T., Baarsch, F., Cartsburg, M., Coumou, D., Eden, A., Ludi, E., Marcus, R., Mengel, M., Mosello, B., Robinson, A., Schleussner, C.F., Serdeczny, O., Stagl, J., 2015. Climate change impacts in central asia and their implications for development. Regional Environmental Change 17, 1639–1650. URL: https://doi.org/10.1007/s10113-015-0893-z, doi:10.1007/s10113-015-0893-z.

RGI Consortium, 2017. Randolph glacier inventory - a dataset of global glacier outlines, version 6. URL: https://nsidc.org/data/nsidc-0770/versions/6, doi:10.7265/4M1F-GD79.

Richardson, K., Steffen, W., Lucht, W., Bendtsen, J., Cornell, S.E., Donges, J.F., Drüke, M., Fetzer, I., Bala, G., von Bloh, W., Feulner, G., Fiedler, S., Gerten, D., Gleeson, T., Hofmann, M., Huiskamp, W., Kummu, M., Mohan, C., Nogués-Bravo, D., Petri, S., Porkka, M., Rahmstorf, S., Schaphoff, S., Thonicke, K., Tobian, A., Virkki, V., Wang-Erlandsson, L., Weber, L., Rockström, J., 2023. Earth beyond six of nine planetary boundaries. Science Advances 9. URL: http://dx.doi.org/10.1126/sciadv.adh2458, doi:10.1126/sciadv.adh2458.

Richardson, L.F., 1922. Weather Prediction by Numerical Process. Cambridge University Press.

Ritchie, H., Roser, M., 2018. Water use and stress. Our World in Data URL: https://ourworldindata.org/water-use-stress.

Rodell, M., Houser, P.R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J.K., Walker, J.P., Lohmann, D., Toll, D., 2004. The global land data assimilation system. Bulletin of the American Meteorological Society 85, 381–394. URL: https://doi.org/10.1175/bams-85-3-381, doi:10.1175/bams-85-3-381.

Rounce, D.R., Hock, R., Maussion, F., Hugonnet, R., Kochtitzky, W., Huss, M., Berthier, E., Brinkerhoff, D., Compagno, L., Copland, L., Farinotti, D., Menounos, B., McNabb, R.W., 2023. Global glacier change in the 21st century: Every increase in temperature matters. Science 379, 78–83. URL: http://dx.doi.org/10.1126/science.abo1324, doi:10.1126/science.abo1324.

Saks, T., Pohl, E., Machguth, H., Dehecq, A., Barandun, M., Kenzhebaev, R., Kalashnikova, O., Hoelzle, M., 2022. Glacier runoff variation since 1981 in the upper naryn river catchments, central tien shan. Frontiers in Environmental Science 9. URL: https://doi.org/10.3389/fenvs.2021.780466, doi:10.3389/fenvs.2021.780466.

Saltelli, A., Gigerenzer, G., Hulme, M., Katsikopoulos, K.V., Melsen, L.A., Peters, G.P., Pielke, R., Robertson, S., Stirling, A., Tavoni, M., Puy, A., 2024. Bring digital twins back to earth. WIREs Climate Change URL: http://dx.doi.org/10.1002/wcc.915, doi:10.1002/wcc.915.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. Global sensitivity analysis: the primer. John Wiley & Sons.

Samavati, A., Babamiri, O., Rezai, Y., Heidarimozaffar, M., 2022. Investigating the effects of climate change on future hydrological drought in mountainous basins using swat model based on cmip5 model. Stochastic Environmental Research and Risk Assessment 37, 849–875. URL: http://dx.doi.org/10.1007/s00477-022-02319-7, doi:10.1007/s00477-022-02319-7.

Sarrazin, F., Pianosi, F., Wagener, T., 2016. Global sensitivity analysis of environmental models: Convergence and validation. Environmental Modelling &amp Software 79, 135–152. URL: https://doi.org/10.1016/j.envsoft.2016.02.005, doi:10.1016/j.envsoft.2016.02.005.

