WRF-based dynamical downscaling over High Mountain Asia

vorgelegt von
M. Sc.
Xun Wang
ORCID: 0000-0003-3294-747X

an der Fakultät VI - Planen Bauen Umwelt
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften
- Dr. rer. nat. -

genehmigte Dissertation

Promotionsausschuss:

Vorsitzende:	Prof. Dr. Eva Paton
Gutachter:	Prof. Dr. Dieter Scherer
Gutachter:	Prof. Dr. Jörg Bendix
Gutachter:	Prof. Dr. Thomas Mölg

Tag der wissenschaftlichen Aussprache: 14. Dezember 2021

Berlin, 2022

Acknowledgements

I want to express my deep and sincere gratitude to a number of people, who have helped and supported me during my PhD.

First of all, I would like to thank my supervisor Prof. Dr. Dieter Scherer for giving me the opportunity to work on two wonderful projects: "Climatic and Tectonic Natural Hazards in Central Asia (CaTeNA)" and "Quaternary Tipping Points of Lake Systems in the Arid Zone of Central Asia (Q-TiP)". I appreciate his enthusiasm for every new result and his guidance and encouragement, which helped me to overcome difficulties along this journey.

I would also like to thank Prof. Dr. Jörg Bendix and Prof. Dr. Thomas Mölg for acting as external examiners and reviewing my PhD thesis.

I deeply thank all my colleagues at the Chair of Climatology, TU Berlin. Special thanks go to Marco for reviewing my manuscripts, project reports, and this thesis; to Tom for technical support of our cluster; to Ben for the collaboration on the Q-TiP manuscript and for the help in programming; and to Daniel for his help in different instances.

I am grateful to my parents for their love and support. I thank all my friends, here in Berlin and back in China. Finally, I would like to thank Sebastian for standing by my side and always cheering me up.

Abstract

High Mountain Asia (HMA) is a mountainous area including the Tibetan Plateau (TP) and surrounding mountain ranges. Due to its unique climatic and tectonic settings, the HMA region is highly vulnerable to natural hazards, such as landslides and Glacial Lake Outburst Floods (GLOFs). Under climate change, the frequency of natural hazards is expected to increase, posing further threats to society and human lives in HMA.

Due to the lack of meteorological data, the triggering mechanisms of atmospherically induced natural hazards, especially landslides, are still not fully understood in HMA, which hinders the development of early-warning systems. To overcome this issue, a new atmospheric data set: the High Asia Refined analysis version 2 (HAR v2), was developed and is presented in this thesis. The HAR v2 was generated by dynamical downscaling of ERA5 reanalysis data using the Weather Research and Forecasting Model (WRF). The HAR v2 provides atmospheric data at 10 km grid spacing and hourly temporal resolution. It is currently available from 2000 to 2020 and will be extended back to 1979. Compared to the old version, the HAR v2 covers a broader area and a longer temporal range.

To find the optimal model configuration of the HAR v2, several sensitivity experiments were conducted. Validation of the HAR v2 against in-situ stations from the Global Surface Summary of the Day (GSOD) shows that the HAR v2 fits well with observations and outperforms its forcing data ERA5. In addition, the HAR v2 and a version of the HAR v2 run with 2 km grid spacing (HAR v2 2 km) were compared with other commonly used gridded precipitation data sets (reanalysis data, satellite retrieval, and interpolated in-situ observations), over a sub-region in HMA with rugged terrain. Results indicate the added value of the HAR v2 and HAR v2 2 km since they are the only products that can reproduce orographic precipitation and capture more extreme events.

One of the primary goals of the HAR v2 is to provide atmospheric data for landslide researches in Kyrgyzstan and Tajikistan, one of the landslide hot spots in the HMA region. The newly developed HAR v2 was combined with historical landslide inventories to investigate the atmospheric triggering mechanisms of landslides in this region. Results reveal the crucial role of snowmelt in landslide triggering in this region and the added

value of climatic disposition derived from atmospheric triggering conditions in landslide susceptibility mapping. Furthermore, the majority of previous studies applied rainfall estimates from in-situ gauges or satellite retrievals. This study also highlights the potential of dynamical downscaling products generated by regional climate models in landslide prediction.

Dynamical downscaling has already been extensively applied to understand the present-day climate and also future climate. In the last part of this thesis, the applicability of dynamical downscaling in the context of paleoclimate is demonstrated. Here, two global climate simulations for the present day and the mid-Pliocene ($\sim$3 Ma) were dynamically downscaled to 30 km grid spacing over HMA, using the same model configuration as the HAR v2. By keeping land surface conditions the same in both downscaling experiments, this study was able to isolate the influence of large-scale climate states and reveal its role in maintaining the Qaidam mega-lake system during the mid-Pliocene. An increase of the water balance (ΔS), i.e., the change in terrestrial water storage was found, when the mid-Pliocene climate is imposed on the Qaidam Basin (QB) with its modern land surface settings. This imbalance of ΔS induced solely by the changes in large-scale climate state would lead to an increase of lake extent until a new equilibrium state is reached. The estimated equilibrium lake extent is 12%-21% of the maximum lake extent approximated from proxy data.

Zusammenfassung

Hochasien (engl. High Mountain Asia, HMA) ist eine gebirgige Region, die das Tibetische Plateau (TP) und die umliegenden Gebirgsketten umfasst. Aufgrund der einzigartigen klimatischen und tektonischen Gegebenheiten ist die HMA-Region sehr anfällig für Naturgefahren wie Erdrutsche und Gletscherseeausbrüche (engl. Glacier Lake Outburst Floods, GLOFs). Im Zuge des Klimawandels wird erwartet, dass die Häufigkeit von Naturgefahren zunehmen wird, was eine weitere Bedrohung für die Gesellschaft und Menschenleben in der HMA-Region darstellt.

Aufgrund des Mangels an meteorologischen Daten sind die Auslösemechanismen atmosphärisch induzierter Naturgefahren, insbesondere von Erdrutschen, in der HMA-Region noch nicht vollständig verstanden. Dies behindert die Entwicklung von Frühwarnsystemen. Um dieses Problem zu überwinden, wurde ein neuer atmosphärischer Datensatz, die High Asia Refined analysis Version 2 (HAR v2), entwickelt und wird in dieser Arbeit vorgestellt. Die HAR v2 wurde durch dynamisches Downscaling von ERA5-Reanalysedaten unter Verwendung des Weather Research and Forecasting Model (WRF) generiert. Die HAR v2 liefert atmosphärische Daten in einem 10 km Gitterabstand und stündlicher Zeitauflösung. Der Datensatz ist derzeit von 2000 bis 2020 verfügbar und wird bis 1979 zurück erweitert. Im Vergleich zur alten Version hat die HAR v2 eine größere räumliche Ausdehnung und einen längeren zeitlichen Bereich.

Um die optimale Modellkonfiguration der HAR v2 zu finden, wurden mehrere Sensitivitätsexperimente durchgeführt. Die Validierung der HAR v2 mit Beobachtungen aus der Global Surface Summary of the Day (GSOD), dass die HAR v2 mit den Beobachtungen gut übereinstimmt und die Antriebsdaten ERA5 übertrifft. Zusätzlich wurden die HAR v2 und eine Version der HAR v2 mit 2 km Gitterabstand (HAR v2 2 km) mit anderen häufig verwendeten gerasterten Niederschlagsdatensätzen (Reanalyse-, Satellitendaten und interpolierte In-situ-Beobachtungen) verglichen. Die Ergebnisse demonstrieren die Notwendigkeit von höher aufgelösten HAR v2- und HAR v2 2 km-Daten für Regionen mit komplexer Topographie, da sie die einzigen Datensätze sind, die den orographischen Niederschlag reproduzieren und mehr extreme Ereignisse erfassen können.

Eines der Hauptziele der HAR v2 ist die Bereitstellung von atmosphärischen Daten für die Erdrutschforschung in Kirgisistan und Tadschikistan, einem der Erdrutsch-Hotspots in der HMA-Region. Die neu entwickelte HAR v2 wurde mit historischen Erdrutschinventaren kombiniert, um die atmosphärischen Auslösemechanismen von Erdrutschen in dieser Region zu untersuchen. Die Ergebnisse zeigen die entscheidende Rolle der Schneeschmelze bei der Auslösung von Erdrutschen in dieser Region und den zusätzlichen Wert der aus den atmosphärischen Auslösebedingungen abgeleiteten klimatischen Disposition. Die meisten früheren Studien verwendeten Niederschlagsschätzungen aus In-situ-Messgeräten oder Satellitenabruf. Diese Studie unterstreicht auch das Potenzial von dynamischen Downscaling-Produkten, die von regionalen Klimamodellen erzeugt werden, für die Vorhersage von Erdrutschen.

Dynamisches Downscaling wurde bereits ausgiebig angewandt, um das heutige Klima und auch das zukünftige Klima zu verstehen. Im letzten Teil dieser Arbeit wird die Anwendbarkeit des dynamischen Downscaling im Kontext des Paläoklimas demonstriert. Hier wurden zwei globale Klimasimulationen für die Gegenwart und das mittlere Pliozän ($\sim$3 Ma) auf 30 km Gitterabstand verfeinert, wobei dieselbe Modellkonfiguration wie bei der HAR v2 verwendet wurde. Indem die Landoberflächenbedingungen in beiden Downscaling-Experimenten gleich gehalten wurden, war diese Studie in der Lage, den Einfluss großräumiger Klimazustände zu isolieren und ihre Rolle bei der Aufrechterhaltung des Qaidam-Megaseesystems während des mittleren Pliozäns aufzuzeigen. Es wurde eine Zunahme der Wasserbilanz (ΔS), d.h. der Veränderung der terrestrischen Wasserspeicherung, gefunden, wenn das mittelpliozäne Klima dem Qaidam-Becken (QB) mit seinen modernen Landoberflächeneinstellungen aufgezwungen wird. Dieses Ungleichgewicht von ΔS, das allein durch die Änderungen des großräumigen Klimazustands induziert wird, würde zu einer Zunahme der Seeausdehnung führen, bis ein neuer Gleichgewichtszustand erreicht ist. Die geschätzte Gleichgewichts-Seeausdehnung beträgt 12%-21% der maximalen Seeausdehnung, die aus Proxydaten approximiert wurde.

Contents

Contents

List of Symbols

ΔS	Water balance
Δz	Rise in lake level
A_{lake}	Lake extent
d	Euclidean distance to the optimal point
ET	Evapotranspiration
I_{max}	Maximum intensity
I_{mean}	Mean intensity
P	Precipitation
$P - ET$	Net precipitation
Q	Accumulated amount
Q_2	Specific humidity at 2 m
r_s	Spearman's rank correlation coefficient
T_2	Air temperature at 2 m
z_{lake}	Lake level

List of Abbreviations

AUC	area under the receiver operating characteristic curve
AWT	Atmospheric Water Transport
BMBF	Bundesministerium für Bildung und Forschung
CaTeNA	Climatic and Tectonic Natural Hazards in Central Asia
DFG	Deutsche Forschungsgemeinschaft
EASM	East Asian Summer Monsoon
FAR	False Alarm Rate
FNL	Final Operational Global Analysis
GCM	General Circulation Model
GFLD	Global Fatal Landslide Database
GLC	Global Landslide Catalog
GLOF	Glacial Lake Outburst Flood
GSOD	Global Surface Summary of the Day
HAR	High Asia Refined analysis
HAR v2	High Asia Refined analysis version 2
HMA	High Mountain Asia
HR	Hit Rate
IMS	Interactive Multi-Sensor Snow and Ice Mapping System
ISM	Indian Summer Monsoon
JRA-55	Japanese 55-year Reanalysis
LGM	Last Glacial Maximum
MAE	Mean Absolute Error

MB	Mean Bias
NCEI	National Centres of Environmental Information
NCEP	National Centers for Environmental Prediction
NPSH	Northwest Pacific Subtropical High
PPS	Physical Parameterization Scheme
PRISM	Pliocene Research Interpretation and Synoptic Mapping
PSS	Peirce Skill Score
Q-TiP	Quaternary Tipping Points of Lake Systems in the Arid Zone of Central Asia
QB	Qaidam Basin
RCM	Regional Climate Model
RGI 6.0	Randolph Glacier Inventory version 6.0
RMSE	Root Mean Square Error
TOPSIS	Technique for Order Preference by Similarity to the Ideal Solution
TP	Tibetan Plateau
TWS	Terrestrial Water Storage
UTL	up to landslide occurrence
WRF	Weather Research and Forecasting Model

List of Figures

List of Tables

1 Introduction

1.1 Background and motivation

High Mountain Asia (HMA) is a broad mountainous area that includes the Tibetan Plateau (TP) and its surrounding mountain ranges, such as the Himalayas, the Hindu Kush, the Tien Shan, and the Kunlun Mountain (Figure 1.1). The region holds the largest concentration of high-elevation lakes in the world (Lei et al., 2013) and is also the largest reservoir of glacier and ice outside the polar region (Qiu, 2008), which serves as the water source for major Asian rivers and for over one billion of people (Immerzeel et al., 2010). HMA is under the influence of mid-latitude westerlies and Asian monsoon systems (e.g., Schiemann et al., 2009; Bolch et al., 2012; Yao et al., 2012; Mölg et al., 2014; Gao et al., 2014).

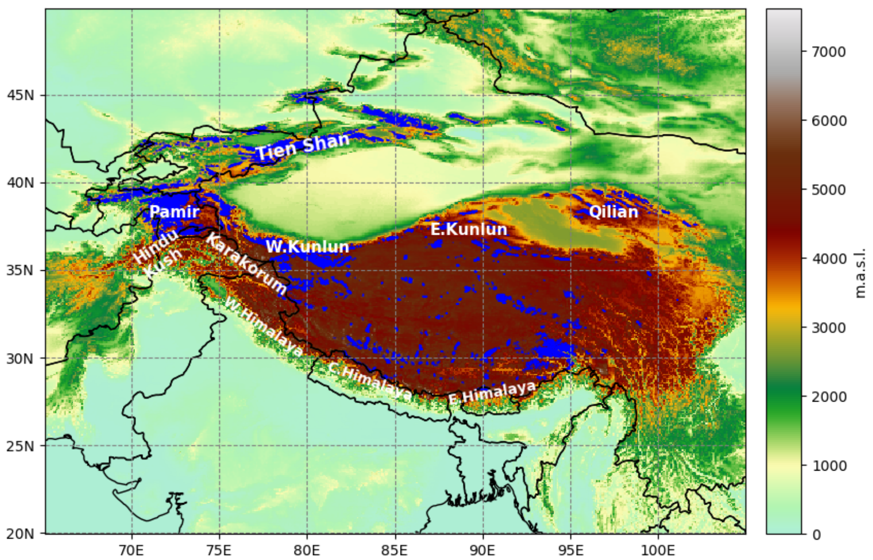

Figure 1.1: Overview of HMA and locations of major mountain ranges. Letters "W", "C", and "E" denote "western", "central", and "eastern". Glaciers from Randolph Glacier Inventory version 6.0 (RGI 6.0) are marked in blue.

HMA is extremely vulnerable to natural hazards (Kirschbaum et al., 2019). HMA presents a global hot spot of rainfall-triggered landslides due to its special tectonic and climatic settings. Global Landslide Catalog (GLC) (Kirschbaum et al., 2010; Kirschbaum et al., 2015) recorded more than 1600 rainfall-triggered landslides in HMA from 2007 to 2018 with more than 1000 fatalities per year (Figure 1.2). According to the Randolph Glacier Inventory version 6.0 (RGI 6.0), HMA encompasses 97 973 glaciers with a total area of 98 768.86 km² (Figure 1.1). Glaciers have retreated on average over HMA as a response to anthropogenic climate change (Cogley, 2016), except in the Pamir–Karakorum–western Himalaya region (Bazai et al., 2020). In 2018, 30 121 (~2080.12 km²) glacial lakes were identified over HMA (Wang et al., 2020). The sudden discharge of water from glacial lakes, known as Glacial Lake Outburst Floods (GLOFs), can damage the downstream areas. A global study (Carrivick and Tweed, 2016) found that Bhutan and Nepal have the greatest economic consequences of GLOFs. These climate-triggered hazards already pose a threat to society and human lives. Under global warming, glacier retreat will lead to an expansion of glacier lakes and increase the risk of GLOFs. On the other hand, increases in extreme precipitation are expected to enhance landslide activities in HMA, with the highest increase rate over areas covered by glaciers and glacial lakes (Kirschbaum et al., 2020). Therefore, understanding the triggering mechanisms of these hazards is crucial and fundamental to developing early-warning systems.

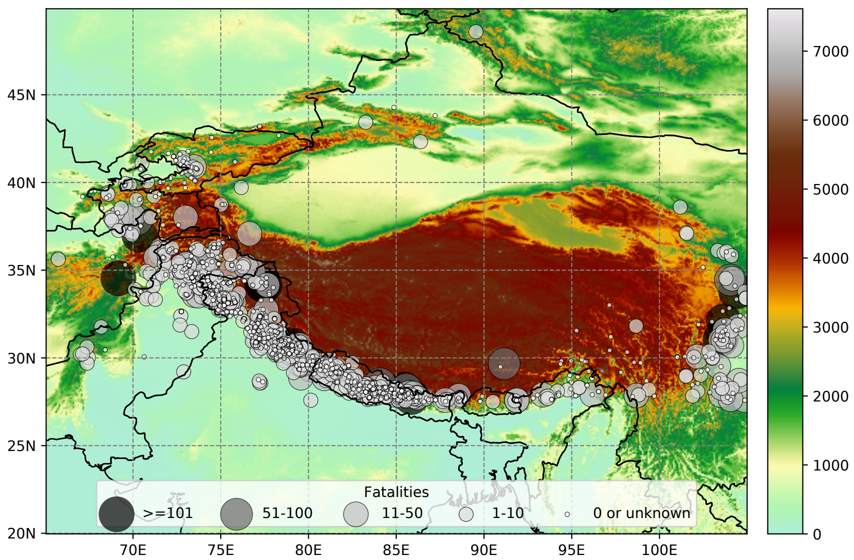

Figure 1.2: *An example of natural hazards that affect HMA. Rainfall-triggered landslides in HMA and the corresponding fatalities from 2007-2018 extracted from the GLC (Kirschbaum et al., 2010; Kirschbaum et al., 2015).*

Due to the complex terrain with harsh environments and limited access, in-situ meteorological stations are sparsely and unevenly distributed in HMA. Scarce meteorological data availability in HMA constitutes a challenge, and the atmospheric triggering mechanisms of the aforementioned localized hazards are not yet fully understood. For instance, as hot spots for landslide activities, no rainfall threshold for landslide triggering has been defined in Kyrgyzstan and Tajikistan (Segoni et al., 2018), even though this topic has already been thoroughly investigated in other parts of the world (e.g., Berti et al., 2012; Gariano et al., 2015; Giannecchini et al., 2016; Leonarduzzi et al., 2017). The number of in-situ observation stations in Kyrgyzstan and Tajikistan decreased sharply in the 1990s due to reduced funding. Currently, there are eight stations in Kyrgyzstan and 26 stations in Tajikistan available from Global Surface Summary of the Day (GSOD), which is a publicly available data set. These numbers are already significantly below the recommendation of the World Meteorological Organization, even for flat areas (Ilyasov et al., 2013). Despite the sparse distribution, most GSOD stations are located in low-lying valleys and are not fully representative of the area.

Reanalysis data sets derived from forecast models and data assimilation systems provide a valuable source of homogeneous atmospheric data. Unlike operational analyses, which typically suffer from inconsistency due to frequent updates of operational analysis systems, a reanalysis employs a consistent forecast model and data assimilation system (Dee et al., 2011). However, most reanalysis data sets are so far too coarse to realistically represent the complex terrain and associated processes, such as orography-induced precipitation, orographic drag, and atmospheric water transport (Feser et al., 2011; Zhou et al., 2021).

Downscaling can provide atmospheric data with high spatial resolution and has the potential to overcome the data paucity problem over HMA. Downscaling is a technique used to derive regional climate information from global data sets (von Storch, 1995). It can be achieved by two fundamentally different methods: statistical and dynamical (von Storch et al., 2000). Statistical downscaling is based on empirical relationships between the large-scale information and local variables established for the present-day climate. A recent application of statistical downscaling of precipitation over the TP can be found in Jiang et al. (2021). Dynamical downscaling employs coarse-resolution global data sets, such as analyses, General Circulation Model (GCM) simulations, and reanalyses, as initial and lateral boundary conditions to drive Regional Climate Models (RCMs). Instead of relying on statistical relationships that may not be valid under different conditions, dynamical downscaling physically resolves regional-scale processes and thus, can be applied to a broader range.

Dynamical downscaling bridges the gap between large scale and regional-to-local scale. There exist some controversies regarding the added value of dynamical downscaling over coarser global simulations since the ability of dynamical downscaling is strongly impacted by several factors, such as domain size and position, forcing strategies, and Physical Parameterization Schemes (PPSs) (Xue et al., 2014). Nevertheless, over complex terrain, the added value of dynamical downscaling is generally well agreed (Prömmel et al., 2010; Lin et al., 2018; Zhou et al., 2021), and dynamical downscaling has been widely applied in the HMA region (e.g., Maussion et al., 2011; Pan et al., 2014; Maussion et al., 2014; Gao et al., 2015; Gao et al., 2017; Karki et al., 2018; Ou et al., 2020; Li et al., 2021; Zhou et al., 2021).

The High Asia Refined analysis (HAR) (Maussion et al., 2011; Maussion et al., 2014) is a good example to illustrate the potential and scientific value of dynamical downscaling in data-scarce regions. It was developed at the Chair of Climatology, Technische Universität Berlin, to overcome the data paucity problem over the TP. The HAR was generated by dynamical downscaling of the Final Operational Global Analysis (FNL) from the National Centers for Environmental Prediction (NCEP) using the Weather Research and Forecasting Model (WRF). Several studies indicate the high accuracy and quality of the HAR (Maussion et al., 2014; Pritchard et al., 2019; Li et al., 2020). Over the last few years, the HAR was comprehensively analyzed to provide a better insight of e.g., the seasonality, variability, and dynamic controls of precipitation over the TP (Maussion et al., 2014; Curio and Scherer, 2016); the atmospheric water transport on and to the TP (Curio et al., 2015); the large-scale driver of local climate variability (Mölg et al., 2014; Mölg et al., 2017). The HAR was also applied as forcing data for glacier mass-balance modeling (Mölg et al., 2014), snow and energy balance modeling (Huintjes et al., 2015), and hydrological modeling (Biskop, 2016). However, there are two limitations of the HAR: (1) the incomplete coverage of the 10 km domain over HMA; (2) the relatively short temporal coverage (15 years). This calls for the need for a new version of the HAR.

Dynamical downscaling has been extensively applied to understand the present-day climate and also future climate (e.g., Pierce et al., 2013; Hong and Kanamitsu, 2014; Jacob et al., 2014; Xue et al., 2014; Bao et al., 2015; Dosio et al., 2015; Di Luca et al., 2016). However, the application of dynamical downscaling is less common in paleoclimate (Ludwig et al., 2019). Most paleoclimate studies rely on proxy-based reconstructions or output from GCMs. GCMs have been applied to reveal the mechanisms and climate drivers of global and regional climate changes, which can not be explained by proxy data (e.g., Kitoh et al., 2001; Tao et al., 2010; Stepanek and Lohmann, 2012; Haywood et al., 2016; Mutz et al., 2018; Botsyun et al., 2020). GCMs are usually operated with coarse resolutions (normally larger than 1°) to reduce computational cost and acquire long-term simulations. But therefore, they are not able to reproduce key processes at regional scales realistically. To the author's knowledge, only a few paleoclimate studies applied the dynamical downscaling technique, e.g., for the Last Glacial Maximum (LGM) (Yoo et al., 2016; Strandberg et al., 2011; Ludwig et al., 2017), the Holocene (Patricola and Cook, 2007; Russo and Cubasch, 2016), the last millennium (Gómez-Navarro et al., 2011; Gómez-Navarro et al., 2013; Gómez-Navarro et al., 2015).

Ludwig et al. (2019) summarized the difficulties of regional paleoclimate modeling as follows: (1) high computational cost; (2) lack of atmospheric fields from global model simulations; (3) technical problems. Within the "Quaternary Tipping Points of Lake Systems in the Arid Zone of Central Asia (Q-TiP)" project, we were able to work with scientists with expertise in global paleoclimate modeling and overcome the difficulties mentioned above in regional paleoclimate modeling. The primary goal of the Q-TiP project was to explore the control factors of water resources in Central Asia, conditioned by climate and other processes, for the geological past. Our particular focus within the project was to examine the role of climate in maintaining the Qaidam mega-lake system during the mid-Pliocene. The lower part of the Qaidam Basin (QB) is featured with hyperarid conditions today. But the QB once held a mega-lake system during the Pliocene Epoch (5.33 to 2.58 Ma).

1.2 Objectives

The principal objectives of this thesis are to:

I apply the dynamical downscaling method to develop a new version of the HAR to provide long-term and high-quality atmospheric data over HMA;

II explore the added values of this new data set over other commonly used gridded data sets with regards to precipitation;

III apply the new version of the HAR to understand the atmospheric triggering mechanisms of landslides in Kyrgyzstan and Tajikistan, one of the hot spots of landslide activities in HMA;

IV extend the application of the dynamical downscaling method in paleoclimate context and investigate the role of climate states on the maintenance of the Qaidam mega-lake system during the mid-Pliocene.

The research covered in this thesis was conducted within the framework of the "Climatic and Tectonic Natural Hazards in Central Asia (CaTeNA)" project and the Q-TiP project funded by Bundesministerium für Bildung und Forschung (BMBF).

Objective I was motivated by the success and limitations of the HAR. Here, an atmospheric data set: the High Asia Refined analysis version 2 (HAR v2) with a spatial resolution of 10 km and temporal coverage from 2000 to 2020 is presented. For generating the HAR v2, ERA5 reanalysis data set was chosen as the forcing data. Sensitivity studies to different forcing strategies, PPSs, and initial snow depth conditions were conducted to find the optimal model configurations for the HAR v2. Then, the HAR v2 was validated against in-situ observations from GSOD and compared with ERA5 in terms of precipitation and air temperature at 2 m.

More and more gridded precipitation products are available over the HMA region derived from different sources and with different time spans and spatial resolutions. The question arises what is the added value of the newly developed HAR v2 over other commonly used data sets? To answer this question and address Objective II, HAR v2 was compared with seven other commonly used gridded precipitation data sets, including reanalysis data, satellite-based precipitation retrieval, and interpolated ground observations, over a subregion of the Central Himalaya and the Southwestern TP with highly complex terrain. In addition, to examine the influence of horizontal grid spacing on precipitation simulation, ERA5 was dynamically downscaled further to a grid spacing of 2 km, which was also included in the intercomparison study.

One primary goal for developing the HAR v2 is to provide high-resolution atmospheric data for investigating the atmospheric triggers of landslides in Kyrgyzstan and Tajikistan within the framework of the CaTeNA project. Landslide is one of the most severe natural hazards in Kyrgyzstan and Tajikistan, but little attention has been paid to the atmospheric triggering mechanism of landslide in this region due to the lack of available atmospheric data. The newly developed HAR v2 has the potential to bridge this gap. Addressing Objective III, the applicability of the HAR v2 in landslide predicting is demonstrated.

To achieve Objective IV, the sensitivity of water balance in the QB to different climate states was examined by two WRF dynamical downscaling experiments of global climate simulations for the present day and the mid-Pliocene. Through climate modeling, the climatic and non-climatic factors influencing the water balance in the QB can be separated.

1.3 Structure of the thesis

This thesis is presented in a cumulative form and based on the following four peer-reviewed papers, which are reprinted in their original form in the appendix:

I **Wang, X.**, Tolksdorf, V., Otto, M. and Scherer, D., 2021: WRF-based dynamical downscaling of ERA5 reanalysis data for High Mountain Asia: Towards a new version of the High Asia Refined analysis. *Int. J. Climatol.*, 41, 743–762, https://doi.org/10.1002/joc.6686.

II Hamm, A., Arndt, A., Kolbe, C., **Wang, X.**, Thies, B., Boyko, O., Reggiani, P., Scherer, D., Bendix, J. and Schneider, C., 2020: Intercomparison of Gridded Precipitation Datasets over a Sub-Region of the Central Himalaya and the Southwestern Tibetan Plateau. *Water*, 12(11), 3271, https://doi.org/10.3390/w12113271.

III **Wang, X.**, Otto, M. and Scherer, D., 2021: Atmospheric triggering conditions and climatic disposition of landslides in Kyrgyzstan and Tajikistan at the beginning of the 21st century. *Nat. Hazards Earth Syst. Sci.*, 21, 2125–2144, https://doi.org/10.5194/nhess-21-2125-2021.

IV **Wang, X.**[1], Schmidt, B.[1], Otto, M., Ehlers, A.T., Mutz, S., Botsyun, S. and Scherer, D., 2021: Sensitivity of Water Balance in the Qaidam Basin to the Mid-Pliocene Climate. *J. Geophys. Res. Atmos.*, 126(16), https://doi.org/10.1029/2020JD033965.

These four thesis papers are outcomes of the collaborative work of several authors. Each of them addresses each of the four objectives listed in section 1.2. This thesis is organized as follows:

Chapter 2, The High Asia Refined analysis version 2 (HAR v2) presents the development and validation results of the HAR v2 (Objective I, Paper I). In addition, a case study over a sub-region of the Central Himalaya and the Southwestern TP is presented to compare the HAR v2 with other gridded precipitation products (Objective II, Paper II).

Chapter 3, Atmospheric triggers of landslides in Kyrgyzstan and Tajikistan shows an application example of the HAR v2 in investigating the triggering mechanism of landslides in Kyrgyzstan and Tajikistan (Objective III, Paper III).

Chapter 4, Sensitivity of water balance in the Qaidam Basin demonstrates the application of dynamical downscaling in paleoclimate context. Here, the global climate simulations from ECHAM5 of the present day and the mid-Pliocene were downscaled to understand the role of climate in maintaining the Qaidam mega-lake system during the mid-Pliocene (Objective IV, Paper IV).

Chapter 5, Conclusions and outlook summarizes the main conclusions of this thesis and provides perspectives for future researches in HMA.

[1]Schmidt, B. and Wang, X. contributed equally to this work

2 The High Asia Refined analysis version 2 (HAR v2)

This chapter summarizes the development of the HAR v2, including the design and configuration choices (Section 2.1), as well as validation results (Section 2.2). In addition, precipitation from the HAR v2 was compared to other commonly used gridded precipitation products (Section 2.3).

2.1 Development

2.1.1 General setups

The HAR is a regional atmospheric data set generated by dynamical downscaling using the WRF model version 3.3.1. The FNL from the NCEP was used as forcing data. The HAR covers the period from October 2000 to October 2014 and is available in 30 km (3-hourly interval) and 10 km (hourly interval) resolution. Due to its relatively short temporal coverage and incomplete spatial coverage over the TP of the 10 km domain (Figure 2.1a), a new version of the HAR was developed. For generating a new version with longer temporal coverage, FNL is not suitable as forcing data anymore because (1) it is only available from 1999; (2) as other analysis data set, it suffers from inconsistency because the operational analysis system has been updated frequently. For the HAR v2, the newly developed ERA5 reanalysis data (Hersbach et al., 2020) is chosen as forcing data because it provides consistent data from 1950 to near real-time.

WRF version 4.1 was employed for generating the HAR v2. The WRF model configurations for the HAR v2 are summarized in Table 2.1. The domain setup (Figure 2.1b) consists of two-way nested domains with 30 km and 10 km grid spacing. Model output from the 30 km domain was discarded, and only the output from the 10 km domain was used to generate the HAR v2 products. The forcing strategy is daily re-initialization adopted from the HAR: each run started at 12:00 UTC and contained 36 h, with the first 12 h as spin-up

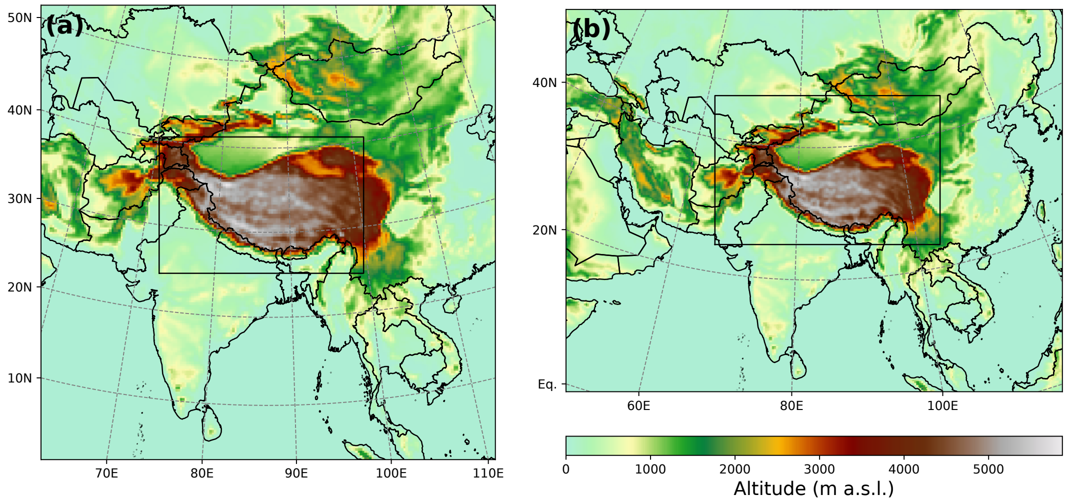

Figure 2.1: *WRF domain configuration for (a) the HAR; (b) the HAR v2. In both (a) and (b), the larger map shows 30 km domain, while the small black box represents 10 km domain.*

time. This strategy prevents the model from deviating too far from the forcing data and provides computational flexibility since daily runs are totally independent of each other and can be computed in parallel and in any sequence.

2.1.2 Sensitivity studies

Nesting strategy

The HAR v2 still applies a 30 km domain as the parent domain in the dynamical downscaling process, even though ERA5 already has a grid spacing of near 32 km. The reason is that a one-day experiment, which directly downscaled ERA5 to 10 km resolution, shows that the large-scale circulation patterns are distorted in the direct downscaling approach, and the 500 hPa wind field from the two-way nesting approach is closer to the forcing data ERA5 (Figure 2.2). Thus, the 30 km was kept as the parent domain.

Physical Parameterization Scheme (PPS)

Changes in forcing data and domain configuration have significant impacts on model output (e.g., Miguez-Macho et al., 2004; Leduc and Laprise, 2009; Kala et al., 2015; Huang and Gao, 2018), which means the original PPSs applied in the HAR might not be suitable for the HAR v2. Therefore, different cumulus, microphysics, planetary boundary layer, and

Table 2.1: WRF Model configurations for the HAR v2. Table taken and modified after Paper I.

Dynamics	
Dynamical solver	Advanced Research WRF (ARW), non-hydrostatic
Maps and grids	
Map projection	Lambert conformal conic
Horizontal grid spacing	30 km (281 x 217 grid points), 10 km (382 x 253 grid points)
Vertical levels	28 Eta-level
Forcing strategy	
Forcing data	ERA5 (0.25°, hourly)
Nesting Strategy	Two-way nesting
Lake surface temperature	Substituted by daily mean surface air temperature
Snow depth	Corrected using JRA-55
Initialization	Daily
Runs starting time	Daily at 12:00 UTC
Runs duration	36 h
Spin-up time	12 h
Physical parameterization schemes	
Longwave radiation	RRTM scheme (Mlawer et al., 1997)
Shortwave radiation	Dudhia scheme (Dudhia, 1989)
Cumulus	Kain-Fritsch cumulus potential scheme (Berg et al., 2013)
Microphysics	Morrison 2-moment scheme (Morrison et al., 2009)
Planetary boundary layer	Yonsei University scheme (Hong et al., 2006)
Land surface model	Unified Noah land surface model (Tewari et al., 2004)
Surface layer	Revised MM5 surface layer scheme (Jiménez et al., 2012)

land surface model schemes were tested in terms of precipitation (P) and air temperature at 2 m (T_2) for January and July 2011 (Section 2.2 in Paper I). The results of PPS sensitivity experiments (Table 3 in Paper I) show that no single PPS performs ideally in both seasons and for both quantities. To solve this problem, the Technique for Order Preference by Similarity to the Ideal Solution (TOPSIS) (Tzeng and Huang, 2011), which is a multiple criteria decision-making method, was applied to identify the best scheme. The basic concept of TOPSIS is to determine the best alternative that has the shortest and longest distance from the positive and negative ideal solution. The cumulus, microphysics, planetary boundary layer, and land surface model schemes listed in Table 2.1 were selected as the best schemes by TOPSIS.

Snow depth correction

Orsolini et al. (2019) evaluated snow depth and snow cover from several global reanalysis data sets over the TP against in-situ and remote-sensing observations. It was found that

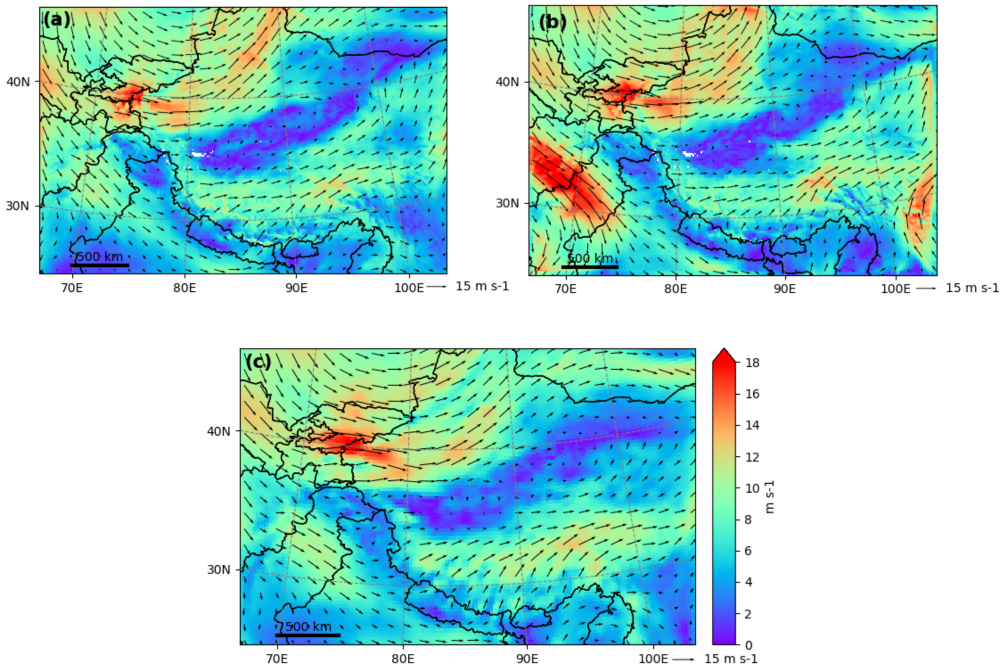

Figure 2.2: Comparison of 500 hPa wind vectors and wind speed (contour) from (a) two-way nesting approach; (b) direct downscaling approach; and (c) ERA5. Figure taken from supplementary of Paper I.

ERA5 largely overestimates the snow depth over the TP, especially during winter. The reason is that ERA5 did not assimilate snow cover from the Interactive Multi-Sensor Snow and Ice Mapping System (IMS) for areas above 1500 m. WRF already exhibits a systematic underestimation of T_2 over high elevated areas (e.g., Gao et al., 2015; Karki et al., 2017; Bonekamp et al., 2018; Maussion, 2014; Pritchard et al., 2019). Overestimation of surface snow depth in the forcing data can lead to additional cold bias in the WRF output (Meng et al., 2018).

To overcome this problem, the snow depth initialized from ERA5 was corrected using the Japanese 55-year Reanalysis (JRA-55) (Kobayashi et al., 2015). JRA-55 was chosen because it has an excellent performance among reanalyses regarding snow depth (Orsolini et al., 2019) and also has a relatively longer temporal coverage. The snow depth correction approach is based on linear scaling. The detailed methodology can be found in Paper I Section 2.4. Figure 2.3 presents the bias score (model minus observation) of monthly mean T_2 from WRF simulations with and without snow depth correction. It can be seen that snow depth

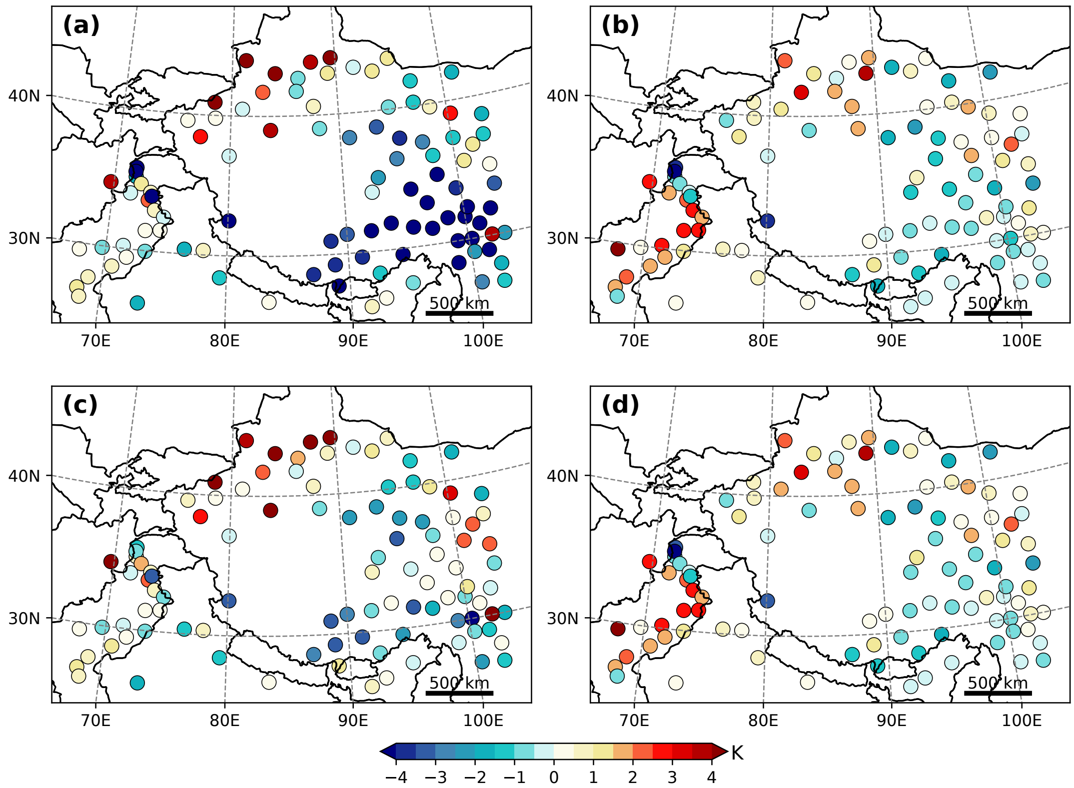

Figure 2.3: Bias scores of monthly mean T_2 for WRF simulations without snow depth correction (a, b), and with snow depth correction (c, d) in January (a, c) and July (b, d) at 103 GSOD stations. Figure modified after Paper I .

correction significantly reduces cold bias, especially in winter over the TP and northern Pakistan.

2.2 Validation

In Paper I, the HAR v2 was validated against in-situ observations and compared with the HAR for 2011 in terms of P and T_2. The results show that compared to the old version, the HAR v2 generally produces slightly higher P amounts, but the spatial distribution of seasonal P matches better to observations. With regards to T_2, the HAR v2 performs better. At the time of writing, the production of the HAR v2 is finished from 2000 to 2020. Here, the HAR v2 and ERA5 for these 21 years are compared to each other and validated against in-situ observations.

2.2.1 Method

In-situ observations of P and T_2 from GSOD provided by the National Centres of Environmental Information (NCEI) were used for validation. Stations with more than 90% of records within the validation period were selected. After filtering, 89 stations are left within the 10 km domain of the HAR v2 from 2000 to 2020. Some GSOD stations contain problematic precipitation values. For example, stations Da-Qaidam, Doulan, and Delingha in the Qaidam Basin in the northeastern TP have shown 486 mm m^{-1}, 764 mm m^{-1}, and 1108 mm m^{-1} precipitation amounts in January 2019 (Da-Qaidam, Doulan) and January 2020 (Delingha), which is unrealistic for the arid condition in the Qaidam Basin. These unrealistic monthly values were masked out for validation.

Statistical measures were applied to assess the model performance, which includes Mean Bias (MB), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the distribution-free Spearman's rank correlation coefficient (r_s). r_s is calculated based on daily values of P and T_2. For T_2, a constant temperature lapse rate of 6.5 K km^{-1} was applied to transfer the T_2 from the HAR v2 and ERA5 to the station altitudes.

2.2.2 Results

Table 2.2 summarizes the statistic metrics. Figure 2.4 and Figure 2.5 present the monthly time series derived from the HAR v2, ERA5, and GSOD, as well as station-wise MB and r_s for P and T_2, respectively.