Schaefli, B., Gupta, H.V., 2007. Do nash values have value? Hydrological Processes 21, 2075–2080. URL: http://dx.doi.org/10.1002/hyp.6825, doi:10.1002/hyp.6825.

Schaefli, B., Hingray, B., Niggli, M., Musy, A., 2005. A conceptual glacio-hydrological model for high mountainous catchments. Hydrology and Earth System Sciences 9, 95–109. URL: http://dx.doi.org/10.5194/hess-9-95-2005, doi:10.5194/hess-9-95-2005.

Schaefli, B., Huss, M., 2011. Integrating point glacier mass balance observations into hydrologic model identification. Hydrology and Earth System Sciences 15, 1227–1241. URL: http://dx.doi.org/10.5194/hess-15-1227-2011, doi:10.5194/hess-15-1227-2011.

Schaffhauser, T., 2023. SWAT-G demo model martelltal [Data set]. URL: https://doi.org/10.5281/zenodo.8068725, doi:10.5281/zenodo.8068725.

Schaffhauser, T., Hofmeister, F., Chiogna, G., Merk, F., Tuo, Y., Machnitzke, J., Alcamo, L., Huang, J., Disse, M., 2024a. Merits and limits of swat-gl: Application in contrasting glaciated catchments. Hydrology and Earth System Sciences Discussions 2024, 1–40. URL: https://hess.copernicus.org/preprints/hess-2024-89/, doi:10.5194/hess-2024-89.

Schaffhauser, T., Lange, S., Tuo, Y., Disse, M., 2023. Shifted discharge and drier soils: Hydrological projections for a central asian catchment. Journal of Hydrology: Regional Studies 46, 101338. URL: https://www.sciencedirect.com/science/article/pii/S2214581823000253, doi:https://doi.org/10.1016/j.ejrh.2023.101338.

Schaffhauser, T., Tuo, Y., Hofmeister, F., Chiogna, G., Huang, J., Merk, F., Disse, M., 2024b. Swat-gl: A new glacier routine for the hydrological model swat. JAWRA Journal of the American Water Resources Association URL: http://dx.doi.org/10.1111/1752-1688.13199, doi:10.1111/1752-1688.13199.

Schleussner, C.F., Lissner, T.K., Fischer, E.M., Wohland, J., Perrette, M., Golly, A., Rogelj, J., Childers, K., Schewe, J., Frieler, K., Mengel, M., Hare, W., Schaeffer, M., 2016. Differential climate impacts for policy-relevant limits to global warming: the case of 1.5°c and 2°c. Earth System Dynamics 7, 327–351. URL: http://dx.doi.org/10.5194/esd-7-327-2016, doi:10.5194/esd-7-327-2016.

Schürz, C., Hollosi, B., Matulla, C., Pressl, A., Ertl, T., Schulz, K., Mehdi, B., 2019. A comprehensive sensitivity and uncertainty analysis for discharge and nitrate-nitrogen loads involving multiple discrete model inputs under future changing conditions. Hydrology and Earth System Sciences 23, 1211–1244. URL: https://doi.org/10.5194/hess-23-1211-2019, doi:10.5194/hess-23-1211-2019.

Schwemmle, R., Demand, D., Weiler, M., 2021. Technical note: Diagnostic efficiency – specific evaluation of model performance. Hydrology and Earth System Sciences 25, 2187–2198. URL: http://dx.doi.org/10.5194/hess-25-2187-2021, doi:10.5194/hess-25-2187-2021.

Seibert, J., Vis, M.J.P., Kohn, I., Weiler, M., Stahl, K., 2018a. Technical note: Representing glacier geometry changes in a semi-distributed hydrological model. Hydrology and Earth System Sciences 22, 2211–2224. URL: https://doi.org/10.5194/hess-22-2211-2018, doi:10.5194/hess-22-2211-2018.

Seibert, J., Vis, M.J.P., Lewis, E., van Meerveld, H., 2018b. Upper and lower benchmarks in hydrological modelling. Hydrological Processes 32, 1120–1125. URL: http://dx.doi.org/10.1002/hyp.11476, doi:10.1002/hyp.11476.