Table 2.2: *Statistical metrics of P (unit: mm d^{-1} except for r_s) and T_2 (unit: K except for r_s) for the HAR v2 and ERA5.*

	P				T2			
	MB	MAE	RMSE	r_s	MB	MAE	RMSE	r_s
HAR v2	0.29	0.47	0.80	0.41	-0.38	0.94	1.28	0.86
ERA5	0.52	0.61	1.06	0.46	-0.25	0.95	1.29	0.85

The precipitation seasonality and annual variability are generally well reproduced by the HAR v2 (Figure 2.4a). The temporal variability of monthly P from the HAR v2 follows ERA5 closely. Over the whole 10 km domain, both the HAR v2 and ERA5 overestimate P, especially in summer months, but the HAR v2 has a smaller overestimation. In winter, the spatial pattern of MB from the HAR v2 generally follows that from ERA5, but with lower MBs. In summer, the HAR v2 shows a drying in the southern periphery of the TP but a wetting over the middle-eastern TP, compared to ERA5 (Figure 2.4c, f). ERA5 shows

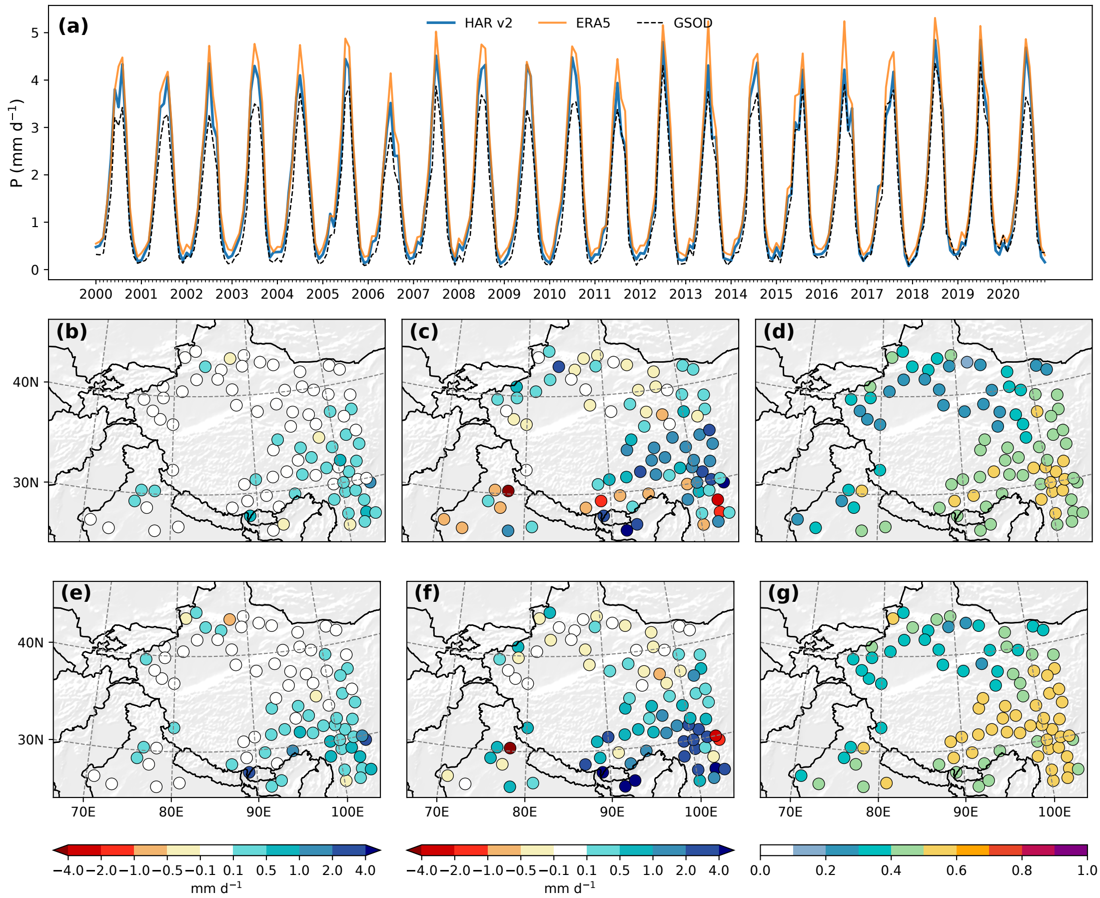

Figure 2.4: *Comparison of the HAR v2, ERA5, and GSOD from 2000 to 2020 in terms of P. Time series averaged over 89 GSOD stations as shown in the following subfigures (a); MB of the HAR v2 (b, c) and ERA5 (e, f) at each station for winter (b, e) and summer (c, f); Spearman's correlation coefficient of the HAR v2 (d) and ERA5 (g) at each station calculated based on daily P.*

more consistency with in-situ observations with regards to temporal variations of daily P (Figure 2.4d, g, Table 2.2). Previous studies also found that ERA5 performs better than WRF simulations in reproducing daily temporal variations of P (e.g., Zhou et al., 2021; Jiang et al., 2021).

For T_2, both the HAR v2 and ERA5 exhibit a systematic cold bias over HMA (Table 2.2). ERA5 shows an overall smaller cold bias due to the overestimation of T_2 in summer (Figure 2.5f). The HAR v2 and ERA5 exhibit different spatial patterns of station-wise biases, especially in summer. Over stations in the southeastern TP, the HAR v2 shows warming in winter, but cooling in summer, compared to ERA5. The warming in winter could result from the snow correction approach (Section 2.1.2). The cooling in summer over the southern TP produced by WRF has also been found in Zhou et al. (2021), according to which the reason could be the decreased latent heat release caused by weakened water vapor transport from South Asia due to higher spatial resolution.

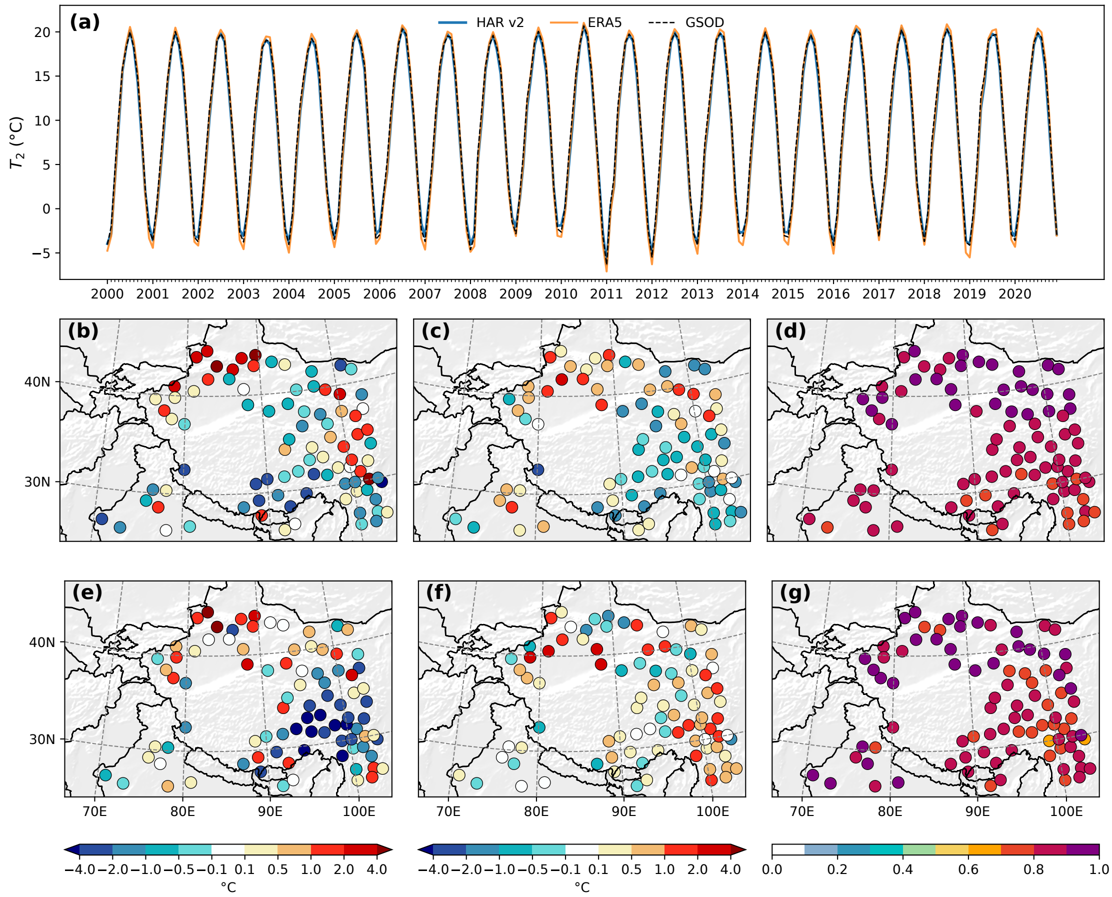

Figure 2.5: *Comparison of the HAR v2, ERA5, and GSOD from 2000 to 2020 in terms of T_2. Time series averaged over 89 GSOD stations as shown in the following subfigures (a); MB of the HAR v2 (b, c) and ERA5 (e, f) at each station for winter (b, e) and summer (c, f); Spearman's correlation coefficient of the HAR v2 (d) and ERA5 (g) at each station calculated based on daily T_2.*

2.3 Comparison of the HAR v2 with other gridded precipitation products

Over complex terrain, precipitation observation is challenging due to the harsh environment and limited accessibility (Wang and Zeng, 2012). Precipitation measurements are usually error-prone. Gridded products, on the other hand, provide continuous estimations of precipitation both spatially and temporally, and thus, can overcome the limitations of rain-gauge observations.

In Paper II, an intercomparison of gridded precipitation datasets was conducted over a sub-region of the Central Himalaya and the Southwestern TP. The study area (81 °E-88 °E, 28 °N-32 °N) includes different topographic features, ranging from low-lying southern slopes of the Himalayas, followed by the extreme relief of the Himalayas, and to the less complex TP terrain (Figure 1 in Paper II). The HAR v2 were compared to seven commonly used gridded data sets, including weather model-derived analysis and reanalysis data sets, satellite-based precipitation retrieval, and interpolated ground observations. Details of these data sets can be found in Paper II Section 2.2. In addition, to investigate the influence of horizontal grid spacing on precipitation simulation, ERA5 has been further downscaled to 2 km grid spacing using WRF V4.1 for the study area from April 2017 to October 2017. The 30 km and 10 km domains of the HAR v2 (Figure 2.1b) were used as the two-way nesting parent domains. To distinguish this new dynamical downscaled product from the original HAR v2, hereinafter they are referred to as HAR v2 2 km and HAR v2 10 km. The model setup for the HAR v2 2 km is the same as the HAR v2 10 km, except that no cumulus parameterization scheme was used, and cumulus convection was thus explicitly resolved in the HAR v2 2 km. The study period for the intercomparison is May to September 2017, covering a full Indian Summer Monsoon (ISM) season.

Figure 2.6 shows that the HAR v2 products have the best representation of orographic precipitation over complex terrains and show improvement over their forcing data ERA5. Interestingly, the ERA5-Land with a similar spatial resolution does not depict as many spatial details as the HAR v2 10 km. The reason is that precipitation is purely interpolated to the ERA5-Land grid from ERA5 and is not obtained by running the land surface model[1]. Climdex indices in Figure 2.7 reveal precipitation extreme for each product. Generally, the higher the spatial resolution, the larger is the data range among all grid cells. Comparison of R10 and R20 (number of wet days with $P > 10\,\mathrm{mm}$ and $P > 20\,\mathrm{mm}$) show that both HAR v2 10 km and HAR v2 2 km with higher resolution return the overall highest maximum values and median values. This indicates that higher resolved products experience more extreme precipitation events in multiple grid cells than coarser products, and they can resolve locally confined heavy precipitation events. HAR v2 10 km has an extremely high precipitation amount of over 500 mm in a single day (Rx1). The much lower Rx1 in HAR v2 2 km could imply the added value of convection-permitting scale simulations. However, this needs to be further investigated in the future since 500 mm d^{-1} is not impossible in our study area. A maximum precipitation amount of 528 mm per 24 h was

[1] https://confluence.ecmwf.int/display/CKB/ERA5-Land

observed during an extreme precipitation event on 14-15 August 2014 at Chisapani station in southwestern Nepal (Karki et al., 2018).

In summary, this intercomparison study highlights the similarities and differences of gridded precipitation products from various sources. It demonstrates the potential and added value of high-resolution dynamical downscaling products (HAR v2 10 km and HAR v2 2 km) over complex terrain.

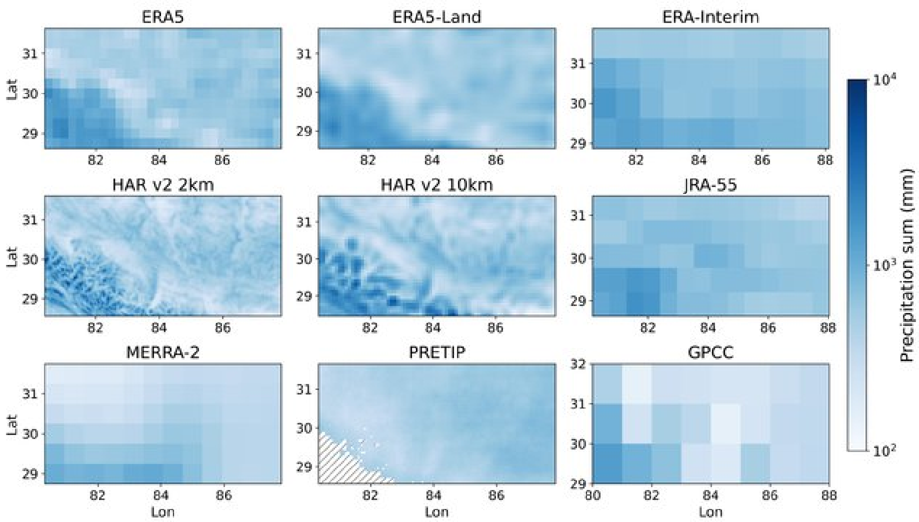

Figure 2.6: Spatial log-scaled per-grid-cell sum over the study period for each of the precipitation product. Figure taken from Paper II.

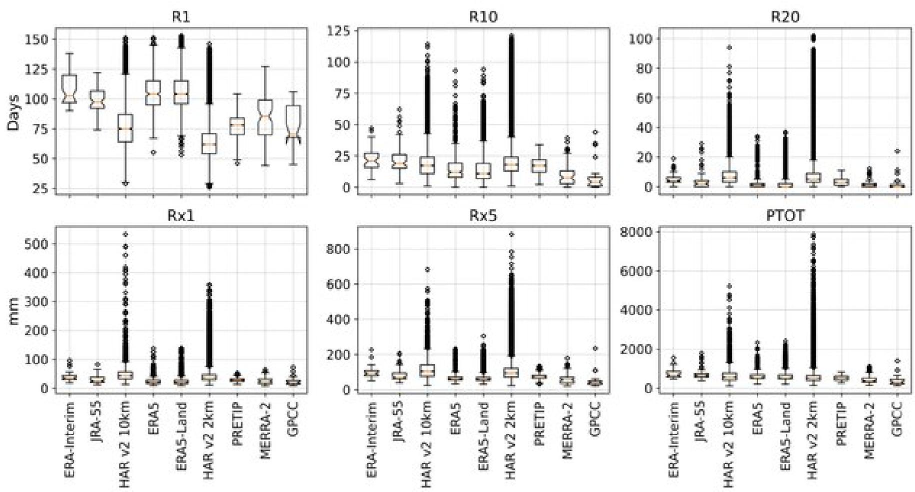

Figure 2.7: Visualization of the selected climdex indices R1 (number of wet days with P > 1 mm), R10 (number of wet days with P > 10 mm), R20 (number of wet days with P > 20 mm), Rx1 (maximum 1-day precipitation), Rx5 (maximum consecutive 5-day precipitation), and PTOT (total precipitation) as boxplots. Figure taken from Paper II.

3 Atmospheric triggers of landslides in Kyrgyzstan and Tajikistan

In this chapter, the HAR v2 is applied to investigate the atmospheric triggering conditions of landslides in Kyrgyzstan and Tajikistan. The results shown here were published in Paper III.

3.1 Background

Landslide is one of the most severe natural hazards in Kyrgyzstan and Tajikistan. While the majority of landslide research in Kyrgyzstan and Tajikistan focused on characterizing landslide susceptibility, i.e., "where" landslides are prone to occur (e.g., Braun et al., 2015; Saponaro et al., 2015; Havenith et al., 2015), little attention has been paid to the atmospheric triggering conditions, and our knowledge of "when" landslides are likely to occur is limited in this region. The reasons are the lack of landslide inventories with the exact date of landslide occurrence and the lack of atmospheric data.

Rainfall is the most common trigger of landslide all over the world (Wieczorek, 1996). Over snow-covered regions, snowmelt is recognized as another common trigger of shallow landslides and debris flows (Wieczorek, 1996; Mostbauer et al., 2018). In Kyrgyzstan and Tajikistan, more than half of the annual precipitation falls in the form of snow. Snow cover duration over high mountain ranges in the Tien Shan and the Pamir is more than 200 days per year (Dietz et al., 2014). A large amount of water stored in snowpacks is released during the melting season. Snowmelt is another critical source of water infiltrating into the soil that increases slope instability. Thus, in Kyrgyzstan and Tajikistan, snowmelt might also play a role in landslide triggering besides rainfall. But snowmelt is not as easy to observe as rainfall and might often be neglected as a landslide trigger, especially when co-occurring with rainfall.

The newly developed, freely available HAR v2 was combined with 96 historical landslide events extracted from GLC and Global Fatal Landslide Database (GFLD) (Froude and

Petley, 2018) from 2004 to 2018 (Figure 3.1) to analyze the atmospheric triggering conditions of landslides and to generate climatic disposition maps in Kyrgyzstan and Tajikistan.

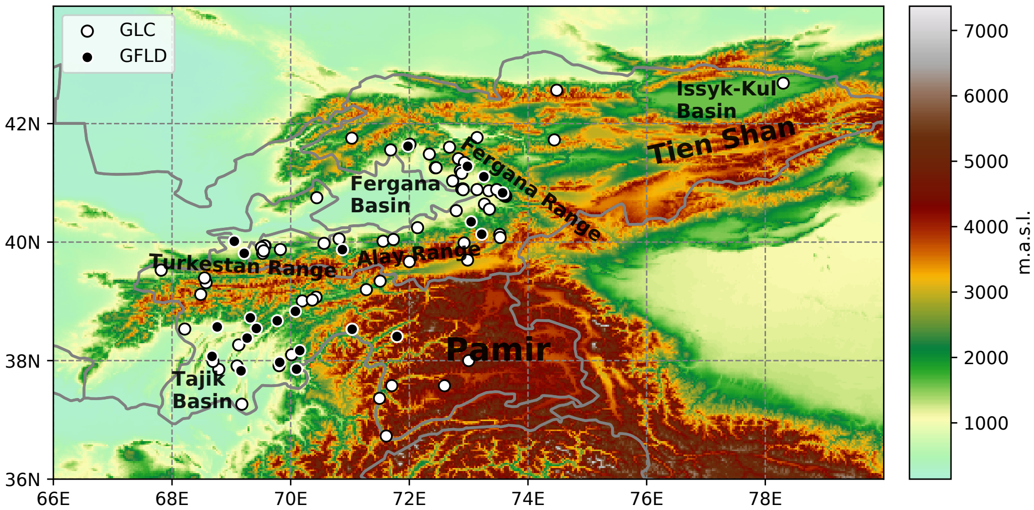

Figure 3.1: Landslide events from 2004 to 2018 extracted from the GLC (white points) and the GFLD (black points). Figure taken from Paper III.

3.2 Atmospheric triggers of landslides

The atmospheric trigger of a landslide event is determined by the co-occurrence of the landslide event with rainfall and snowmelt events. If a landslide event only occurred within or one day after a rainfall (snowmelt) event, then this landslide event is defined as rainfall (snowmelt) triggered. If there are both a rainfall event and a snowmelt event on the day or one day before the landslide occurrence day, then the atmospheric trigger of this landslide event is mixed. The results are presented in Figure 3.2. Nine landslide events did not occur within any rainfall or snowmelt event. This mismatch between landslide information and weather information stems from the uncertainties in landslide locations and timing, as well as the uncertainties in the HAR v2. These nine events are referred to as "not detected" (white points in Figure 3.2) and are excluded for further analysis.

Comparing the annual cycles of rainfall, snowmelt, and the sum of rainfall and snowmelt (hereinafter referred to as rainfall+snowmelt) with monthly landslide occurrences, it can be seen in Figure 3.3 that the peak of landslide activity in April and May corresponds well

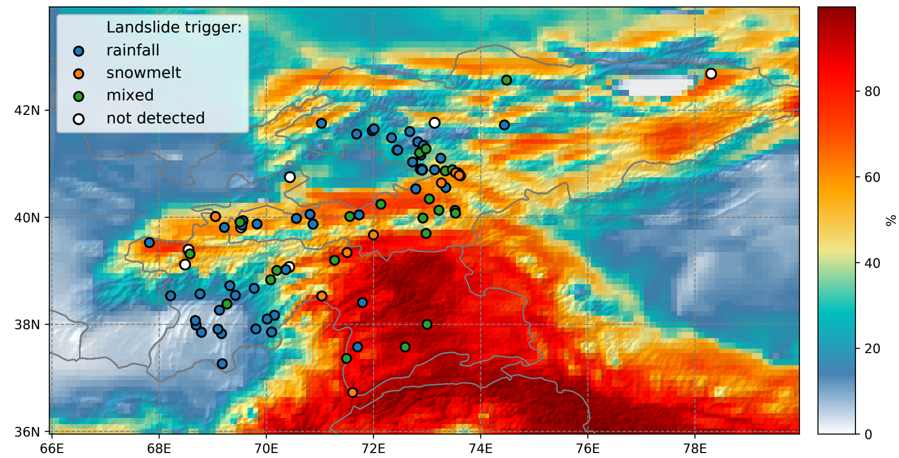

Figure 3.2: *Contribution (%) of snowmelt to annual sum of rainfall and snowmelt (background contour) and atmospheric triggers of 96 landslide events extracted from the GLC and the GFLD (points). Figure taken from Paper III.*

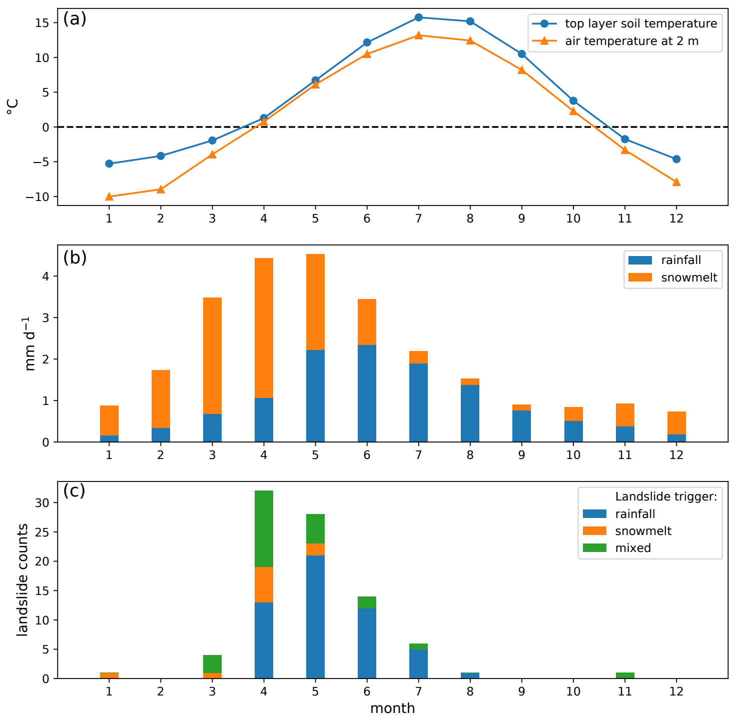

Figure 3.3: *(a) Mean monthly soil temperature at the top soil layer (0-0.1m) and T_2 averaged over Kyrgyzstan and Tajikistan extracted from the HAR v2; (b) mean monthly rainfall and snowmelt averaged over Kyrgyzstan and Tajikistan extracted from the HAR v2; (c) mean monthly landslide occurrences in Kyrgyzstan and Tajikistan from 2004 to 2018. Figure taken from Paper III.*

with the peak of rainfall+snowmelt. Snowmelt contributes to triggering 40% of landslide events (35 out of 87). There are 29% of landslide events (25 out of 87) that are attributed to the combined effect of rainfall and snowmelt. Most snowmelt-contributing events occurred in April when snowmelt amount is the highest. March and June have almost the same amount of rainfall+snowmelt, but more landslide occurred in June. The reason could be the still frozen soil in March (negative soil temperature and T_2 as shown in Figure 3.3a), which makes slopes more stable.

3.3 Thresholds of landslide triggering

The threshold model from Leonarduzzi et al. (2017) was utilized to define the triggering threshold, which can best separate the atmospheric conditions that resulted and did not result in landslides. Firstly, landslide-triggering events and non-triggering events were determined. For each event, three event properties were calculated: mean intensity I_{mean}, maximum intensity I_{max}, and the accumulated amount for the entire event Q. For landslide-triggering events, these three properties were also computed by only considering the period up to landslide occurrence (UTL). Then, the thresholds of rainfall, snowmelt, and rainfall+snowmelt were defined based on maximizing the predictive performance using 2×2 contingency tables. The Peirce Skill Score (PSS), i.e., the difference between the Hit Rate (HR) and the False Alarm Rate (FAR), was chosen to measure the predictive performance because it is unbiased even when the numbers of triggering events and non-triggering events are not equally presented (Woodcock, 1976). Details of threshold model can be found in Paper III Section 2.2.2. The calibrated thresholds are presented in Table 3.1.

Predictive performance is better when using the entire period than just using the UTL period, which was also concluded by Leonarduzzi et al. (2017). One of the reasons is that by considering a longer period, I_{mean}, I_{max}, and especially Q of triggering events generally increase, making it easier to distinguish triggering events from non-triggering events. Another reason could be the uncertainty in landslide timing. Although it is more typical that a landslide was reported after its actual occurrence (positive errors), negative errors are also possible depending on the interpretation of historical landslide information from an analyst (Peres et al., 2018). The better performance by considering the entire period could be an indication of negative errors in the landslide timing.

Despite different performances, the defined thresholds of entire events and UTL events for rainfall+snowmelt are very similar. It can be seen from Table 3.1 that rainfall+snowmelt

Table 3.1: *Calibrated thresholds of mean intensity I_{mean} (mm d^{-1}), maximum intensity I_{max} (mm d^{-1}), and accumulated amount Q (mm) for entire events and UTL events of rainfall, snowmelt, and the sum of rainfall and snowmelt (rainfall+snowmelt), as well as corresponding performance statistics. Table taken from Paper III.*

Predictor	Property	Threshold	HR	FAR	d	PSS	AUC
Rainfall (entire event)	I_{mean}	3.60	0.62	0.35	0.51	0.27	0.62
	I_{max}	11.20	0.49	0.18	0.54	0.32	0.65
	Q	16.95	0.52	0.18	0.52	0.34	0.67
Snowmelt (entire event	I_{mean}	7.05	0.23	0.06	0.77	0.17	0.31
	I_{max}	13.45	0.24	0.04	0.76	0.20	0.32
	Q	119.60	0.24	0.03	0.76	0.21	0.33
Rainfall+snowmelt (entire event)	I_{mean}	4.95	0.71	0.25	0.38	0.46	0.78
	I_{max}	12.80	0.67	0.15	0.37	0.51	0.81
	Q	17.15	0.74	0.23	0.35	0.50	0.81
Rainfall (UTL event)	I_{mean}	3.05	0.60	0.40	0.57	0.20	0.59
	I_{max}	12.40	0.34	0.16	0.67	0.19	0.58
	Q	9.25	0.52	0.31	0.57	0.21	0.59
Snowmelt (UTL event)	I_{mean}	7.40	0.22	0.05	0.78	0.17	0.31
	I_{max}	12.80	0.24	0.05	0.76	0.19	0.32
	Q	98.30	0.24	0.04	0.76	0.20	0.32
Rainfall+snowmelt (UTL event)	I_{mean}	5.05	0.68	0.25	0.41	0.43	0.76
	I_{max}	14.05	0.59	0.14	0.44	0.45	0.77
	Q	15.65	0.66	0.25	0.43	0.40	0.76

has the best predictive performance for both entire events and UTL events. The predictive performance indicated by d (euclidean distance to the optimal point, where HR=1 and FAR=0), PSS, and area under the receiver operating characteristic curve (AUC) of the three event properties (I_{mean}, I_{max}, and Q) are quite similar, but using I_{max} as a predictor leads to a lower FAR but also a lower HR when compared with Q and I_{mean}.

3.4 Climatic disposition

The occurrence of landslides depends on disposition and triggering events. Disposition refers to the general settings that make slopes prone to failure without actually initiating it, such as slope gradient and aspect, geology, vegetation cover, and climate (Dai et al., 2002). Using the thresholds defined in Section 3.3 for rainfall+snowmelt UTL events, Figure 3.4 presents the annual number of rainfall+snowmelt events that exceed I_{mean} =5.05 mm d^{-1}, I_{max} =14.05 mm d^{-1}, and Q =15.65 mm. These mean annual exceedance maps derived from triggering thresholds depict the climatic disposition of landslides. Historical landslide locations generally correspond with high mean annual exceedance. High false alarms in remote areas, such as the Tien Shan, could be because landslides extracted from media reports are under-reported in remote regions. The majority of landslide susceptibility studies only considered non-climatic factors or often simply applied averaged annual precipitation as the climatic factor. Compared to a landslide susceptibility map (Stanley and Kirschbaum, 2017) extracted from non-climatic factors (Figure 8 in Paper III), these mean annual exceedance maps show similarity with regards to topographic features but also exhibit discrepancies, which suggests that both climatic and non-climatic aspects need to be considered for landslide susceptibility mapping.

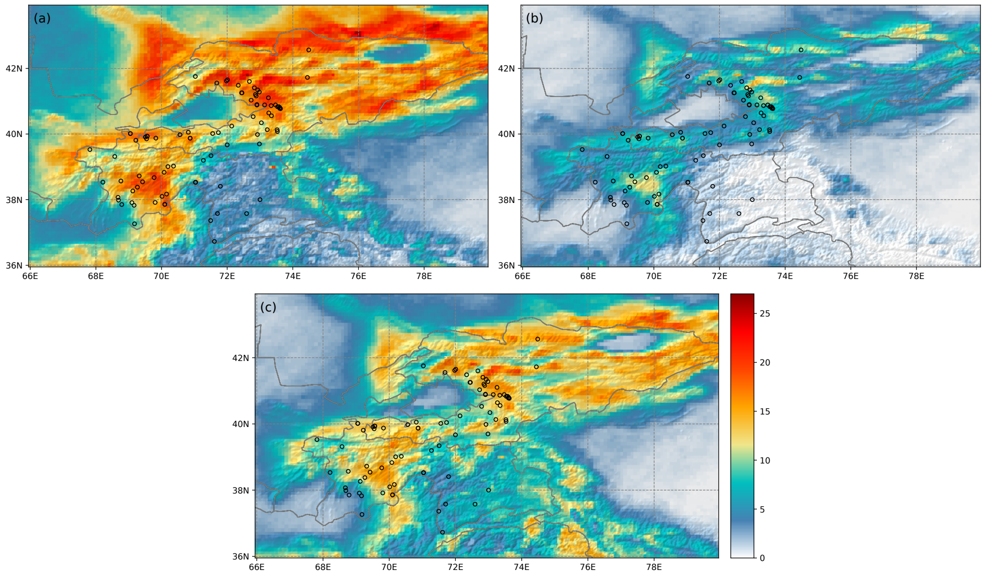

Figure 3.4: *Mean annual exceedance (number of events per year) of (a) I_{mean} =5.05 mm d^{-1}; (b) I_{max} =14.05 mm d^{-1}; and (c) Q =15.65 mm for the rainfall+snowmelt UTL events. Black circles: landslide events from the GLC and the GFLD. Figure modified after Paper III.*

4 Sensitivity of water balance in the Qaidam Basin

This chapter summarizes Paper IV and demonstrates the application of dynamical downscaling in the context of paleoclimate. The research area is the QB located in the northeastern TP, which once held a mega-lake system during the Pliocene Epoch (5.33 to 2.58 Ma). Here, the dynamical downscaling method was applied to investigate the role of the climate in the maintenance of the mega-lake system during the Pliocene.

4.1 Study region and background

The QB is an intermontane endorheic drainage basin with a total area of around 254 000 km^2 located in the northeastern TP (Figure 4.1). Hyperarid condition prevails its central and lower part today. But the QB contained a freshwater mega-lake system throughout the Pliocene as revealed by paleogeographic studies (Chen and Bowler, 1986; Mischke et al., 2010; Wang et al., 2012), even though during this period, the basin and surrounding mountain areas experienced a general aridification process (Rieser et al., 2009; Miao et al., 2013). With the beginning of the Pleistocene, the mega-lake system began to shrink. This process continued throughout the Pleistocene. Today, only a few playas and saline lakes remain in the QB (Wang et al., 2012).

To understand how the Qaidam mega-lake system could survive the continuous aridification for millions of years, Scherer (2020) applied the HAR to investigate the present-day water balance (ΔS) in the QB and its climate drivers. It was found that the ΔS in the QB is close to zero under the present-day climate condition, and specific humidity at 2 m (Q_2) is the climate driver of the annual ΔS in the QB. The annual ΔS in the QB is positively correlated with the annual Q_2. Thus, the Qaidam mega-lake system was sustained by the wetter climate condition during the Pliocene. Following these findings of Scherer (2020), the principal objective of Paper IV is to understand to what extent climate plays a role in the maintenance of the mega-lake system during the mid-Pliocene.

In the mid-Pliocene (∼3 Ma), the global climate was warmer and wetter than today (Ravelo et al., 2004), while paleogeographic features, continental configuration, and ocean bathymetries share similarities to those of today (Dowsett et al., 2010). Therefore, the mid-Pliocene is often considered a past analog of near-future climate (e.g., Zubakov and Borzenkova, 1988; Burke et al., 2018).

The steady-state ΔS is zero in an endorheic basin (Broecker, 2010; Ibarra et al., 2018). Following Scherer (2020), ΔS is defined as the total change in Terrestrial Water Storage (TWS) within all the reservoirs inside the basin. For a large endorheic basin like the QB, ΔS can be calculated as the spatial average of net precipitation ($P - ET$), i.e., the difference between precipitation (P) and evapotranspiration (ET). However, changes in climate state can alter the TWS and cause an imbalance in the ΔS (e.g., Jiao et al., 2015; Wang et al., 2018; Li et al., 2019). Based on the findings of Scherer (2020), the basic hypothesis of Paper IV is that the wetter mid-Pliocene climate state would lead to a positive imbalance of the ΔS in the QB, which would result in recharging of groundwater reservoirs and eventually to a rising lake level and an extension of the lake area. This readjustment of lake extent would continue until a new equilibrium state is reached, where evaporation over lake areas compensates for the input by runoff and precipitation.

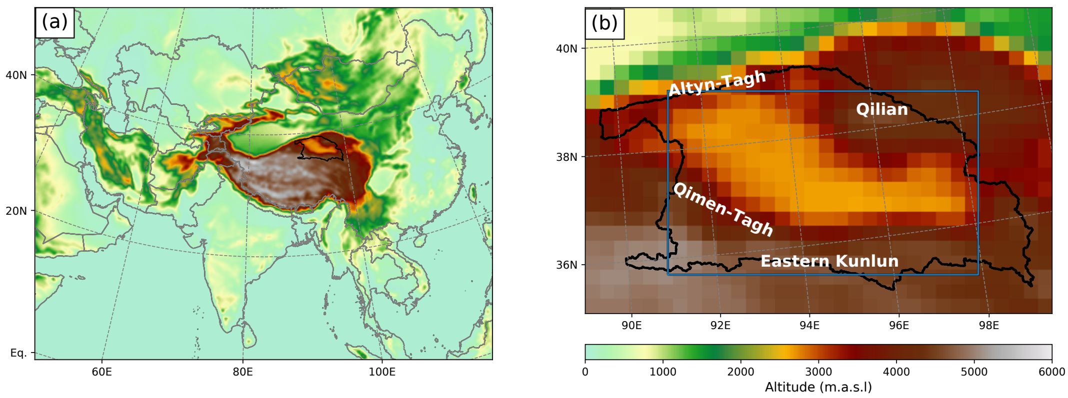

Figure 4.1: *(a) Map of WRF model domain; (b) overview of the QB. Black line: boundary of the QB (Lehner and Grill, 2013). Blue rectangle: Qaidam box for atmospheric moisture budget analysis. Figure taken from Paper IV.*

4.2 Modeling concept

RCMs provide the opportunity to isolate the climate forcing from other factors, such as land use and vegetation cover. By imposing the mid-Pliocene climate state on the QB with its modern land surface settings, an imbalance of ΔS will be induced and can be translated into an equilibrium lake extent needed to remove this imbalance. Comparing this equilibrium lake extent with the estimated lake extent from proxies can give us an insight into the influence of climate on the existence of the mega-lake system during the mid-Pliocene. To achieve this, WRF was applied for dynamical downscaling of present-day and mid-Pliocene global climate simulations from ECHAM5. These two regional downscaling simulations are referred to as PD and PLIO in the following text. Details of the global simulations from ECHAM5 can be found in Mutz et al. (2018) and Botsyun et al. (2020). Except for the forcing data, the WRF model configurations are the same as the HAR v2. Note that, such downscaling is not to reconstruct the regional climate in the QB during the mid-Pliocene. Instead, we simulated the regional climate of the QB and its surrounding areas with modern surface conditions. In this way, we can analyze the large-scale controls of ΔS in the QB independently from other land surface controls.

4.3 Difference in water balance and its implication

Since the PLIO simulation can not be considered a reconstruction of the regional climate in the QB for the mid-Pliocene, it should not be interpreted on its own but in conjunction with the PD simulation. Interpretation of the simulation results should focus on the differences between PLIO and PD simulations. The annual ΔS averaged over the QB is 26 mm a^{-1} higher in PLIO. Seasonally, PLIO has a higher ΔS in the QB in winter, spring, and autumn. In summer, the difference in ΔS between PLIO and PD is negative (Table 4.1). Higher ΔS together with higher Q_2 is in accordance with the conclusion from Scherer (2020) that annual Q_2 is the climate driver of annual ΔS in the QB.

Table 4.1: *Seasonal and annual P (mm season^{-1} or mm a^{-1}), ET (mm season^{-1} or mm a^{-1}), ΔS (mm season^{-1} or mm a^{-1}), T_2 (°C) and Q_2 (g kg^{-1}) averaged over the QB for PLIO, PD, and the difference between PLIO and PD (PLIO-PD). Table taken from Paper IV.*

	PLIO					PD					PLIO-PD				
	P	ET	ΔS	T_2	Q_2	P	ET	ΔS	T_2	Q_2	P	ET	ΔS	T_2	Q_2
DJF	51	22	29	-13.9	1.3	43	25	19	-12.2	1.3	8	-3	11	-1.7	0.0
MAM	131	87	43	-2.7	3.4	111	83	27	0.0	3.1	20	4	17	-2.7	0.3
JJA	163	152	12	8.7	7.4	151	122	29	9.9	6.0	13	30	-17	-1.2	1.4
SON	74	62	12	-3.3	3.3	52	56	-4	-1.0	3.0	22	6	16	-2.3	0.3
Annual	419	322	96	-2.8	3.9	356	286	70	-0.8	3.4	63	36	26	-2.0	0.5

Applying the lake mass balance equation for an endorheic basin, the imbalance in ΔS can be transformed into the equilibrium lake extent (details in Paper IV Section 2.3). Three projections of lake extent were calculated based on three different values of lake evaporation:

600 mm a^{-1}, 800 mm a^{-1}, and 1000 mm a^{-1}. These values represent lower, medium, and upper estimations of lake evaporation rates over the TP based on previous studies (Haginoya et al., 2009; Xu et al., 2009; Yu et al., 2011; Lazhu et al., 2016; Li et al., 2016). Figure 4.2 summarizes the lake extents (A_{lake}), lake levels (z_{lake}), and the rise in lake levels (Δz) of these three equilibrium lake states under the increase in mean annual ΔS of 26 mm a^{-1}. Figure 4.2 also illustrates these three projected states, where lake extents are estimated using WRF model topography for accumulations of net precipitation and subsequent runoff originating from land in grid points at or below equilibrium lake levels. An increase in mean annual ΔS of 26 mm a^{-1} would be sufficient to sustain a lake in the QB with an extent ranging from 7298 km^2 to 12 260 km^2, which is much larger than Lake Qinghai, the largest lake over the TP (4317 km^2). The maximum lake extent of 59 000 km^2 approximated from proxy data (Chen and Bowler, 1986) is still much larger than the estimations presented here. This indicates that factors besides large-scale climate state also contributed to sustaining the mega-lake system during the Pliocene.

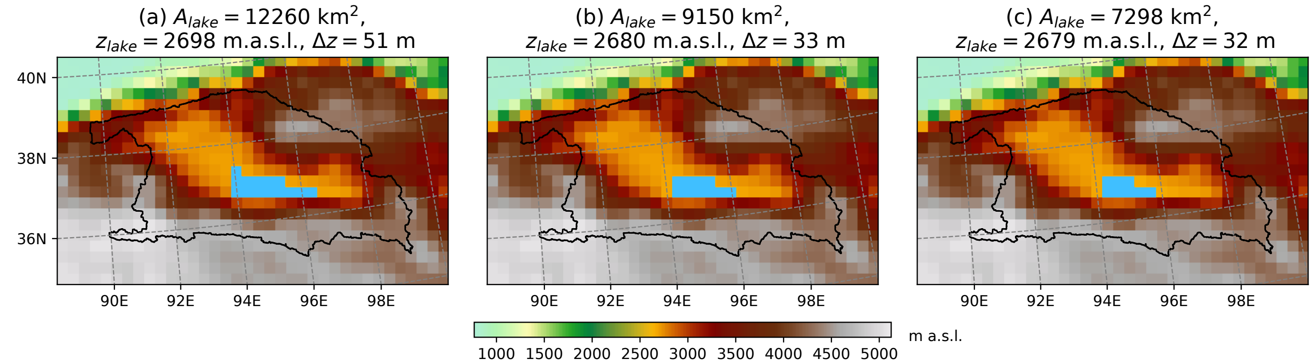

Figure 4.2: *Illustrations of lake extents of equilibrium lake states using the present-day model topography from WRF for accumulation of net precipitation and subsequent runoff originating from land in areas at or below equilibrium lake levels (marked in blue). The equilibrium lake extent (A_{lake}), lake level (z_{lake}), and the rise in lake level (Δz) were estimated using 26 mm a^{-1} as input change in mean annual ΔS and (a) 600 mm a^{-1}, (b) 800 mm a^{-1}, (c) 1000 mm a^{-1} as input lake evaporation. Figure taken from Paper IV.*

4.4 Large-scale systems controlling water balance in the Qaidam Basin

On the climatological scale, ΔS and net moisture transport into and out of an endorheic basin through its lateral boundary are in balance (Brubaker et al., 1993). Due to the irregular shape of the QB, a simple rectangle covering the QB (hereafter referred to as Qaidam box) was defined to perform moisture budget analysis on the Atmospheric Water Transport

(AWT) across the four borders (blue rectangle in Figure 4.1b). The AWT was calculated following the method described in Curio et al. (2015). Table 4.2 presents the seasonal and annual AWT through each border of the Qaidam box, as well as the sum of AWT from all borders, i.e., the atmospheric moisture budget, converted to the theoretical precipitation amount. Figure 4.3 illustrates the route and the magnitude of seasonal moisture propagation across the whole domain.

Table 4.2: *15-year average of seasonal and annual atmospheric water flux converted to theoretical precipitation amount (mm season^{-1} or mm a^{-1}) through each border of the Qaidam box (blue rectangle in Figure 4.1b). Positive values indicate moisture input into the QB, while negative values represent moisture output. Table taken from Paper IV.*

	PLIO					PD					PLIO-PD				
	West	East	South	North	Sum	West	East	South	North	Sum	West	East	South	North	Sum
DJF	327	-298	-38	28	19	277	-248	-7	-2	19	50	-50	-31	30	0
MAM	433	-399	50	-49	36	390	-368	19	-14	28	43	-31	31	-35	8
JJA	344	-283	120	-174	8	407	-459	159	-95	12	-63	176	-39	-79	-4
SON	415	-414	71	-79	-7	368	-394	95	-80	-11	47	-20	-24	1	4
Annual	1519	-1393	204	-274	56	1441	-1469	267	-191	48	78	76	-63	-83	8

The western and eastern borders of the Qaidam box serve as the dominant moisture input and output channel in both PD and PLIO. The higher annual moisture budget in PLIO derives from (1) the increased moisture influx across the western border, and (2) the decreased moisture outflux at the eastern border (Table 4.2). This indicates that moisture budget and ΔS in the QB are related to the large-scale systems that influence the AWT at the western and eastern borders. In PLIO, the increased moisture influx at the western border occurs in winter, spring, and autumn, while a strong reduction of moisture export at the eastern border occurs in summer.

The AWT at both the western and eastern borders of the Qaidam Box is under the influence of the mid-latitude westerlies throughout the year (Figure 4.3). Following the methods used by Sun et al. (2020), we defined a strength index as the average maximum zonal wind speed at 200 hPa at each longitude over the Qaidam box to quantify the strength of the westerlies over the QB. In PLIO, the westerlies over the QB are stronger in all seasons except for summer (Figure 4.4), which explains the larger AWT through the western and eastern borders (Table 4.2).