Senent-Aparicio, J., Pérez-Sánchez, J., Carrillo-García, J., Soto, J., 2017. Using swat and fuzzy topsis to assess the impact of climate change in the headwaters of the segura river basin (se spain). Water 9, 149. URL: http://dx.doi.org/10.3390/w9020149, doi:10.3390/w9020149.

Shafeeque, M., Luo, Y., Wang, X., Sun, L., 2019. Altitudinal distribution of meltwater and its effects on glacio-hydrology in glacierized catchments, central asia. JAWRA Journal of the American Water Resources Association 56, 30–52. URL: https://doi.org/10.1111/1752-1688.12805, doi:10.1111/1752-1688.12805.

Shannon, S., Payne, A., Freer, J., Coxon, G., Kauzlaric, M., Kriegel, D., Harrison, S., 2023. A snow and glacier hydrological model for large catchments – case study for the naryn river, central asia. Hydrology and Earth System Sciences 27, 453–480. URL: http://dx.doi.org/10.5194/hess-27-453-2023, doi:10.5194/hess-27-453-2023.

Shannon, S.R., Payne, A., Freer, J., Coxon, G., Kauzlaric, M., Kriegel, D., Harrison, S., 2022. A snow and glacier hydrological model for large catchments – case study for the naryn river, central asia URL: https://doi.org/10.5194/hess-2022-51, doi:10.5194/hess-2022-51.

Sharannya, T.M., Mudbhatkal, A., Mahesha, A., 2018. Assessing climate change impacts on river hydrology – a case study in the western ghats of india. Journal of Earth System Science 127. URL: http://dx.doi.org/10.1007/s12040-018-0979-3, doi:10.1007/s12040-018-0979-3.

Sherwood, S.C., Webb, M.J., Annan, J.D., Armour, K.C., Forster, P.M., Hargreaves, J.C., Hegerl, G., Klein, S.A., Marvel, K.D., Rohling, E.J., Watanabe, M., Andrews, T., Braconnot, P., Bretherton, C.S., Foster, G.L., Hausfather, Z., Heydt, A.S., Knutti, R., Mauritsen, T., Norris, J.R., Proistosescu, C., Rugenstein, M., Schmidt, G.A., Tokarska, K.B., Zelinka, M.D., 2020. An assessment of earth's climate sensitivity using multiple lines of evidence. Reviews of Geophysics 58. URL: https://doi.org/10.1029/2019rg000678, doi:10.1029/2019rg000678.

Shukla, S., Jain, S.K., Kansal, M.L., 2021. Hydrological modelling of a snow/glacier-fed western himalayan basin to simulate the current and future streamflows under changing climate scenarios. Science of The Total Environment 795, 148871. URL: https://doi.org/10.1016/j.scitotenv.2021.148871, doi:10.1016/j.scitotenv.2021.148871.

Sin, G., Gernaey, K.V., 2009. Improving the morris method for sensitivity analysis by scaling the elementary effects, in: Computer Aided Chemical Engineering. Elsevier, pp. 925–930. URL: https://doi.org/10.1016/s1570-7946(09)70154-3, doi:10.1016/s1570-7946(09)70154-3.

Singh, P., Kumar, N., Arora, M., 2000. Degree–day factors for snow and ice for dokriani glacier, garhwal himalayas. Journal of Hydrology 235, 1–11. URL: https://doi.org/10.1016/s0022-1694(00)00249-3, doi:10.1016/s0022-1694(00)00249-3.

Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., Xu, C., 2015. Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications. Journal of Hydrology 523, 739–757. URL: http://dx.doi.org/10.1016/j.jhydrol.2015.02.013, doi:10.1016/j.jhydrol.2015.02.013.

Stephens, C., Marshall, L., Johnson, F., Lin, L., Band, L., Ajami, H., 2020. Is past variability a suitable proxy for future change? a virtual catchment experiment. Water Resources Research 56. URL: https://doi.org/10.1029/2019wr026275, doi:10.1029/2019wr026275.

Stoll, E., Hanzer, F., Oesterle, F., Nemec, J., Schöber, J., Huttenlau, M., Förster, K., 2020. What can we learn from comparing glacio-hydrological models? Atmosphere 11, 981. URL: https://doi.org/10.3390/atmos11090981, doi:10.3390/atmos11090981.