In the summer months, the moisture transport over eastern Asia by the East Asian Summer Monsoon (EASM) is stronger in PLIO (Figure 4.3g-i). Analysis of the daily zonal water transport on the eastern border of the Qaidam box shows moisture input from the eastern border into the QB in PLIO from the middle of July to the beginning of August (Figure 4.5), which indicates the influence of the EASM in PLIO. This pattern can not be observed in PD. The additional moisture input by the EASM in PLIO also contributes to the lower

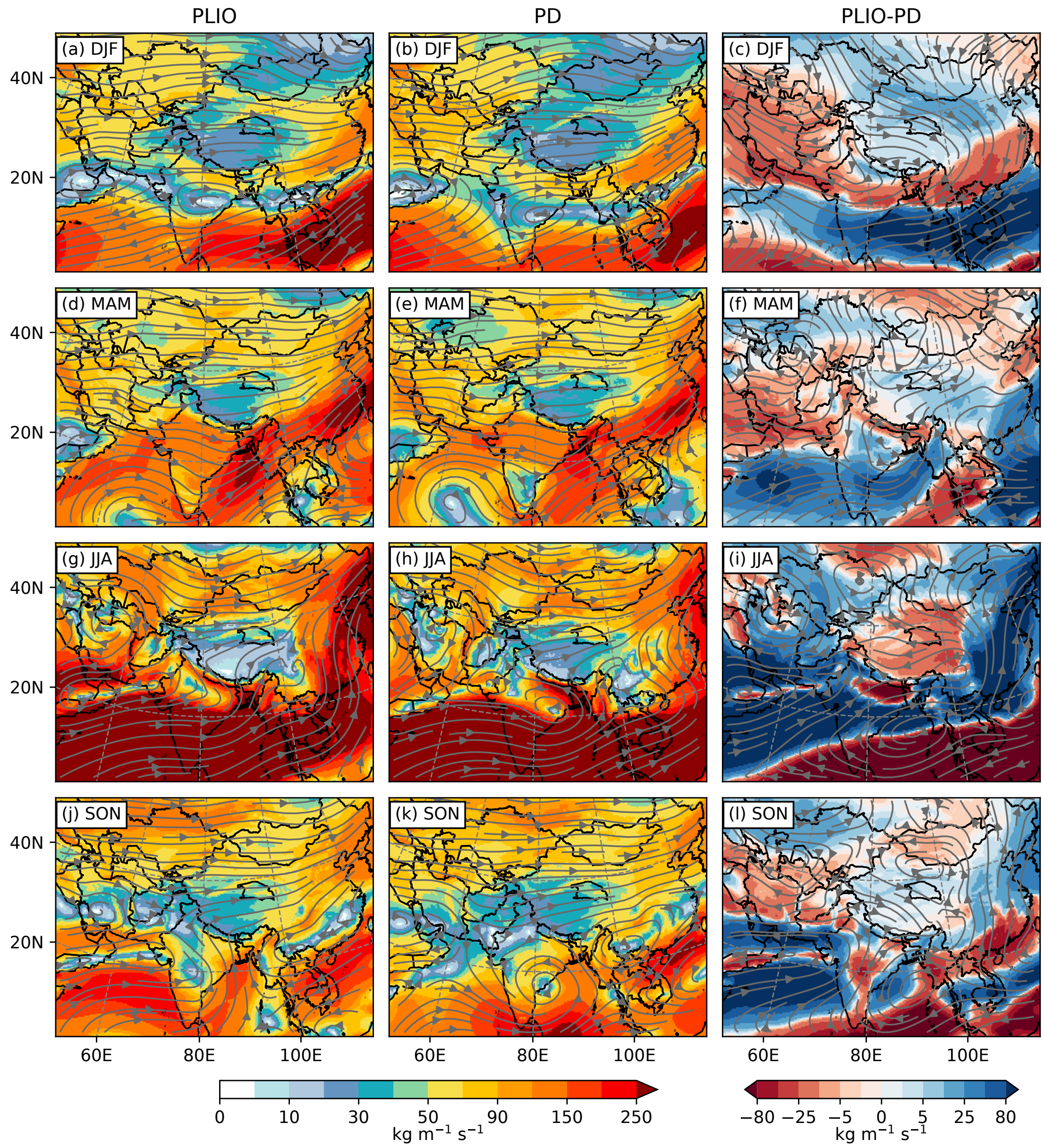

Figure 4.3: *15-year mean seasonal AWT (kg m^{-1} s^{-1}) in DJF (a, b, c), MAM (d, e, f), JJA (g, h, i), and SON (j, k, l) for PLIO (a, d, g, j), PD (b, e, h, k), and the difference between PLIO and PD (c, f, i, j). Colors represent the strength of water flux; arrows indicate transport direction. Figure taken from Paper IV .*

values for total moisture output at the eastern border in summer. As a major component of the subtropical EASM (Huang et al., 2019), the Northwest Pacific Subtropical High (NPSH) intensifies and extends westwards in PLIO, indicated by higher geopotential and stronger anti-cyclonic circulation over eastern China (Figure 4.6a). The strengthening of the EASM in PLIO is coupled with a tilted jet stream axis over northeastern China. To demonstrate the influence of the EASM on the QB, we selected the period from 17 July to 27 July, when the daily AWT averaged over all the grid points at the eastern border of the Qaidam Box is negative in PLIO (i.e., the eastern border receives moisture input on average). During this period, the NPSH intensifies and extends further northwestwards in both PLIO and PD (Figure 4.6d-f), as compared to the climatology in summer (Figure 4.6a-c). In PLIO, the southern part of the QB is clearly under the control of easterly winds during this period (Figure 4.6d).

In summary, the strengthening of the mid-latitude westerlies in all seasons, except for summer, and the intensification of the EASM lead to higher moisture budget and higher ΔS in the QB in PLIO.

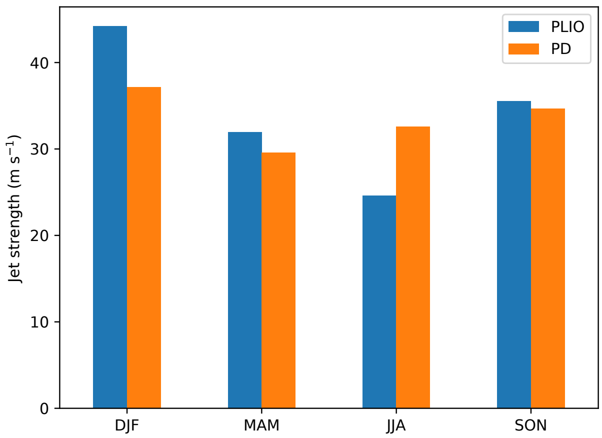

Figure 4.4: *15-year average of seasonal strength index of jet stream ($m\,s^{-1}$) over the QB for PLIO and PD, calculated as defined in Sun et al. (2020). Figure taken from Paper IV.*

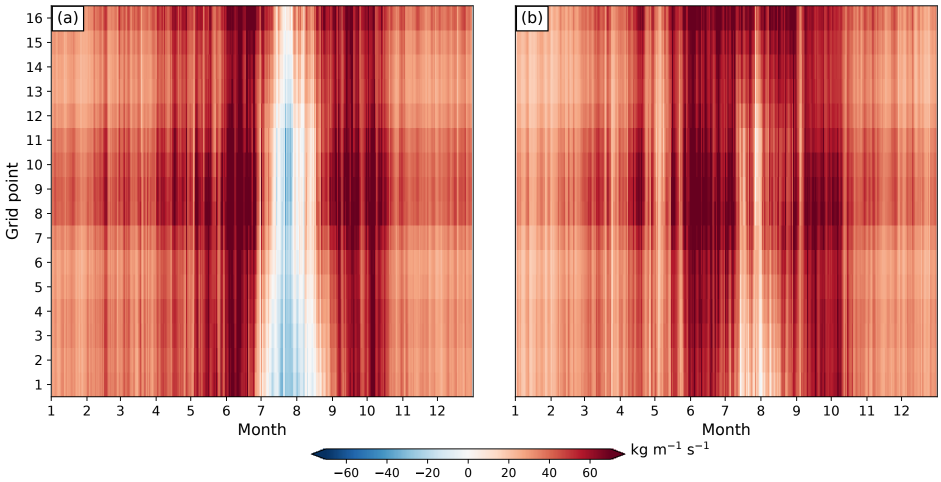

Figure 4.5: *15-year average of daily zonal AWT (kg m^{-1} s^{-1}) of 16 grid points at the eastern border of the Qaidam box (blue rectangle in Figure 4.1b) for (a) PLIO and (b) PD. Grid point 1 and grid point 16 represent the south-most and north-most grid points. Reddish colors indicate water transport away from the QB, while blueish colors indicate water transport towards the QB. Figure taken from Paper IV.*

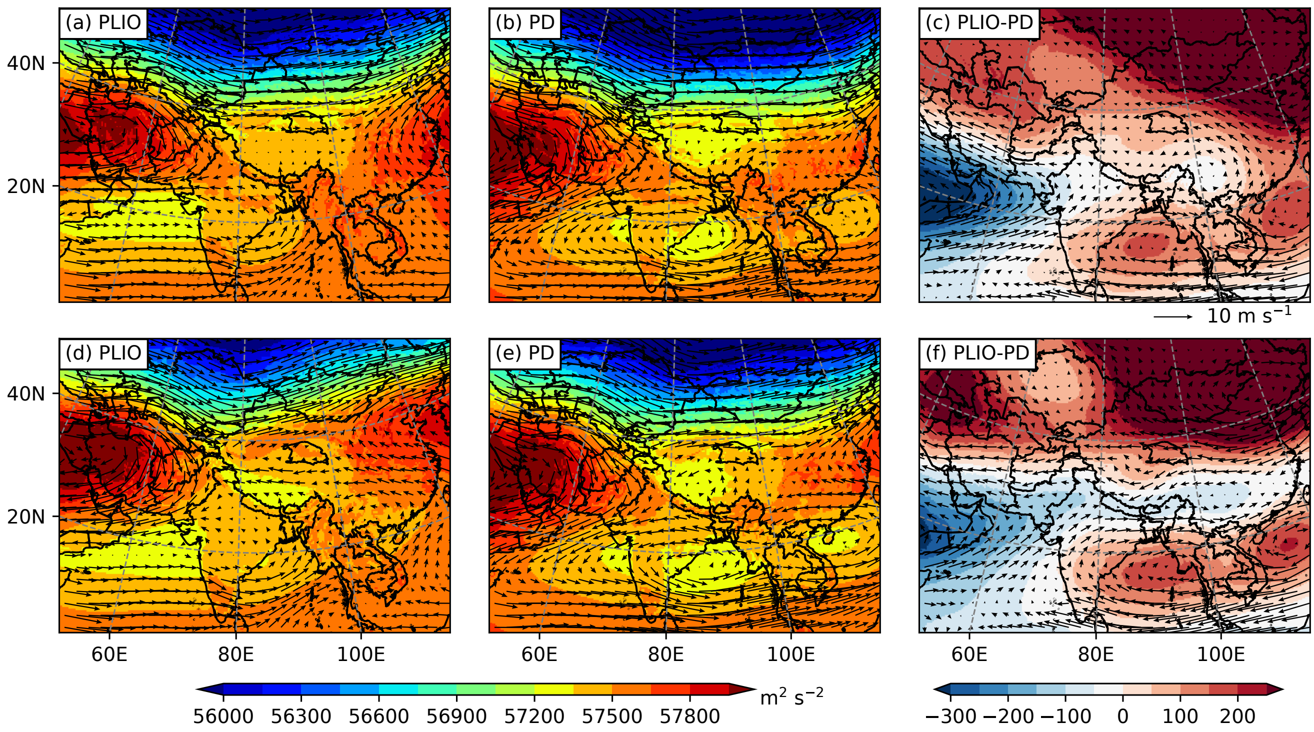

Figure 4.6: *15-year climatology of geopotential (shading, m^{2} s^{-2}) and wind field (arrows, m s^{-1}) at 500 hPa averaged in JJA (a, b, c) and from 17 July to 27 July (d, e, f) for PLIO, PD, and PLIO-PD. Figure taken from Paper IV.*

5 Conclusions and outlook

In this thesis, dynamical downscaling based on WRF was employed over the HMA region to overcome the data paucity problem. The newly developed WRF-downscaled data was utilized to provide an insight into atmospheric triggering mechanisms of landslide activities. Furthermore, the dynamical downscaling method was applied in the paleoclimate context to examine the mechanisms of regional water balance changes.

A new version of the HAR with a larger 10 km domain and longer temporal coverage was developed. ERA5 reanalysis data set was chosen as the forcing data. As the first step, sensitivity experiments regarding the nesting strategies, PPSs, and initial snow depth were conducted to optimize the model configuration for the HAR v2 in terms of P and T_2 (Paper I). Validation against publicly available in-situ observations from GSOD from 2000 to 2020 shows that the HAR v2 can reproduce the seasonal and annual variability of P and T_2 very well over HMA and fits better with observations than its forcing data ERA5. The HAR v2 shows an overall cold bias and wet bias over the HMA region. Station-wise biases exhibit regional and seasonal patterns. Due to the long validation period (21 years), only 89 stations with more than 90% records left for validation over the whole HMA. Some areas of HMA, such as the western TP, the western Tien Shan, and the Pamir, are not covered with available GSOD stations, and thus, the validation presented here does not show the full aspect of the performance of the HAR v2.

Precipitation is a central quantity for hydrometeorological researches. Several gridded precipitation products derived from different sources are available over HMA. Paper II presents an intercomparison of commonly used gridded precipitation products, including WRF-downscaled products (HAR v2 10 km and HAR v2 2 km), reanalysis data, satellite retrieval, and interpolated in-situ observations. Results reveal that these products show general agreement in precipitation variability, but there exist considerable differences among products spatially and temporally. Thus, one should carefully select the product based on the type of application. For applications requiring regional spatial details, such as glacier mass-balance modeling, high-resolution HAR v2 10 km and HAR v2 2 km are most suitable.

By the time of writing, the data production of the HAR v2 10 km was finished from 2000 to 2020, and the data was made publicly available at: www.klima.tu-berlin.de/HARv2. It is planned to generate the HAR v2 10 km data back to 1979. In addition, the HAR v2 2 km used in Paper II has been extended to the period from 2017 to 2020 and will be utilized to investigate GLOFs in the Halji region in northwestern Nepal (Figure 5.1). The Halji village is home to one of the oldest Tibetan Buddhist monasteries of Nepal. From 2004 to 2014, the village has been severely affected by six GLOFs from a glacial lake, which formed episodically during the snowmelt season in the Halji Glacier (Kropáček et al., 2015). The "Glacial lake outburst floods in the Halji region, Nepal" project funded by the Deutsche Forschungsgemeinschaft (DFG) will explore the preconditions of GLOFs, the controlling factor of the temporal variability of GLOFs, and the linkage of GLOF occurrence with large-scale climate change. The research will be carried out by combining numerical simulations with remote sensing analysis and field observations. In particular, the HAR v2 data will be compared to field measurements to assess its uncertainty and will be analyzed to determine the triggers and large-scale controls of GLOFs. It will also be applied to force glacier mass-balance simulations (Sauter et al., 2020; Arndt et al., 2021) and snowdrift simulations to quantify meltwater production rates and to investigate the blocking effect of snow on the outlet of the glacial lake basin.

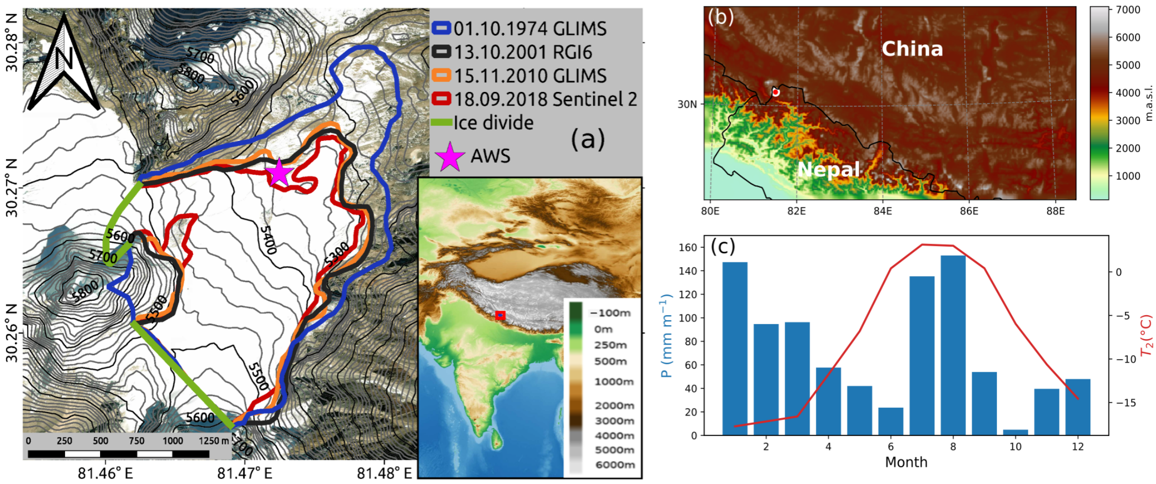

Figure 5.1: (a) Overview of the Halji glacier with multi-temporal glacier outlines, figure taken from Arndt et al. (2021); (b) domain and topography of HAR v2 2 km; red point indicates the location of the Halji Glacier; (c) monthly P and T_2 at the Halji Glacier extracted from the nearest grid point of the HAR v2 2 km averaged over 2017-2020.

As a newly developed data set, the field of applications for the HAR v2 is rich but still unexplored. Paper III demonstrates the application of the HAR v2 in landslide prediction. Combining the HAR v2 with historical landslide events in Kyrgyzstan and Tajikistan, the atmospheric triggers of landslides in this region were determined, and the triggering thresholds were defined. Mean annual exceedance maps derived from the defined thresholds display the climatic disposition and have added values in landslide susceptibility mapping. The majority of studies in this field used precipitation data from rain gauges, which are rarely located at landslide initial points and can lead to considerable uncertainty (Nikolopoulos et al., 2014; Nikolopoulos et al., 2015; Marra et al., 2016). Different interpolation methods have been applied to extract precipitation information from rain gauges (Nikolopoulos et al., 2015). Using gridded data can avoid this allocation problem, but uncertainty still exists since gridded data only represent the grid-mean value but not the "true" weather condition at landslide sites. Nevertheless, it is still essential that the gridded data can accurately represent the grid-mean value. The careful selections of WRF model configurations in Paper I ensure the quality of the HAR v2. Intercomparison results from Paper II also indicate the suitability of the HAR v2 in regional researches.

Most regional climate simulations have been performed at a grid spacing of 10 km or coarser, which cannot resolve convective processes explicitly. In recent years, more attention has been paid to convection-permitting scale (a few km grid spacing) simulations over HMA (e.g., Lin et al., 2018; Ou et al., 2020; Cai et al., 2021; Li et al., 2021; Zhou et al., 2021). Convection-permitting scale simulations show improvement over simulations applying cumulus parameterization schemes in several aspects. For example, they can (1) reproduce the timing of precipitation peaks more accurately (Ou et al., 2020; Zhou et al., 2021); (2) reduce wet bias over the TP due to a more realistic representation of topography (Lin et al., 2018); (3) capture more extreme events as shown in Paper II. In a feasibility study, three WRF simulations (with 30 km, 10 km, and 2 km spatial resolution) were generated for one week before a landslide event occurred in the Jalal-Abad region in Kyrgyzstan on 2011-05-11. Results reveal that the 2 km WRF simulation captures more accumulated precipitation than simulations with coarser resolutions (Figure 5.2). This implies that applying finer resolution atmospheric data could lead to a higher triggering threshold of landslides and consequently a lower FAR. However, simulations with a finer spatial resolution have a lower tolerance to the uncertainty in the landslide location. Landslide information from the GLC and the GFLD used in Paper III was primarily derived from media reports. Most media reports just recorded the location affected by landslides rather than the landslide initiation point. The potential of a kilometer-scale simulation cannot be realized if the landslide location uncertainty is larger than the grid size. Therefore, for Kyrgyzstan and Tajikistan, future studies should not only focus on acquiring high-resolution and high-quality atmospheric data but also on developing landslide inventories with higher location accuracy.

Using the same model set-up as the HAR v2, Paper IV addresses the applicability of dynamical downscaling in the context of paleoclimate. Here, global climate simulations

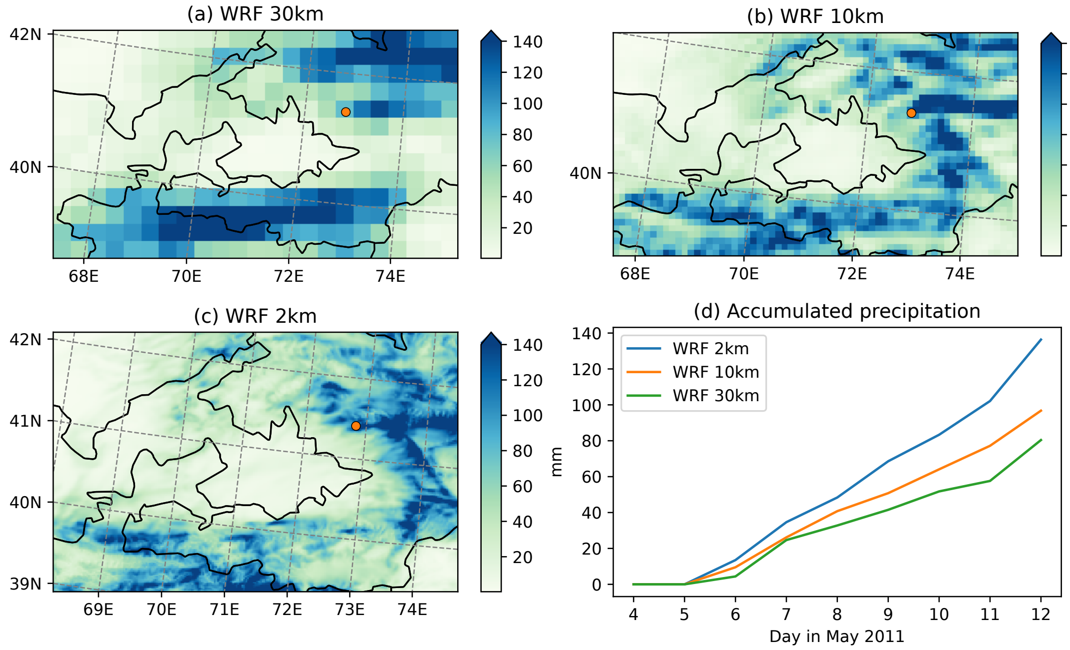

Figure 5.2: A feasibility study of convection-permitting scale simulation in determine landslide triggering conditions. Spatial distribution of accumulated precipitation from 2011-05-04 to 2011-05-11 from WRF simulations at a grid spacing of 30 km (a), 10 km (b), and 2 km (c); Orange point indicates the location of a landslide event occurred on 2011-05-11 recorded by the GLC. Accumulated precipitation at the nearest grid point of the landslide event from three WRF simulations (d).

for the present day and the mid-Pliocene from ECHAM5 were dynamically downscaled to 30 km resolution over the HMA region. By keeping the land surface condition the same, this study was able to isolate the influence of large-scale climate states and to reveal its role on the maintenance of the Qaidam mega-lake system during the Pliocene. A strengthening of the mid-latitude westerlies in all seasons, except for summer, and an intensification of the EASM was found under the mid-Pliocene climate. These changes in large-scale climate would increase ΔS in the QB with its modern land surface settings, which would lead to a readjustment of lake extent until a new equilibrium state is reached. The equilibrium lake extents derived from this imbalance in ΔS take up 12%-21% of the maximum lake extent approximated from proxy data. This implies non-climatic factors also contributed to the maintenance of the Qaidam mega-lake system during the Pliocene. The mid-Pliocene is considered a past analog of the near-future climate (Burke et al., 2018) since both periods feature higher CO_2 concentrations compared to pre-industrial values. Several studies also suggest similarities of large-scale circulations in these two periods, such as the intensification of EASM (Sun et al., 2018). Thus, Paper IV also contributes to a better understanding of lake development in the future in HMA.

The downscaling simulation presented in Paper IV cannot be considered reconstruction of regional climate in the QB during the mid-Pliocene. Reconstruction at a grid spacing of 30 km would require Pliocene surface conditions, such as topography and land cover, at an equivalent or even higher resolution, which are not available to date. The most commonly used surface boundary conditions for mid-Pliocene simulations are from the US Geological Survey's Pliocene Research Interpretation and Synoptic Mapping (PRISM) project (Dowsett et al., 1994). The most recently released PRISM data set has a $1° \times 1°$ grid spacing (Dowsett et al., 2016), which is too coarse for regional climate simulations at a grid spacing of 30 km. Future regional paleoclimate studies could focus on the impact of land surface conditions on the regional ΔS and reconstruction of regional climate in the QB, once high-resolution reconstructions of mid-Pliocene surface boundary conditions are available.

In summary, this thesis illustrates several aspects regarding dynamical downscaling over the HMA region. Atmospheric data generated by dynamical downscaling provides physically based information at spatial scales of relevance for regional climate studies and contributes to a better understanding of atmospheric related processes at different temporal scales (from mean climate conditions to extreme events). This thesis, in particular, demonstrates the application and future potential of downscaling products in the field of atmospherically triggered natural hazards. Furthermore, the dynamical downscaling technique has been proven to be a valuable tool for regional climate studies across different time scales (past, present, and future).

List of References

Arndt, A., D. Scherer, and C. Schneider (2021): Atmosphere Driven Mass-Balance Sensitivity of Halji Glacier, Himalayas. *Atmosphere* 12 (4), 426 (cit. on p. 34).

Bao, J., J. Feng, and Y. Wang (2015): Dynamical downscaling simulation and future projection of precipitation over China. *Journal of Geophysical Research: Atmospheres* 120 (16), 8227–8243 (cit. on p. 4).

Bazai, N. A., P. Cui, P. A. Carling, H. Wang, J. Hassan, D. Liu, G. Zhang, and J. Wen (2020): Increasing glacial lake outburst flood hazard in response to surge glaciers in the Karakoram. *Earth-Science Reviews*, 103432 (cit. on p. 2).

Berg, L. K., W. I. Gustafson Jr, E. I. Kassianov, and L. Deng (2013): Evaluation of a modified scheme for shallow convection: Implementation of CuP and case studies. *Monthly weather review* 141 (1), 134–147 (cit. on p. 11).

Berti, M., M. Martina, S. Franceschini, S. Pignone, A. Simoni, and M. Pizziolo (2012): Probabilistic rainfall thresholds for landslide occurrence using a Bayesian approach. *Journal of Geophysical Research: Earth Surface* 117 (F4) (cit. on p. 3).

Biskop, S. (2016): Advancing the understanding for hydro-climatic controls on water balance and lake-level variability in the Tibetan Plateau: Hydrological modeling in data-scarce lake basins integrating multi-source data. *phdthesis* (cit. on p. 4).

Bolch, T., A. Kulkarni, A. Kääb, C. Huggel, F. Paul, J. G. Cogley, H. Frey, J. S. Kargel, K. Fujita, M. Scheel, et al. (2012): The state and fate of Himalayan glaciers. *Science* 336 (6079), 310–314 (cit. on p. 1).

Bonekamp, P., E. Collier, and W. Immerzeel (2018): The impact of spatial resolution, land use, and spinup time on resolving spatial precipitation patterns in the Himalayas. *Journal of Hydrometeorology* 19 (10), 1565–1581 (cit. on p. 12).

Botsyun, S., T. Ehlers, S. Mutz, K. Methner, E. Krsnik, and A. Mulch (2020): Opportunities and challenges for paleoaltimetry in "small" orogens: Insights from the European Alps. *Geophys. Res. Lett.* ISSN: 0094-8276 (cit. on pp. 4, 27).

Braun, A., T. Fernandez-Steeger, H.-B. Havenith, and A. Torgoev (2015): Landslide Susceptibility Mapping with Data Mining Methods—a Case Study from Maily-Say, Kyrgyzstan. In: *Engineering Geology for Society and Territory-Volume 2*. Springer, 995–998 (cit. on p. 19).

Broecker, W. (2010): Long-term water prospects in the western United States. *Journal of Climate* 23 (24), 6669–6683 (cit. on p. 26).

Brubaker, K. L., D. Entekhabi, and P. Eagleson (1993): Estimation of continental precipitation recycling. *Journal of Climate* 6 (6), 1077–1089 (cit. on p. 28).

Burke, K. D., J. W. Williams, M. A. Chandler, A. M. Haywood, D. J. Lunt, and B. L. Otto-Bliesner (2018): Pliocene and Eocene provide best analogs for near-future climates. *Proc. Natl. Acad. Sci. U. S. A.* 115 (52), 13288–13293. ISSN: 10916490 (cit. on pp. 26, 36).

Cai, S., A. Huang, K. Zhu, B. Yang, X. Yang, Y. Wu, and X. Mu (2021): Diurnal cycle of summer precipitation over the Eastern Tibetan Plateau and surrounding regions simulated in a convection-permitting model. *Climate Dynamics*, 1–22 (cit. on p. 35).

Carrivick, J. L. and F. S. Tweed (2016): A global assessment of the societal impacts of glacier outburst floods. *Global and Planetary Change* 144, 1–16 (cit. on p. 2).

Chen, K. and J. M. Bowler (1986): Late pleistocene evolution of salt lakes in the Qaidam basin, Qinghai province, China. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 54 (1-4), 87–104. ISSN: 00310182 (cit. on pp. 25, 28).

Cogley, J. G. (2016): Glacier shrinkage across High Mountain Asia. *Annals of Glaciology* 57 (71), 41–49 (cit. on p. 2).

Curio, J., F. Maussion, and D. Scherer (2015): A 12-year high-resolution climatology of atmospheric water transport over the Tibetan Plateau. *Earth System Dynamics* 6 (1), 109–124 (cit. on pp. 4, 29).

Curio, J. and D. Scherer (2016): Seasonality and spatial variability of dynamic precipitation controls on the Tibetan Plateau. *Earth System Dynamics* 7 (3), 767–782 (cit. on p. 4).

Dai, F., C. Lee, and Y. Y. Ngai (2002): Landslide risk assessment and management: an overview. *Engineering geology* 64 (1), 65–87 (cit. on p. 23).

Dee, D. P., S. M. Uppala, A. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. Balmaseda, G. Balsamo, d. P. Bauer, et al. (2011): The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society* 137 (656), 553–597 (cit. on p. 3).

Di Luca, A., D. Argüeso, J. P. Evans, R. de Elıa, and R. Laprise (2016): Quantifying the overall added value of dynamical downscaling and the contribution from different spatial scales. *Journal of Geophysical Research: Atmospheres* 121 (4), 1575–1590 (cit. on p. 4).

Dietz, A. J., C. Conrad, C. Kuenzer, G. Gesell, and S. Dech (2014): Identifying changing snow cover characteristics in central Asia between 1986 and 2014 from remote sensing data. *Remote Sensing* 6 (12), 12752–12775 (cit. on p. 19).

Dosio, A., H.-J. Panitz, M. Schubert-Frisius, and D. Lüthi (2015): Dynamical downscaling of CMIP5 global circulation models over CORDEX-Africa with COSMO-CLM: evaluation over the present climate and analysis of the added value. *Climate Dynamics* 44 (9), 2637–2661 (cit. on p. 4).

Dowsett, H., A. Dolan, D. Rowley, R. Moucha, A. M. Forte, J. X. Mitrovica, M. Pound, U. Salzmann, M. Robinson, M. Chandler, et al. (2016): The PRISM4 (mid-Piacenzian) paleoenvironmental reconstruction. *Climate of the Past* 12 (7), 1519–1538 (cit. on p. 37).

Dowsett, H., M. Robinson, A. Haywood, U. Salzmann, D. Hill, L. Sohl, M. Chandler, M. Williams, K. Foley, and D. Stoll (2010): The PRISM3D paleoenvironmental reconstruction. *Stratigraphy* 7 (2-3), 123–139 (cit. on p. 26).

Dowsett, H., R. Thompson, J. Barron, T. Cronin, F. Fleming, S. Ishman, R. Poore, D. Willard, and T. Holtz Jr (1994): Joint investigations of the Middle Pliocene climate I: PRISM paleoenvironmental reconstructions. *Global and Planetary Change* 9 (3-4), 169–195 (cit. on p. 37).

Dudhia, J. (1989): Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *Journal of the atmospheric sciences* 46 (20), 3077–3107 (cit. on p. 11).

Feser, F., B. Rockel, H. von Storch, J. Winterfeldt, and M. Zahn (2011): Regional climate models add value to global model data: a review and selected examples. *Bulletin of the American Meteorological Society* 92 (9), 1181–1192 (cit. on p. 3).

Froude, M. J. and D. N. Petley (2018): Global fatal landslide occurrence from 2004 to 2016. *Natural Hazards and Earth System Sciences* 18 (8), 2161–2181 (cit. on p. 19).

Gao, Y., L. Cuo, and Y. Zhang (2014): Changes in moisture flux over the Tibetan Plateau during 1979–2011 and possible mechanisms. *Journal of Climate* 27 (5), 1876–1893 (cit. on p. 1).

Gao, Y., L. Xiao, D. Chen, F. Chen, J. Xu, and Y. Xu (2017): Quantification of the relative role of land-surface processes and large-scale forcing in dynamic downscaling over the Tibetan Plateau. *Climate Dynamics* 48 (5-6), 1705–1721 (cit. on p. 4).

Gao, Y., J. Xu, and D. Chen (2015): Evaluation of WRF mesoscale climate simulations over the Tibetan Plateau during 1979–2011. *Journal of Climate* 28 (7), 2823–2841 (cit. on pp. 4, 12).

Gariano, S. L., M. T. Brunetti, G. Iovine, M. Melillo, S. Peruccacci, O. Terranova, C. Vennari, and F. Guzzetti (2015): Calibration and validation of rainfall thresholds for shallow landslide forecasting in Sicily, southern Italy. *Geomorphology* 228, 653–665 (cit. on p. 3).

Giannecchini, R., Y. Galanti, G. D. Avanzi, and M. Barsanti (2016): Probabilistic rainfall thresholds for triggering debris flows in a human-modified landscape. *Geomorphology* 257, 94–107 (cit. on p. 3).

Gómez-Navarro, J., J. Montávez, S. Jerez, P. Jiménez-Guerrero, R. Lorente-Plazas, J. F. González-Rouco, and E. Zorita (2011): A regional climate simulation over the Iberian Peninsula for the last millennium. *Climate of the Past* 7 (2), 451–472 (cit. on p. 5).

Gómez-Navarro, J., J. Montávez, S. Wagner, and E. Zorita (2013): A regional climate palaeosimulation for Europe in the period 1500–1990–Part 1: Model validation. *Climate of the Past* 9 (4), 1667–1682 (cit. on p. 5).

Gómez-Navarro, J. J., O. Bothe, S. Wagner, E. Zorita, J. P. Werner, J. Luterbacher, C. C. Raible, and J. Montávez (2015): A regional climate palaeosimulation for Europe in the period 1500–1990–Part 2: Shortcomings and strengths of models and reconstructions. *Climate of the Past* 11 (8), 1077–1095 (cit. on p. 5).

Haginoya, S., H. Fujii, T. Kuwagata, J. Xu, Y. Ishigooka, S. Kang, and Y. Zhang (2009): Air-lake interaction features found in heat and water exchanges over Nam Co on the Tibetan Plateau. *Sola* 5, 172–175 (cit. on p. 28).

Havenith, H.-B., A. Torgoev, R. Schlögel, A. Braun, I. Torgoev, and A. Ischuk (2015): Tien Shan geohazards database: Landslide susceptibility analysis. *Geomorphology* 249, 32–43 (cit. on p. 19).

Haywood, A. M., H. J. Dowsett, A. M. Dolan, D. Rowley, A. Abe-Ouchi, B. Otto-Bliesner, M. A. Chandler, S. J. Hunter, D. J. Lunt, M. Pound, et al. (2016): The Pliocene model intercomparison project (PlioMIP) phase 2: scientific objectives and experimental design. *Climate of the Past* 12 (3), 663–675 (cit. on p. 4).

Hersbach, H., B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, et al. (2020): The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* 146 (730), 1999–2049 (cit. on p. 9).

Hong, S.-Y. and M. Kanamitsu (2014): Dynamical downscaling: Fundamental issues from an NWP point of view and recommendations. *Asia-Pacific Journal of Atmospheric Sciences* 50 (1), 83–104 (cit. on p. 4).

Hong, S.-Y., Y. Noh, and J. Dudhia (2006): A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly weather review* 134 (9), 2318–2341 (cit. on p. 11).

Huang, D. and S. Gao (2018): Impact of different reanalysis data on WRF dynamical downscaling over China. *Atmospheric Research* 200, 25–35 (cit. on p. 10).

Huang, X., D. Jiang, X. Dong, S. Yang, B. Su, X. Li, Z. Tang, and Y. Wang (2019): Northwestward Migration of the Northern Edge of the East Asian Summer Monsoon During the Mid-Pliocene Warm Period: Simulations and Reconstructions. *Journal of Geophysical Research: Atmospheres* 124 (3), 1392–1404 (cit. on p. 31).

Huintjes, E., T. Sauter, B. Schröter, F. Maussion, W. Yang, J. Kropáček, M. Buchroithner, D. Scherer, S. Kang, and C. Schneider (2015): Evaluation of a coupled snow and energy balance model for Zhadang glacier, Tibetan Plateau, using glaciological measurements and time-lapse photography. *Arctic, antarctic, and alpine research* 47 (3), 573–590 (cit. on p. 4).

Ibarra, D. E., J. L. Oster, M. J. Winnick, J. K. Caves Rugenstein, M. P. Byrne, and C. P. Chamberlain (2018): Warm and cold wet states in the western United States during the Pliocene–Pleistocene. *Geology* 46 (4), 355–358 (cit. on p. 26).

Ilyasov, S., O. Zabenko, N. Gaydamak, A. Kirilenko, N. Myrsaliev, V. Shevchenko, and L. Penkina (2013): Climate profile of the Kyrgyz Republic. *The State Agency for Environmental Protection and Forestry under the Government of the Kyrgyz Republic and The United Nations Development Programme: Bishkek, Kyrgyzstan* (cit. on p. 3).

Immerzeel, W. W., L. P. Van Beek, and M. F. Bierkens (2010): Climate change will affect the Asian water towers. *Science* 328 (5984), 1382–1385 (cit. on p. 1).

Jacob, D., J. Petersen, B. Eggert, A. Alias, O. B. Christensen, L. M. Bouwer, A. Braun, A. Colette, M. Déqué, G. Georgievski, et al. (2014): EURO-CORDEX: new high-resolution climate change projections for European impact research. *Regional environmental change* 14 (2), 563–578 (cit. on p. 4).

Jiang, Y., K. Yang, C. Shao, X. Zhou, L. Zhao, Y. Chen, and H. Wu (2021): A downscaling approach for constructing high-resolution precipitation dataset over the Tibetan Plateau from ERA5 reanalysis. *Atmospheric Research* 256, 105574 (cit. on pp. 3, 15).

Jiao, J. J., X. Zhang, Y. Liu, and X. Kuang (2015): Increased water storage in the Qaidam Basin, the North Tibet Plateau from GRACE gravity data. *PLoS One* 10 (10), e0141442 (cit. on p. 26).

Jiménez, P. A., J. Dudhia, J. F. González-Rouco, J. Navarro, J. P. Montávez, and E. Garcıa-Bustamante (2012): A revised scheme for the WRF surface layer formulation. *Monthly Weather Review* 140 (3), 898–918 (cit. on p. 11).

Kala, J., J. Andrys, T. J. Lyons, I. J. Foster, and B. J. Evans (2015): Sensitivity of WRF to driving data and physics options on a seasonal time-scale for the southwest of Western Australia. *Climate Dynamics* 44 (3-4), 633–659 (cit. on p. 10).

Karki, R., L. Gerlitz, U. Schickhoff, T. Scholten, J. Böhner, et al. (2017): Quantifying the added value of convection-permitting climate simulations in complex terrain: a systematic evaluation of WRF over the Himalayas. *Earth System Dynamics* 8 (3), 507–528 (cit. on p. 12).

Karki, R., S. ul Hasson, L. Gerlitz, R. Talchabhadel, E. Schenk, U. Schickhoff, T. Scholten, and J. Böhner (2018): WRF-based simulation of an extreme precipitation event over the Central Himalayas: Atmospheric mechanisms and their representation by microphysics parameterization schemes. *Atmospheric Research* 214, 21–35 (cit. on pp. 4, 18).

Kirschbaum, D., S. Kapnick, T. Stanley, and S. Pascale (2020): Changes in extreme precipitation and landslides over High Mountain Asia. *Geophysical Research Letters* 47 (4), e2019GL085347 (cit. on p. 2).

Kirschbaum, D., T. Stanley, and Y. Zhou (2015): Spatial and temporal analysis of a global landslide catalog. *Geomorphology* 249, 4–15 (cit. on p. 2).

Kirschbaum, D., C. S. Watson, D. R. Rounce, D. Shugar, J. S. Kargel, U. K. Haritashya, P. Amatya, D. Shean, E. R. Anderson, and M. Jo (2019): The State of Remote Sensing Capabilities of Cascading Hazards Over High Mountain Asia. *Frontiers in Earth Science* 7, 197 (cit. on p. 2).

Kirschbaum, D. B., R. Adler, Y. Hong, S. Hill, and A. Lerner-Lam (2010): A global landslide catalog for hazard applications: method, results, and limitations. *Natural Hazards* 52 (3), 561–575 (cit. on p. 2).

Kitoh, A., S. Murakami, and H. Koide (2001): A simulation of the Last Glacial Maximum with a coupled atmosphere-ocean GCM. *Geophysical Research Letters* 28 (11), 2221–2224 (cit. on p. 4).

Kobayashi, S., Y. Ota, Y. Harada, A. Ebita, M. Moriya, H. Onoda, K. Onogi, H. Kamahori, C. Kobayashi, H. Endo, et al. (2015): The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan. Ser. II* 93 (1), 5–48 (cit. on p. 12).

Kropáček, J., N. Neckel, B. Tyrna, N. Holzer, A. Hovden, N. Gourmelen, C. Schneider, M. Buchroithner, and V. Hochschild (2015): Repeated glacial lake outburst flood threatening the oldest Buddhist monastery in north-western Nepal. *Natural Hazards and Earth System Sciences* 15 (10), 2425–2437 (cit. on p. 34).

Lazhu, K. Yang, J. Wang, Y. Lei, Y. Chen, L. Zhu, B. Ding, and J. Qin (2016): Quantifying evaporation and its decadal change for Lake Nam Co, central Tibetan Plateau. *Journal of Geophysical Research: Atmospheres* 121 (13), 7578–7591 (cit. on p. 28).

Leduc, M. and R. Laprise (2009): Regional climate model sensitivity to domain size. *Climate Dynamics* 32 (6), 833–854 (cit. on p. 10).

Lehner, B. and G. Grill (2013): Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes* 27 (15), 2171–2186 (cit. on p. 26).

Lei, Y., T. Yao, B. W. Bird, K. Yang, J. Zhai, and Y. Sheng (2013): Coherent lake growth on the central Tibetan Plateau since the 1970s: Characterization and attribution. *Journal of Hydrology* 483, 61–67 (cit. on p. 1).

Leonarduzzi, E., P. Molnar, and B. W. McArdell (2017): Predictive performance of rainfall thresholds for shallow landslides in Switzerland from gridded daily data. *Water Resources Research* 53 (8), 6612–6625 (cit. on pp. 3, 22).

Li, D., K. Yang, W. Tang, X. Li, X. Zhou, and D. Guo (2020): Characterizing precipitation in high altitudes of the western Tibetan plateau with a focus on major glacier areas. *International Journal of Climatology* 40 (12), 5114–5127 (cit. on p. 4).

Li, P., K. Furtado, T. Zhou, H. Chen, and J. Li (2021): Convection-permitting modelling improves simulated precipitation over the central and eastern Tibetan Plateau. *Quarterly Journal of the Royal Meteorological Society* 147 (734), 341–362 (cit. on pp. 4, 35).

Li, X., Y. Ma, Y. Huang, X. Hu, X. Wu, P. Wang, G. Li, S. Zhang, H. Wu, Z. Jiang, et al. (2016): Evaporation and surface energy budget over the largest high-altitude saline lake on the Qinghai-Tibet Plateau. *Journal of Geophysical Research: Atmospheres* 121 (18), 10–470 (cit. on p. 28).

Li, Y., F. Su, D. Chen, and Q. Tang (2019): Atmospheric water transport to the endorheic Tibetan plateau and its effect on the hydrological status in the region. *Journal of Geophysical Research: Atmospheres* 124 (23), 12864–12881 (cit. on p. 26).

Lin, C., D. Chen, K. Yang, and T. Ou (2018): Impact of model resolution on simulating the water vapor transport through the central Himalayas: implication for models' wet bias over the Tibetan Plateau. *Climate dynamics* 51 (9), 3195–3207 (cit. on pp. 4, 35).

Ludwig, P., J. J. Gómez-Navarro, J. G. Pinto, C. C. Raible, S. Wagner, and E. Zorita (2019): Perspectives of regional paleoclimate modeling. *Annals of the New York Academy of Sciences* 1436 (1), 54 (cit. on pp. 4, 5).

Ludwig, P., J. G. Pinto, C. C. Raible, and Y. Shao (2017): Impacts of surface boundary conditions on regional climate model simulations of European climate during the Last Glacial Maximum. *Geophysical Research Letters* 44 (10), 5086–5095 (cit. on p. 5).

Marra, F., E. Nikolopoulos, J. Creutin, and M. Borga (2016): Space–time organization of debris flows-triggering rainfall and its effect on the identification of the rainfall threshold relationship. *Journal of Hydrology* 541, 246–255 (cit. on p. 35).

Maussion, F. (2014): A new atmospheric dataset for High Asia (cit. on p. 12).

Maussion, F., D. Scherer, R. Finkelnburg, J. Richters, W. Yang, and T. Yao (2011): WRF simulation of a precipitation event over the Tibetan Plateau, China–an assessment using remote sensing and ground observations. *Hydrology and Earth System Sciences* 15 (6), 1795–1817 (cit. on p. 4).

Maussion, F., D. Scherer, T. Mölg, E. Collier, J. Curio, and R. Finkelnburg (2014): Precipitation seasonality and variability over the Tibetan Plateau as resolved by the High Asia Reanalysis. *Journal of Climate* 27 (5), 1910–1927 (cit. on p. 4).

Meng, X., S. Lyu, T. Zhang, L. Zhao, Z. Li, B. Han, S. Li, D. Ma, H. Chen, Y. Ao, et al. (2018): Simulated cold bias being improved by using MODIS time-varying albedo in the Tibetan Plateau in WRF model. *Environmental Research Letters* 13 (4), 044028 (cit. on p. 12).