Storn, R., Price, K., 1995. Differential evolution: A simple and efficient adaptive scheme for global optimization over continuous spaces. Journal of Global Optimization 23.

Su, B., Huang, J., Zeng, X., Gao, C., Jiang, T., 2016. Impacts of climate change on streamflow in the upper yangtze river basin. Climatic Change 141, 533–546. URL: https://doi.org/10.1007/s10584-016-1852-5, doi:10.1007/s10584-016-1852-5.

Swiss Academy of Sciences, 2023. Two catastrophic years obliterate 10% of swiss glacier volume. https://scnat.ch/en/uuid/i/b8d5798e-a75e-5a7d-a858-f7a6613524ed-Two_catastrophic_years_obliterate_10_of_Swiss_glacier_volume.

Tang, G., Clark, M.P., Papalexiou, S.M., 2021. Sc-earth: A station-based serially complete earth dataset from 1950 to 2019. Journal of Climate 34, 6493–6511. URL: http://dx.doi.org/10.1175/JCLI-D-21-0067.1, doi:10.1175/jcli-d-21-0067.1.

Tarasova, L., Gnann, S., Yang, S., Hartmann, A., Wagener, T., 2024. Catchment characterization: Current descriptors, knowledge gaps and future opportunities. Earth-Science Reviews 252, 104739. URL: http://dx.doi.org/10.1016/j.earscirev.2024.104739, doi:10.1016/j.earscirev.2024.104739.

Tebaldi, C., Debeire, K., Eyring, V., Fischer, E., Fyfe, J., Friedlingstein, P., Knutti, R., Lowe, J., O'Neill, B., Sanderson, B., van Vuuren, D., Riahi, K., Meinshausen, M., Nicholls, Z., Tokarska, K.B., Hurtt, G., Kriegler, E., Lamarque, J.F., Meehl, G., Moss, R., Bauer, S.E., Boucher, O., Brovkin, V., Byun, Y.H., Dix, M., Gualdi, S., Guo, H., John, J.G., Kharin, S., Kim, Y., Koshiro, T., Ma, L., Olivié, D., Panickal, S., Qiao, F., Rong, X., Rosenbloom, N., Schupfner, M., Séférian, R., Sellar, A., Semmler, T., Shi, X., Song, Z., Steger, C., Stouffer, R., Swart, N., Tachiiri, K., Tang, Q., Tatebe, H., Voldoire, A., Volodin, E., Wyser, K., Xin, X., Yang, S., Yu, Y., Ziehn, T., 2021. Climate model projections from the scenario model intercomparison project (scenariomip) of cmip6. Earth System Dynamics 12, 253–293. URL: http://dx.doi.org/10.5194/esd-12-253-2021, doi:10.5194/esd-12-253-2021.

van Tiel, M., Aubry-Wake, C., Somers, L., Andermann, C., Avanzi, F., Baraer, M., Chiogna, G., Daigre, C., Das, S., Drenkhan, F., Farinotti, D., Fyffe, C.L., de Graaf, I., Hanus, S., Immerzeel, W., Koch, F., McKenzie, J.M., Müller, T., Popp, A.L., Saidaliyeva, Z., Schaefli, B., Schilling, O.S., Teagai, K., Thornton, J.M., Yapiyev, V., 2024. Cryosphere–groundwater connectivity is a missing link in the mountain water cycle. Nature Water 2, 624–637. URL: http://dx.doi.org/10.1038/s44221-024-00277-8, doi:10.1038/s44221-024-00277-8.

Tiel, M., Stahl, K., Freudiger, D., Seibert, J., 2020. Glacio-hydrological model calibration and evaluation. WIREs Water 7. URL: https://doi.org/10.1002/wat2.1483, doi:10.1002/wat2.1483.

Tokarska, K.B., Stolpe, M.B., Sippel, S., Fischer, E.M., Smith, C.J., Lehner, F., Knutti, R., 2020. Past warming trend constrains future warming in CMIP6 models. Science Advances 6. URL: https://doi.org/10.1126/sciadv.aaz9549, doi:10.1126/sciadv.aaz9549.