Miao, Y. F., X. M. Fang, F. L. Wu, M. T. Cai, C. H. Song, Q. Q. Meng, and L. Xu (2013): Late Cenozoic continuous aridification in the western Qaidam Basin: evidence from sporopollen records. *Clim. Past* 9 (4), 1863–1877. ISSN: 1814-9332 (cit. on p. 25).

Miguez-Macho, G., G. L. Stenchikov, and A. Robock (2004): Spectral nudging to eliminate the effects of domain position and geometry in regional climate model simulations. *Journal of Geophysical Research: Atmospheres* 109 (D13) (cit. on p. 10).

Mischke, S., Z. Sun, U. Herzschuh, Z. Qiao, and N. Sun (2010): An ostracod-inferred large Middle Pleistocene freshwater lake in the presently hyper-arid Qaidam Basin (NW China). *Quat. Int.* 218 (1-2), 74–85. ISSN: 10406182 (cit. on p. 25).

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough (1997): Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research: Atmospheres* 102 (D14), 16663–16682 (cit. on p. 11).

Mölg, T., F. Maussion, E. Collier, J. C. Chiang, and D. Scherer (2017): Prominent midlatitude circulation signature in High Asia's surface climate during monsoon. *Journal of Geophysical Research: Atmospheres* 122 (23), 12–702 (cit. on p. 4).

Mölg, T., F. Maussion, and D. Scherer (2014): Mid-latitude westerlies as a driver of glacier variability in monsoonal High Asia. *Nature Climate Change* 4 (1), 68–73 (cit. on pp. 1, 4).

Morrison, H., G. Thompson, and V. Tatarskii (2009): Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one-and two-moment schemes. *Monthly weather review* 137 (3), 991–1007 (cit. on p. 11).

Mostbauer, K., R. Kaitna, D. Prenner, and M. Hrachowitz (2018): The temporally varying roles of rainfall, snowmelt and soil moisture for debris flow initiation in a snow-dominated system. *Hydrol. Earth Syst. Sci* 22, 3493–3513 (cit. on p. 19).

Mutz, S. G., T. A. Ehlers, M. Werner, G. Lohmann, C. Stepanek, and J. Li (2018): Estimates of late Cenozoic climate change relevant to Earth surface processes in tectonically active orogens. *Earth Surf. Dyn.* 6 (2), 271–301. ISSN: 2196632X (cit. on pp. 4, 27).

Nikolopoulos, E. I., S. Crema, L. Marchi, F. Marra, F. Guzzetti, and M. Borga (2014): Impact of uncertainty in rainfall estimation on the identification of rainfall thresholds for debris flow occurrence. *Geomorphology* 221, 286–297 (cit. on p. 35).

Nikolopoulos, E., M. Borga, J. Creutin, and F. Marra (2015): Estimation of debris flow triggering rainfall: Influence of rain gauge density and interpolation methods. *Geomorphology* 243, 40–50 (cit. on p. 35).

Orsolini, Y., M. Wegmann, E. Dutra, B. Liu, G. Balsamo, K. Yang, P. d. Rosnay, C. Zhu, W. Wang, R. Senan, et al. (2019): Evaluation of snow depth and snow cover over the Tibetan Plateau in global reanalyses using in situ and satellite remote sensing observations. *The Cryosphere* 13 (8), 2221–2239 (cit. on pp. 11, 12).

Ou, T., D. Chen, X. Chen, C. Lin, K. Yang, H.-W. Lai, and F. Zhang (2020): Simulation of summer precipitation diurnal cycles over the Tibetan Plateau at the gray-zone grid spacing for cumulus parameterization. *Climate Dynamics* 54 (7), 3525–3539 (cit. on pp. 4, 35).

Pan, X., X. Li, K. Yang, J. He, Y. Zhang, and X. Han (2014): Comparison of downscaled precipitation data over a mountainous watershed: A case study in the Heihe River Basin. *Journal of Hydrometeorology* 15 (4), 1560–1574 (cit. on p. 4).

Patricola, C. M. and K. H. Cook (2007): Dynamics of the West African monsoon under mid-Holocene precessional forcing: Regional climate model simulations. *Journal of Climate* 20 (4), 694–716 (cit. on p. 5).

Peres, D. J., A. Cancelliere, R. Greco, and T. A. Bogaard (2018): Influence of uncertain identification of triggering rainfall on the assessment of landslide early warning thresholds. *Natural Hazards and Earth System Sciences* 18 (2), 633–646 (cit. on p. 22).

Pierce, D. W., T. Das, D. R. Cayan, E. P. Maurer, N. L. Miller, Y. Bao, M. Kanamitsu, K. Yoshimura, M. A. Snyder, L. C. Sloan, et al. (2013): Probabilistic estimates of future changes in California temperature and precipitation using statistical and dynamical downscaling. *Climate Dynamics* 40 (3-4), 839–856 (cit. on p. 4).

Pritchard, D. M., N. Forsythe, H. J. Fowler, G. M. O'Donnell, and X.-F. Li (2019): Evaluation of Upper Indus near-surface climate representation by WRF in the high Asia refined analysis. *Journal of Hydrometeorology* 20 (3), 467–487 (cit. on pp. 4, 12).

Prömmel, K., B. Geyer, J. M. Jones, and M. Widmann (2010): Evaluation of the skill and added value of a reanalysis-driven regional simulation for Alpine temperature. *International Journal of Climatology: A Journal of the Royal Meteorological Society* 30 (5), 760–773 (cit. on p. 4).

Qiu, J. (2008): China: the third pole. *Nature News* 454 (7203), 393–396 (cit. on p. 1).

Ravelo, A. C., D. H. Andreasen, M. Lyle, A. O. Lyle, and M. W. Wara (2004): Regional climate shifts caused by gradual global cooling in the Pliocene epoch. *Nature* 429 (6989), 263–267 (cit. on p. 26).

Rieser, A. B., A.-V. Bojar, F. Neubauer, J. Genser, Y. Liu, X.-H. Ge, and G. Friedl (2009): Monitoring Cenozoic climate evolution of northeastern Tibet: stable isotope constraints from the western Qaidam Basin, China. *Int. J. Earth Sci.* 98 (5), 1063–1075. issn: 1437-3254 (cit. on p. 25).

Russo, E. and U. Cubasch (2016): Mid-to-late Holocene temperature evolution and atmospheric dynamics over Europe in regional model simulations. *Climate of the Past* 12 (8), 1645–1662 (cit. on p. 5).

Saponaro, A., M. Pilz, M. Wieland, D. Bindi, B. Moldobekov, and S. Parolai (2015): Landslide susceptibility analysis in data-scarce regions: the case of Kyrgyzstan. *Bulletin of Engineering Geology and the Environment* 74 (4), 1117–1136 (cit. on p. 19).

Sauter, T., A. Arndt, and C. Schneider (2020): COSIPY v1. 3–an open-source coupled snowpack and ice surface energy and mass balance model. *Geoscientific Model Development* 13 (11), 5645–5662 (cit. on p. 34).

Scherer, D. (2020): Survival of the Qaidam mega-lake system under mid-Pliocene climates and its restoration under future climates. *Hydrology and Earth System Sciences* 24 (7), 3835–3850 (cit. on pp. 25–27).

Schiemann, R., D. Lüthi, and C. Schär (2009): Seasonality and interannual variability of the westerly jet in the Tibetan Plateau region. *Journal of climate* 22 (11), 2940–2957 (cit. on p. 1).

Segoni, S., L. Piciullo, and S. L. Gariano (2018): A review of the recent literature on rainfall thresholds for landslide occurrence. *Landslides* 15 (8), 1483–1501 (cit. on p. 3).

Stanley, T. and D. B. Kirschbaum (2017): A heuristic approach to global landslide susceptibility mapping. *Natural hazards* 87 (1), 145–164 (cit. on p. 24).

Stepanek, C. and G. Lohmann (2012): Modelling mid-Pliocene climate with COSMOS. *Geoscientific Model Development* 5 (5), 1221–1243 (cit. on p. 4).

Strandberg, G., J. Brandefelt, E. Kjellstro
M, and B. Smith (2011): High-resolution regional simulation of last glacial maximum climate in Europe. *Tellus A: Dynamic Meteorology and Oceanography* 63 (1), 107–125 (cit. on p. 5).

Sun, J., K. Yang, W. Guo, Y. Wang, J. He, and H. Lu (2020): Why has the Inner Tibetan Plateau become wetter since the mid-1990s?. *Journal of Climate*, 1–45 (cit. on pp. 29, 31).

Sun, Y., G. Ramstein, L. Z. Li, C. Contoux, N. Tan, and T. Zhou (2018): Quantifying East Asian Summer Monsoon Dynamics in the ECP4. 5 Scenario With Reference to the Mid-Piacenzian Warm Period. *Geophysical Research Letters* 45 (22), 12–523 (cit. on p. 36).

Tao, W., W. Huijun, and J. Dabang (2010): Mid-Holocene East Asian summer climate as simulated by the PMIP2 models. *Palaeogeography, Palaeoclimatology, Palaeoecology* 288 (1-4), 93–102 (cit. on p. 4).

Tewari, M., F. Chen, W. Wang, J. Dudhia, M. LeMone, K. Mitchell, M. Ek, G. Gayno, J. Wegiel, and R. Cuenca (2004): Implementation and verification of the unified NOAH land surface model in the WRF model. *20th conference on weather analysis and forecasting/16th conference on numerical weather prediction*. Vol. 1115. American Meteorological Society Seattle, WA, 2165–2170 (cit. on p. 11).

Tzeng, G.-H. and J.-J. Huang (2011): *Multiple attribute decision making: methods and applications*. CRC press (cit. on p. 11).

von Storch, H. (1995): Inconsistencies at the interface of climate impact studies and global climate research. *Meteorologische Zeitschrift*, 72–80 (cit. on p. 3).

von Storch, H., H. Langenberg, and F. Feser (2000): A spectral nudging technique for dynamical downscaling purposes. *Monthly weather review* 128 (10), 3664–3673 (cit. on p. 3).

Wang, A. and X. Zeng (2012): Evaluation of multireanalysis products with in situ observations over the Tibetan Plateau. *Journal of geophysical research: Atmospheres* 117 (D5) (cit. on p. 16).

Wang, J., X. Fang, E. Appel, and C. Song (2012): Pliocene-Pleistocene Climate Change At the NE Tibetan Plateau Deduced From Lithofacies Variation In the Drill Core SG-1, Western Qaidam Basin, China. *J. Sediment. Res.* 82 (12), 933–952. issn: 1527-1404 (cit. on p. 25).

Wang, J., C. Song, J. T. Reager, F. Yao, J. S. Famiglietti, Y. Sheng, G. M. MacDonald, F. Brun, H. M. Schmied, R. A. Marston, et al. (2018): Recent global decline in endorheic basin water storages. *Nature geoscience* 11 (12), 926–932 (cit. on p. 26).

Wang, X., X. Guo, C. Yang, Q. Liu, J. Wei, Y. Zhang, S. Liu, Y. Zhang, Z. Jiang, and Z. Tang (2020): Glacial lake inventory of high-mountain Asia in 1990 and 2018 derived from Landsat images. *Earth System Science Data* 12 (3), 2169–2182 (cit. on p. 2).

Wieczorek, G. F. (1996): Landslides: investigation and mitigation. Chapter 4-Landslide triggering mechanisms. *Transportation Research Board Special Report* (247) (cit. on p. 19).

Woodcock, F. (1976): The evaluation of yes/no forecasts for scientific and administrative purposes. *Monthly Weather Review* 104 (10), 1209–1214 (cit. on p. 22).

Xu, J., S. Yu, J. Liu, S. Haginoya, Y. Ishigooka, T. Kuwagata, M. Hara, and T. Yasunari (2009): The implication of heat and water balance changes in a lake basin on the Tibetan Plateau. *Hydrological Research Letters* 3, 1–5 (cit. on p. 28).

Xue, Y., Z. Janjic, J. Dudhia, R. Vasic, and F. De Sales (2014): A review on regional dynamical downscaling in intraseasonal to seasonal simulation/prediction and major factors that affect downscaling ability. *Atmospheric research* 147, 68–85 (cit. on p. 4).

Yao, T., L. Thompson, W. Yang, W. Yu, Y. Gao, X. Guo, X. Yang, K. Duan, H. Zhao, B. Xu, et al. (2012): Different glacier status with atmospheric circulations in Tibetan Plateau and surroundings. *Nature climate change* 2 (9), 663–667 (cit. on p. 1).

Yoo, J., J. Galewsky, S. J. Camargo, R. Korty, and R. Zamora (2016): Dynamical downscaling of tropical cyclones from CCSM4 simulations of the Last Glacial Maximum. *Journal of Advances in Modeling Earth Systems* 8 (3), 1229–1247 (cit. on p. 5).

Yu, S., J. Liu, J. Xu, and H. Wang (2011): Evaporation and energy balance estimates over a large inland lake in the Tibet-Himalaya. *Environmental Earth Sciences* 64 (4), 1169–1176 (cit. on p. 28).

Zhou, X., K. Yang, L. Ouyang, Y. Wang, Y. Jiang, X. Li, D. Chen, and A. Prein (2021): Added value of kilometer-scale modeling over the third pole region: a CORDEX-CPTP pilot study. *Climate Dynamics*, 1–15 (cit. on pp. 3, 4, 15, 35).

Zubakov, V. and I. Borzenkova (1988): Pliocene palaeoclimates: past climates as possible analogues of mid-twenty-first century climate. *Palaeogeography, Palaeoclimatology, Palaeoecology* 65 (1-2), 35–49 (cit. on p. 26).

Paper I

WRF-based dynamical downscaling of ERA5 reanalysis data for High Mountain Asia: Towards a new version of the High Asia Refined analysis

Wang, X., Tolksdorf, V., Otto, M. and Scherer, D., 2021: WRF-based dynamical downscaling of ERA5 reanalysis data for High Mountain Asia: Towards a new version of the High Asia Refined analysis. *Int. J. Climatol.*, 41, 743–762, https://doi.org/10.1002/joc.6686.

Status: Published.

Own contribution:

- design of the study (with co-authors)
- numerical simulations
- data preparation
- data analysis and visualization
- writing and revision of manuscript

Received: 15 October 2019 | Revised: 31 March 2020 | Accepted: 24 May 2020 | Published on: 18 June 2020

DOI: 10.1002/joc.6686

RESEARCH ARTICLE

International Journal of Climatology RMetS

WRF-based dynamical downscaling of ERA5 reanalysis data for High Mountain Asia: Towards a new version of the High Asia Refined analysis

Xun Wang | **Vanessa Tolksdorf** | **Marco Otto** | **Dieter Scherer**

Chair of Climatology, Institute of Ecology, Technische Universität Berlin, Berlin, Germany

Correspondence
Xun Wang, Technische Universität Berlin, Institute of Ecology, Chair of Climatology, Rothenburgstraße 12, Berlin 12165, Germany.
Email: xun.wang@tu-berlin.de

Funding information
German Federal Ministry of Education and Research (BMBF), Grant/Award Number: FKZ 03G0878G

Abstract

The High Asia Refined analysis (HAR) is a regional atmospheric data set generated by dynamical downscaling of the Final operational global analysis (FNL) using the Weather Research and Forecasting (WRF) model. It has been successfully and widely utilized. A new version (HAR v2) with longer temporal coverage and extended domains is currently under development. ERA5 reanalysis data is used as forcing data. This study aims to find the optimal set-up for the production of the HAR v2 to provide similar or even better accuracy as the HAR. First, we conducted a sensitivity study, in which different cumulus, microphysics, planetary boundary layer, and land surface model schemes were compared and validated against in situ observations. The technique for order preference by similarity to the ideal solution (TOPSIS) method was applied to identify the best schemes. Snow depth in ERA5 is overestimated in High Mountain Asia (HMA) and causes a cold bias in the WRF output. Therefore, we used Japanese 55-year Reanalysis (JRA-55) to correct snow depth initialized from ERA5 based on the linear scaling approach. After applying the best schemes identified by the TOPSIS method and correcting the initial snow depth, the model performance improves. Finally, we applied the improved set-up for the HAR v2 and computed a one-year run for 2011. Compared to the HAR, the HAR v2 has a better representation of air temperature at 2 m. It produces slightly higher precipitation amounts, but the spatial distribution of seasonal mean precipitation is closer to observations.

KEYWORDS

cold bias, dynamical downscaling, ERA5, HAR, High Mountain Asia, snow depth, WRF

1 | INTRODUCTION

High Mountain Asia (HMA) is a geographic region that includes the Tibetan Plateau (TP) and its surrounding mountain ranges, such as the Himalayas, the Karakoram, the Tian Shan, and so on. Climate-triggered natural hazards pose a threat to human lives in HMA, for example, big landslides regularly occurring in the Fergana basin along the foothills of Tian Shan (Roessner *et al.*, 2005). Landslides are predetermined by static factors that can be derived from surface characteristics but are triggered by dynamic factors, which are mainly extreme and prolonged rainfall, as well as earthquakes (Dai and Lee 2002; Hong *et al.* 2007; Kirschbaum *et al.*, 2012). Another example of climate-triggered natural hazards is the Glacier Lake Outburst Flood (GLOF). Glacier thinning and retreat in the Himalayas, caused by rising air temperature, have resulted in the formation of new glacier lakes and the enlargement of existing ones (ICIMOD, 2011). The sudden discharge of water from these lakes is known as GLOF and leads to extensive damage in downstream villages (Kropáček *et al.*, 2015).

Availability of climate data with a high spatial and temporal resolution is crucial for a better understanding of climatic triggering mechanisms of these localized hazards. However, in HMA, in situ meteorological observations are sparsely and unevenly distributed (Hasson *et al.*, 2016), for example, only a few stations are available in the western TP due to the harsh environment and complex terrain. Moreover, the existing stations are commonly situated at lower altitudes close to valley-based settlements or airports. Thus, our knowledge of the climate at high elevations, that is, where most of the hazards mentioned above occur, is still limited. Global reanalysis data can provide evenly distributed climate data, but they are still too coarse to resolve fundamental processes over complex terrains, such as orographically-induced precipitation, and therefore, they are not suitable for applications at regional to local scales (Leung *et al.*, 2003; Lo *et al.*, 2008; Feser *et al.*, 2011). Here, regional climate models (RCMs) applying dynamical downscaling method have great potential to overcome this problem.

The High Asia Refined analysis (HAR, Maussion *et al.*, 2011; 2014) is a regional atmospheric data set generated by dynamical downscaling using the Weather Research and Forecasting (WRF) model version 3.3.1 (Skamarock and Klemp, 2008) as RCM. The Final operational global analysis (FNL) from the National Centres of Environmental Prediction (NCEP) was used as forcing data. The HAR covers the period from October 2000 to October 2014 and is available in 30 km (3-hourly interval) and 10 km (hourly interval) resolution. The HAR provides detailed and accurate gridded climate data for HMA region. It has been comprehensively analysed, especially in terms of precipitation and atmospheric water transport (Maussion *et al.*, 2014; Curio *et al.*, 2015; Pritchard *et al.*, 2019; Li *et al.*, 2020) and has been successfully applied in many research fields, such as glacier mass balance modelling (Mölg *et al.*, 2014), snow and energy balance modelling (Huintjes *et al.*, 2015), and so forth.

However, the short temporal coverage of the HAR makes it unsuitable for long-term and climatological studies. Moreover, its 10 km domain does not cover the whole TP and Tian Shan, which further limits its application within these two regions. Therefore, a new version (HAR v2) with extended temporal coverage and a larger 10 km domain is developed. The state-of-the-art ERA5 reanalysis data set (Copernicus Climate Change Service (C3S), 2017) from the European Centre for Medium-Range Weather Forecast (ECMWF) is used as forcing data. We switch to ERA5 because it will eventually cover the period from 1950 to near real time (currently available from 1979 to near real time), which is much longer than FNL (available from 1999 to near real time).

The overall goal of this study is to find an optimal model set-up for the HAR v2 to provide similar or even better accuracy as the HAR. During the development of the HAR, the sensitivity of simulated precipitation to different physical parameterization schemes (PPSs) was already thoroughly tested (Maussion *et al.*, 2011). However, changes in forcing data and domain configuration may have a significant impact on model output (Miguez-Macho *et al.*, 2004; Leduc and Laprise, 2009; Kala *et al.*, 2015; Huang and Gao, 2018). Thus, the PPSs used in the HAR might not be suitable for the HAR v2, and therefore, the first objective of the current study is to investigate the sensitivity of simulated total precipitation (Prcp) and air temperature at 2 m above ground (T2) to different cumulus (CU), microphysics (MP), planetary boundary layer (PBL) and land surface model (LSM) schemes. The technique for order preference by similarity to the ideal solution (TOPSIS) method is applied to determine the best PPSs.

Snow depth in ERA5 over the TP is reported to be largely overestimated (Orsolini *et al.* 2019). Snow depth is an important quantity in the initial condition, which later on determines surface albedo, alters surface energy balance, and influences T2. We assume that the overestimated snow depth in ERA5 leads to an underestimation of T2, and by correcting the bias in snow depth, the cold bias might be reduced. The second objective is to validate this assumption and to examine the model's sensitivity to initial snow conditions. Snow depth from the Japanese 55-year Reanalysis (JRA-55) data set is used to correct snow depth initialized from ERA5. The final objective is to apply the best PPSs identified by the TOPSIS method and the snow correction approach as the final set-up for the HAR v2, and to compare the two versions of the HAR.

2 | METHODOLOGY

2.1 | WRF model set-up of the reference experiment (Ref)

WRF version 4.0.3 (Skamarock *et al.*, 2019) was employed as RCM for our sensitivity studies. We chose January and July 2011 as simulation periods to consider atmospheric conditions in summer and winter, which are under different influence by monsoon and mid-latitude westerlies. Initial and boundary conditions were derived from ERA5 reanalysis data set with 0.25° spatial resolution and hourly temporal resolution. The domain setup (Figure 1) consisted of two-way nested domains with 30 km and 10 km grid spacing (hereinafter: d30km and d10km). Only the output from d10km was used in this study. One might argue that ERA5 already has a high resolution of ~32 km, so the parent domain (d30km) might not be necessary. However, a one-day experiment, which directly downscaled ERA5 to 10 km resolution (see Supporting Information), shows that the large-scale circulation patterns are distorted in the direct downscaling approach, and the 500 hPa wind field from two-way nesting approach is closer to the forcing data ERA5 (Figure S1). Thus, we kept the d30km as parent domain. In the vertical direction, 28 Eta-levels were used. Lake surface temperature was substituted by daily mean surface air temperature using the *avg_tsfc.exe* module in WRF. The forcing strategy was daily re-initialization adopted from the HAR. Each run started at 12:00 UTC and contained 36 hr, with the first 12 hr as spin-up time. This strategy avoids the model from deviating too far from the forcing data and provides computational flexibility since daily runs are totally independent of each other and can be computed in parallel and in any sequence. PPSs used in Ref are the same as in the HAR. The model set-up for Ref are summarized in Table 1.