Tuo, Y., Duan, Z., Disse, M., Chiogna, G., 2016. Evaluation of precipitation input for SWAT modeling in alpine catchment: A case study in the adige river basin (italy). Science of The Total Environment 573, 66–82. URL: https://doi.org/10.1016/j.scitotenv.2016.08.034, doi:10.1016/j.scitotenv.2016.08.034.

Tuo, Y., Marcolini, G., Disse, M., Chiogna, G., 2018. Calibration of snow parameters in SWAT: comparison of three approaches in the upper adige river basin (italy). Hydrological Sciences Journal 63, 657–678. URL: https://doi.org/10.1080/02626667.2018.1439172, doi:10.1080/02626667.2018.1439172.

Ullrich, A., Volk, M., 2009. Application of the soil and water assessment tool (SWAT) to predict the impact of alternative management practices on water quality and quantity. Agricultural Water Management 96, 1207–1217. URL: https://doi.org/10.1016/j.agwat.2009.03.010, doi:10.1016/j.agwat.2009.03.010.

Unger-Shayesteh, K., Vorogushyn, S., Farinotti, D., Gafurov, A., Duethmann, D., Mandychev, A., Merz, B., 2013. What do we know about past changes in the water cycle of central asian headwaters? a review. Global and Planetary Change 110, 4–25. URL: https://doi.org/10.1016/j.gloplacha.2013.02.004, doi:10.1016/j.gloplacha.2013.02.004.

United Nations Framework Convention on Climate Change, 2015. Paris agreement. https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement.

U.S. Geological Survey, 1994. Usgs water data for the nation. URL: https://waterdata.usgs.gov/nwis, doi:10.5066/F7P55KJN.

Vaze, J., Post, D., Chiew, F., Perraud, J.M., Viney, N., Teng, J., 2010. Climate non-stationarity – validity of calibrated rainfall–runoff models for use in climate change studies. Journal of Hydrology 394, 447–457. URL: https://doi.org/10.1016%2Fj.jhydrol.2010.09.018, doi:10.1016/j.jhydrol.2010.09.018.

Vetter, T., Huang, S., Aich, V., Yang, T., Wang, X., Krysanova, V., Hattermann, F., 2015. Multi-model climate impact assessment and intercomparison for three large-scale river basins on three continents 6, 17–43. URL: https://doi.org/10.5194/esd-6-17-2015, doi:10.5194/esd-6-17-2015.

Vetter, T., Reinhardt, J., Flörke, M., van Griensven, A., Hattermann, F., Huang, S., Koch, H., Pechlivanidis, I.G., Plötner, S., Seidou, O., Su, B., Vervoort, R.W., Krysanova, V., 2016. Evaluation of sources of uncertainty in projected hydrological changes under climate change in 12 large-scale river basins. Climatic Change 141, 419–433. URL: https://doi.org/10.1007/s10584-016-1794-y, doi:10.1007/s10584-016-1794-y.

Vicente-Serrano, S.M., Miralles, D.G., McDowell, N., Brodribb, T., Domínguez-Castro, F., Leung, R., Koppa, A., 2022. The uncertain role of rising atmospheric co2 on global plant transpiration. Earth-Science Reviews 230, 104055. URL: http://dx.doi.org/10.1016/j.earscirev.2022.104055, doi:10.1016/j.earscirev.2022.104055.

Vidal, J.P., Hingray, B., Magand, C., Sauquet, E., Ducharne, A., 2016. Hierarchy of climate and hydrological uncertainties in transient low-flow projections. Hydrology and Earth System Sciences 20, 3651–3672. URL: https://doi.org/10.5194/hess-20-3651-2016, doi:10.5194/hess-20-3651-2016.

Villani, L., Castelli, G., Yimer, E.A., Chawanda, C.J., Nkwasa, A., Van Schaeybroeck, B., Penna, D., van Griensven, A., Bresci, E., 2024. Impacts of climate change and vegetation response on future aridity in a mediterranean catchment. Agricultural Water Management 299, 108878. URL: http://dx.doi.org/10.1016/j.agwat.2024.108878, doi:10.1016/j.agwat.2024.108878.

Viney, N., Perraud, J., Vaze, J., Chiew, F., Post, D., Yang, A., 2009. The usefulness of bias constraints in model calibration for regionalisation to ungauged catchments, in: Proc. 18th World IMACS Congress and MODSIM09 Int. Congress on Modelling and Simulation, Citeseer. pp. 3421–3427.

Wagener, T., 2022. On the Evaluation of Climate Change Impact Models for Adaptation Decisions. Springer International Publishing. p. 33–40. URL: http://dx.doi.org/10.1007/978-3-030-86211-4_5, doi:10.1007/978-3-030-86211-4_5.

Wang, X., Zhang, Y., Luo, Y., Sun, L., Shafeeque, M., 2018. Combined use of volume-area and volume-length scaling relationships in glacio-hydrological simulation. Hydrology Research 49, 1753–1772. URL: https://doi.org/10.2166/nh.2018.137, doi:10.2166/nh.2018.137.

Węglarczyk, S., 1998. The interdependence and applicability of some statistical quality measures for hydrological models. Journal of Hydrology 206, 98–103. URL: http://dx.doi.org/10.1016/S0022-1694(98)00094-8, doi:10.1016/s0022-1694(98)00094-8.

Wen, S., Su, B., Wang, Y., Zhai, J., Sun, H., Chen, Z., Huang, J., Wang, A., Jiang, T., 2020. Comprehensive evaluation of hydrological models for climate change impact assessment in the upper yangtze

river basin, china. Climatic Change 163, 1207–1226. URL: https://doi.org/10.1007/s10584-020-02929-6, doi:10.1007/s10584-020-02929-6.

Wiersma, P., Aerts, J., Zekollari, H., Hrachowitz, M., Drost, N., Huss, M., Sutanudjaja, E.H., Hut, R., 2022. Coupling a global glacier model to a global hydrological model prevents underestimation of glacier runoff. Hydrology and Earth System Sciences 26, 5971–5986. URL: https://doi.org/10.5194/hess-26-5971-2022, doi:10.5194/hess-26-5971-2022.

Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biometrics Bulletin 1, 80. URL: http://dx.doi.org/10.2307/3001968, doi:10.2307/3001968.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR guiding principles for scientific data management and stewardship. Scientific Data 3. URL: https://doi.org/10.1038/sdata.2016.18, doi:10.1038/sdata.2016.18.

Willmott, C.J., 1981. On the validation of models. Physical Geography 2, 184–194. URL: http://dx.doi.org/10.1080/02723646.1981.10642213, doi:10.1080/02723646.1981.10642213.

Willmott, C.J., Ackleson, S.G., Davis, R.E., Feddema, J.J., Klink, K.M., Legates, D.R., O'Donnell, J., Rowe, C.M., 1985. Statistics for the evaluation and comparison of models. Journal of Geophysical Research: Oceans 90, 8995–9005. URL: http://dx.doi.org/10.1029/JC090iC05p08995, doi:10.1029/jc090ic05p08995.

Willmott, C.J., Robeson, S.M., Matsuura, K., 2011. A refined index of model performance. International Journal of Climatology 32, 2088–2094. URL: http://dx.doi.org/10.1002/joc.2419, doi:10.1002/joc.2419.

Wilson, A.M., Gladfelter, S., Williams, M.W., Shahi, S., Baral, P., Armstrong, R., Racoviteanu, A., 2017. High asia: The international dynamics of climate change and water security. The Journal of Asian Studies 76, 457–480. URL: https://doi.org/10.1017/s0021911817000092, doi:10.1017/s0021911817000092.

WMO World Meteorological Organization, 2024. Wmo confirms that 2023 smashes global temperature record. Accessed Feb 5, 2024. https://wmo.int/news/media-centre/wmo-confirms-2023-smashes-global-temperature-record.

Wortmann, M., Bolch, T., Krysanova, V., Buda, S., 2016. Bridging glacier and river catchment scales: an efficient representation of glacier dynamics in a hydrological model. Hydrology and Earth System Sciences Discussions URL: https://doi.org/10.5194/hess-2016-272, doi:10.5194/hess-2016-272.

Wortmann, M., Bolch, T., Menz, C., Tong, J., Krysanova, V., 2018. Comparison and correction of high-mountain precipitation data based on glacio-hydrological modeling in the tarim river headwaters (high asia). Journal of Hydrometeorology 19, 777–801. URL: https://doi.org/10.1175/jhm-d-17-0106.1, doi:10.1175/jhm-d-17-0106.1.

Wu, F., Zhan, J., Wang, Z., Zhang, Q., 2015. Streamflow variation due to glacier melting and climate change in upstream heihe river basin, northwest china. Physics and Chemistry of the Earth, Parts A/B/C 79-82, 11–19. URL: https://doi.org/10.1016/j.pce.2014.08.002, doi:10.1016/j.pce.2014.08.002.

Xenarios, S., Gafurov, A., Schmidt-Vogt, D., Sehring, J., Manandhar, S., Hergarten, C., Shigaeva, J., Foggin, M., 2018. Climate change and adaptation of mountain societies in central asia: uncertainties, knowledge gaps, and data constraints. Regional Environmental Change 19, 1339–1352. URL: https://doi.org/10.1007/s10113-018-1384-9, doi:10.1007/s10113-018-1384-9.

Xing, X., Qian, J., Chen, X., Chen, C., Sun, J., Wei, S., Yimamaidi, D., Zhanar, Z., 2022. Analysis of effects of recent changes in hydrothermal conditions on vegetation in central asia. Land 11, 327. URL: https://doi.org/10.3390/land11030327, doi:10.3390/land11030327.

Xu, B., Lu, Z., Liu, S., Li, J., Xie, J., Long, A., Yin, Z., Zou, S., 2015. Glacier changes and their impacts on the discharge in the past half-century in tekes watershed, central asia. Physics and Chemistry of the Earth, Parts A/B/C 89-90, 96–103. URL: https://doi.org/10.1016/j.pce.2015.03.003, doi:10.1016/j.pce.2015.03.003.

jie Xu, H., ping Wang, X., xiao Zhang, X., 2016. Decreased vegetation growth in response to summer drought in central asia from 2000 to 2012. International Journal of Applied Earth Observation and Geoinformation 52, 390–402. URL: https://doi.org/10.1016/j.jag.2016.07.010, doi:10.1016/j.jag.2016.07.010.

Yang, C., Xu, M., Fu, C., Kang, S., Luo, Y., 2022. The coupling of glacier melt module in SWAT model based on multi-source remote sensing data: A case study in the upper yarkant river basin. Remote Sensing 14, 6080. URL: https://doi.org/10.3390/rs14236080, doi:10.3390/rs14236080.

Zhang, G., Biradar, C.M., Xiao, X., Dong, J., Zhou, Y., Qin, Y., Zhang, Y., Liu, F., Ding, M., Thomas, R.J., 2018. Exacerbated grassland degradation and desertification in central asia during 2000-2014. Ecological Applications 28, 442–456. URL: https://doi.org/10.1002/eap.1660, doi:10.1002/eap.1660.

Zhang, J., Sanderson, A.C., 2009. Jade: Adaptive differential evolution with optional external archive. IEEE Transactions on Evolutionary Computation 13, 945–958. doi:10.1109/TEVC.2009.2014613.

Zhang, X., Srinivasan, R., Hao, F., 2007. Predicting hydrologic response to climate change in the luohe river basin using the swat model. Transactions of the ASABE 50, 901–910. URL: http://dx.doi.org/10.13031/2013.23154, doi:10.13031/2013.23154.

Zhou, Y., Zhang, L., Fensholt, R., Wang, K., Vitkovskaya, I., Tian, F., 2015. Climate contributions to vegetation variations in central asian drylands: Pre- and post-USSR collapse. Remote Sensing 7, 2449–2470. URL: https://doi.org/10.3390/rs70302449, doi:10.3390/rs70302449.