2.2 | Design of sensitivity experiments and evaluation methods

At first, we conducted four sets of experiments to examine the performance of different CU, MP, PBL and LSM schemes (Table 2, hereinafter, PPS experiments). In each experiment set, except for the reference scheme adopted from the HAR, we chose two additional schemes. The selected schemes fulfil at least one of the following criteria: (a) they are commonly used in the WRF community; (b) they have excellent performance according to previous studies and (c) they were not tested in Maussion *et al.* (2011). Except for the corresponding PPS, all the other set-ups are the same as for Ref. At the end of the PPS experiments, statistical measures for model

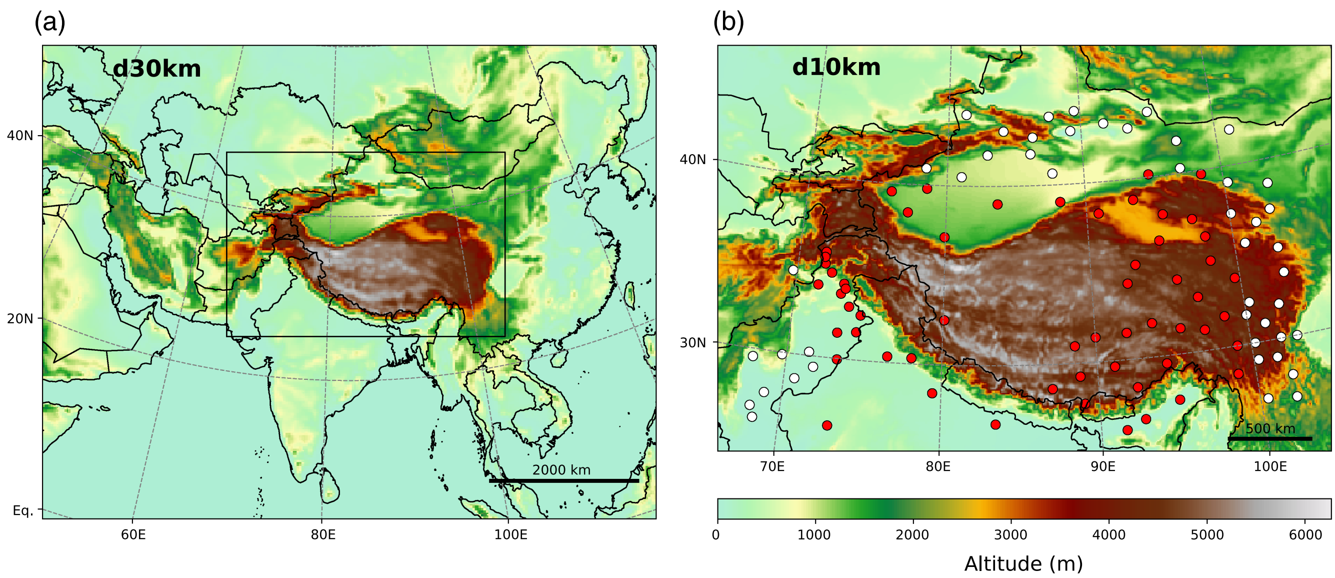

FIGURE 1 Maps of (a) 30 km resolution domain (d30km, 281 × 217 grid points); (b) 10 km resolution domain (d10km, 382 × 253 grid points). The position of stations from global surface summary of the day (GSOD) are marked by points. For the comparison of simulation results with the HAR, only the stations marked by red points are used, because the stations marked by white points are outside the 10 km domain of the HAR

TABLE 1 WRF model setup for the reference experiment (Ref)

Timing	
Simulation period	January and July 2011
Time step	120 s, 40 s
Maps and grids	
Map projection	Lambert conformal conic
Horizontal grid spacing	30 km (281 × 217 grid points), 10 km (382 × 253 grid points)
Vertical levels	28 eta-levels
Model top	50 hPa
Forcing strategy	
Forcing data	ERA5 (0.25°, hourly)
Lake surface temperature	Substituted by daily mean surface air temperature
Initialization	Daily
Runs starting time	Daily at 12:00 UTC
Runs duration	36 hr
Spin-up time	12 hr
Physical parameterization schemes	
Longwave radiation	RRTM scheme (Mlawer *et al.*, 1997)
Shortwave radiation	Dudhia scheme (Dudhia, 1989)
Cumulus (CU)	Grell 3D scheme (Grell, 1993; Grell and Devenyi, 2002)
Microphysics (MP)	Thompson scheme (Thompson *et al.*, 2008)
Planetary boundary layer (PBL)	Mellor–Yamada–Janjic scheme (Janjic, 1994)
Land surface model (LSM)	Unified Noah land surface model (Tewari *et al.*, 2004)
Surface layer	Eta similarity scheme (Janjic, 1994)

TABLE 2 Summary of set-ups used in all the sensitivity experiments

Experiments	Difference from Ref	Description
CU1	CU	Kain-Fritsch cumulus potential scheme (Berg *et al.*, 2013)
CU2	CU	Grell-Freitas ensemble scheme (Grell and Freitas, 2014)
MP1	MP	Purdue Lin scheme (Chen and Sun, 2002)
MP2	MP	Morrison 2-moment scheme (Morrison *et al.*, 2009)
PBL1	PBL	Bougeault and Lacarrere scheme (BouLac, Bougeault and Lacarrere, 1989)
PBL2	PBL	Yonsei University scheme (YSU, Hong *et al.*, 2006)[a]
LSM1	LSM	Unified Noah LSM with mosaic approach (Li *et al.*, 2013)
LSM2	LSM	Noah-MP land surface model (Niu *et al.*, 2001; Yang *et al.*, 2011)[b]
COMB	CU, MP, PBL	CU, MP and PBL schemes from CU1, MP2 and PBL2
COMB_S	CU, MP, PBL and snow correction	PPSs same as COMB, but the initial snow depth and snow water equivalent are corrected

[a]YSU scheme only works with revised MM5 surface layer scheme (Jiménez *et al.*, 2012).
[b]All the options used in Noah-MP are default.

performance were calculated, which include mean bias (MB), mean absolute error (MAE), root mean square error (RMSE) and Spearman correlation coefficient (r_s) for Prcp and T2, respectively. They are defined as follows (Wilks, 2006):

$$\mathrm{MB}=\frac{1}{N}\sum_{i=1}^{N}\bar{P}_i-\bar{O}_i \tag{1}$$

$$\mathrm{MAE}=\frac{1}{N}\sum_{i=1}^{N}\left|\bar{P}_i-\bar{O}_i\right| \tag{2}$$

$$\mathrm{RMSE}=\frac{1}{N}\sum_{i=1}^{N}\left(\bar{P}_i-\bar{O}_i\right)^2 \tag{3}$$

$$r_s=1-\frac{6\sum_{i=1}^{M}\left(R_{pi}-R_{oi}\right)^2}{M\left(M^2-1\right)} \tag{4}$$

Where $\bar{P}_i$ and $\bar{O}_i$ are simulated monthly averages and observed monthly averages at each station, and $\bar{P}_i-\bar{O}_i$ is the bias score at each station; N is the total number of stations; R_{pi} and R_{oi} are the ranks of simulated and observed daily averages over all stations; M is the total number of days. MB demonstrates the systematic deviation of the model from the observations. However, MB could sometimes be misleading due to the offset of positive and negative values. MAE shows the average magnitude of the error. Since RMSE gives more weight to errors with larger absolute values and is more sensitive to outliers, it provides information about the variability of the error distribution (Chai and Draxler, 2014; Willmott and Matsuura, 2005). The distribution-free Spearman rank correlation coefficient (r_s) measures the extent to which simulated and observed values tend to change together. In order to identify the best scheme for each experiment set, MB, MAE, RMSE and r_s were utilized as criteria in the TOPSIS method (Section 2.3). In the next step, the identified best schemes were combined to conduct two runs (Table 2): one without modifying initial snow depth (COMB) and one with bias correction of snow depth (COMB_S). We used JRA-55 to correct snow depth initialized from ERA5 (Section 2.4). Finally, the improved set-up was applied as the best set-up for the HAR v2 for a one-year run in 2011. The resulting one-year data set was then compared with observations and the HAR.

2.3 | Technique for order preference by similarity to the ideal solution (TOPSIS)

The TOPSIS method was applied to identify the best scheme among the sensitivity experiments. It is a multiple criteria decision-making method and aims at finding the optimal decision when the alternatives are numerous and conflicting. It was proposed by Hwang and Yoon (1981) and later applied in several studies, such as ranking general circulation models (Raju and Kumar, 2014; Jena *et al.*, 2015; Li *et al.*, 2019) and identifying the best PPSs in WRF (Sikder and Hossain, 2016; Stergiou *et al.*, 2017). The basic concept of TOPSIS is to determine the best alternative that has the shortest and longest distance from the positive and negative ideal solution. We applied this method for each experiment set. The process was carried out following the description in Tzeng and Huang (2011):

1 Define the weighted normalized evaluation matrix containing m alternatives and n criteria:

$$r_{ij}=w_j\frac{a_{ij}}{\sqrt{\sum_{i=1}^{m}a_{ij}^2}} \tag{5}$$

where a_{ij} represents the criterion j for alternative i; w_j is the weight for each criterion. In our case, there are three alternatives (Ref and other two schemes, i = 1, 2, 3) and 16 criteria (MB, MAE, RMSE and r_s for Prcp and T2 in January and July, j = 1, 2, ...16).

Determine the positive ideal solution (PIS_j) and the negative ideal solution (NIS_j) for each criterion:

$$\text{PIS}_j=\left\{\left(\max\left(r_{ij}\right)|j\in J_1\right),\left(\min\left(r_{ij}\right)|j\in J_2\right)\mid i=1,2,3\right\} \tag{6}$$

$$\text{NIS}_j=\left\{\left(\min\left(r_{ij}\right)|j\in J_1\right),\left(\max\left(r_{ij}\right)|j\in J_2\right)\mid i=1,2,3\right\} \tag{7}$$

where J_1 and J_2 represent the benefit criteria (larger is better) and the cost criteria (smaller is better).

Calculate the Euclidean distances between each alternative to PISs (D_i^+) and to NISs (D_i^-):

$$D_i^+=\sqrt{\sum_{j=1}^{n}\left(r_{ij}-\text{PIS}_j\right)^2} \tag{8}$$

$$D_i^-=\sqrt{\sum_{j=1}^{n}\left(r_{ij}-\text{NIS}_j\right)^2} \tag{9}$$

Define the closeness coefficient (CC_i) for each alternative as:

$$\text{CC}_i=\frac{D_i^-}{D_i^++D_i^-} \tag{10}$$

Rank the alternatives by their CC in descending order. The alternative with the highest CC is chosen as the best scheme.

2.4 | Bias correction of snow depth

A systematic cold bias has been found over high elevated areas in WRF simulations (e.g., Gao *et al.*, 2015; Karki *et al.*, 2017; Bonekamp *et al.*, 2018), including the HAR (Maussion, 2014; Pritchard *et al.*, 2019). Several studies addressed this cold bias to snow-related processes (Tomasi *et al.*, 2017; Meng *et al.*, 2018). Orsolini *et al.* (2019) compared snow depth over the TP in five

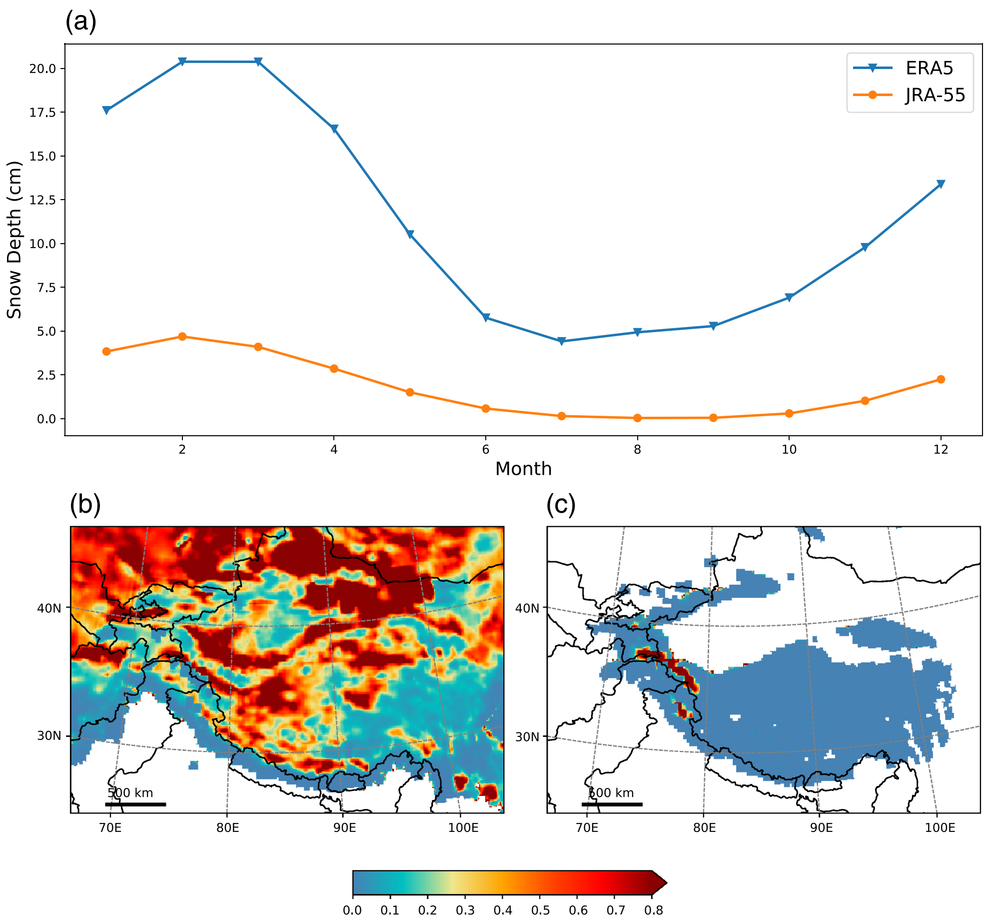

FIGURE 2 (a) 35-year (1979–2013) mean annual cycle of snow depth from ERA5 and JRA-55 averaged over the whole area of d10km of the HAR v2. Map of scaling factor for d10km of HAR v2 in January (b) and July (c). The grid points where the value is equal to one are masked out

recent global reanalyses using in-situ and satellite remote sensing observations. They found that ERA5 largely overestimates snow depth, especially during winter. Compared to its predecessor ERA-interim, snow depth in ERA5 is still much higher, because ERA5 does not assimilate snow cover from Interactive Multi-Sensor Snow and Ice Mapping System (IMS) for areas above 1,500 m (Orsolini *et al.*, 2019). A realistic representation of snow depth in the forcing data is crucial to accurately simulate snow cover, surface albedo, surface energy balance and T2. To overcome this issue, we could use snow depth from another data set to initialize WRF. But as mentioned before, we chose ERA5 as forcing data due to its long temporal coverage. If we introduce a third-party data set, the production of the HAR v2 would be dependent on it. To avoid this, we applied an alternative strategy that is already used in WRF, where land surface characteristic, such as snow-free albedo, leaf area index, and so forth, are given in the form of static data represented as 12-month climatology. Here, analogously, we used this concept of static data and introduced a gridded 12-month climatology of a scaling factor to correct snow-related variables in the initial conditions.

JRA-55 (Kobayashi *et al.*, 2015) developed by the Japan Meteorological Agency was utilized to calculate the scaling factors. JRA-55 has an excellent performance among reanalyses regarding snow depth (Orsolini *et al.*, 2019) and also has a relatively longer temporal coverage. Monthly means of snow depth at the model resolution (~55 km) available from 1958 to 2013 were downloaded from NCAR Research Data Archive. We used a linear approach (Lenderink *et al.* 2007) to scale the initial snow depth by the ratio of long-term monthly means of JRA-55 and ERA5. First, snow depth of ERA5 was derived from snow water equivalent and snow density. The 12-month climatological snow depth of JRA-55 and ERA5 was calculated from 35-year monthly means (1979–2013). Between 1979 and 2013, the two data sets overlap. Second, the 12-month climatological snow depth was reprojected and linearly resampled to the grid of d30km and d10km, respectively. Then the gridded 12-month climatology of the scaling factor was calculated for every domain as the ratio of reprojected climatological snow depth between JRA-55 and ERA5. The ratio of grid points where reprojected ERA5 snow depth was zero was set to one. Figure 2a depicts the 35-year mean annual cycle from ERA5 and JRA-55 averaged over the whole area of d10km of the HAR v2. Figure 2b,c show the map of scaling factors applied for d10km in January and July, respectively. After running the *real.exe* program in WRF, files containing the initial conditions are generated. In the initial conditions, snow depth is represented by two variables: physical snow depth (SNOWH) and snow water equivalent (SNOW). We modified these two variables in the initial files by multiplying them with the scaling factor of the corresponding month.

2.5 | Observational data

To calculate the skill scores described in Section 2.2, in-situ observations of Prcp and T2 from Global Surface Summary of the Day (GSOD) provided by National Centres of Environmental Information (NCEI) were used. We selected stations with more than 80% of records within the simulation period. After filtering, 103 stations are available within d10km of the HAR v2 (all points in Figure 1b), and 57 stations are left within d10km of the HAR (red points in Figure 1b). We compared the station data with the nearest grid cell from the model. Because of the difference between station elevation and the elevation of grid cells in the model, we applied a constant lapse rate of $-6.5\ \mathrm{K{\cdot}km^{-1}}$ to correct simulated T2. No correction was applied for Prcp due to its high spatial and temporal variability.

Due to the sparse and uneven distribution of GSOD stations in HMA, for example, hardly any station located in western TP and Pamir-Karakorum (Figure 1b), we also made use of a satellite-based gridded precipitation product from Tropical Rainfall Measuring Mission (TRMM). Here, Multi-Satellite Precipitation Analysis (TMPA, Huffman and Bolvin, 2015) Rainfall Estimate Product 3B42 Version 7 was applied to examine the spatial distribution of simulated Prcp in the HAR and the HAR v2. TMPA merges satellite measurement with gauge data. It has a spatial resolution of 0.25° and a quasi-global coverage between 50°N and 50°S.

3 | RESULTS

3.1 | Sensitivity to PPSs

Statistical scores for all the PPS experiments and Ref are listed in Table 3. Besides, we use box-whisker plot to visualize the spatial distribution of bias scores (WRF-observation) over 103 GSOD stations. White triangles in each box in Figure 3 (Prcp) and Figure 4 (T2) correspond to the MB scores in Table 3.

For Prcp, Ref features a wet bias, with MB of $0.14\ \mathrm{mm{\cdot}day^{-1}}$ and $0.70\ \mathrm{mm{\cdot}day^{-1}}$ in January and July, respectively (Table 3). The HAR outperforms Ref with respect to Prcp (Table S1). Sensitivities of Prcp to PPSs vary seasonally. The fluctuation of the boxes of all experiments implies that PPSs have a stronger influence on Prcp in July than in January (Figure 3). MB of Prcp ranges from $0.27\ \mathrm{mm{\cdot}day^{-1}}$ to $1.15\ \mathrm{mm{\cdot}day^{-1}}$ in July, and only from $0.10\ \mathrm{mm\ day^{-1}}$ to $0.18\ \mathrm{mm{\cdot}day^{-1}}$ in January (Table 3). The main reason for this seasonal variability is that most GSOD stations are located in areas receiving more precipitation in summer than in winter (Maussion *et al.*, 2014; Curio and Scherer, 2016). Wider boxes in July indicate a stronger spatial variation of bias score in summer (Figure 3). Compared to the other two ensemble CU schemes, CU1 (Kain-Fritsch-Cumulus Potential scheme) shows the best performance regarding MAE and MB scores. In contrary, RMSE of CU1 in July is higher than Ref, which implies that the station-wise bias in CU1 is more scattered. CU1 is a modified version of the Kain-Fritsch scheme with a better treatment of shallow cumuli (Berg *et al.*, 2013). According to Qian *et al.* (2016), CU1 tends to suppress deep convections and consequently produces lower Prcp. All three MP schemes consider five hydrometeor species: cloud, rain, ice, snow and graupel. MP1 (Purdue Lin scheme) is a single-moment scheme only predicting the mixing ratio for these hydrometeor species (Chen and Sun, 2002). Ref (Thompson scheme) uses a double-moment description for rain and ice (Thompson *et al.*, 2008), while MP2 (Morrison two-moment scheme) uses a double-moment description for rain, ice, snow and graupel (Morrison *et al.*, 2009). MP2 performs the best (Table 3), probably due to its double-moment prediction in all ice-phase particles (Orr *et al.*, 2017). The nonlocal scheme PBL2 (YSU) has the best skill compared to the other two local schemes. PBL2 largely improves summer Prcp with MB of $0.27\ \mathrm{mm{\cdot}day^{-1}}$ compared to MB of $0.70\ \mathrm{mm{\cdot}day^{-1}}$ in Ref. As for LSM schemes, LSM1 uses the same scheme as Ref (Noah LSM) but with a mosaic approach (Li *et al.*, 2013), which considers sub-grid variability of land use. However, it does not improve Prcp simulation. LSM2 is Noah LSM with multi-parameterization options (Noah-MP). It captures winter Prcp better but produces more Prcp in summer than Ref (Table 3).

The same analysis is performed for T2. The HAR has a better performance than Ref in winter (Table S1). All PPS experiments including Ref produce larger T2 bias in January than in July, with MB of T2 ranging from -0.06 K to 0.70 K in July and -0.39 K to -2.69 K in January (Table 3). The spatial variability of the bias score is also stronger in winter (Figure 4). Independently from PPSs applied, all experiments show an overall underestimation of T2 in winter. CU schemes hardly influence winter T2. CU1, which shows some improvement over

TABLE 3 Statistical scores for evaluation of PPS experiments over 103 GSOD stations within d10km of the HAR v2 (red and white points in Figure 1b) for Prcp (mm·day^{-1}) and T2 (K)

	Prcp January				Prcp July				T2 January				T2 July			
	MAE	MB	RMSE	r_s	MAE	MB	RMSE	r_s	MAE	MB	RMSE	r_s	MAE	MB	RMSE	r_s
Ref	0.22	0.14	0.42	0.67	1.53	0.70	2.68	0.48	2.89	−1.21	3.99	0.80	1.20	0.35	1.81	0.83
CU1	0.21	**0.12**	**0.34**	0.67	1.48	**0.47**	**(2.99)**	0.47	2.89	−1.19	3.98	0.81	1.27	**(0.39)**	1.87	0.82
CU2	0.23	0.14	0.43	0.68	**(1.87)**	**(1.15)**	**(3.62)**	**(0.26)**	2.89	−1.20	3.99	0.81	1.30	**0.17**	1.88	0.85
MP1	0.22	0.14	0.40	0.71	1.59	**0.56**	2.81	**(0.38)**	2.91	**(−1.38)**	4.05	0.81	1.17	**0.21**	1.78	0.83
MP2	0.22	0.13	0.41	0.64	1.53	**0.45**	2.79	0.50	2.79	−1.12	3.89	0.81	1.21	**0.20**	1.79	0.83
PBL1	**(0.25)**	**(0.18)**	0.45	0.65	1.52	0.71	2.58	0.45	**2.55**	**−0.39**	**3.50**	0.85	1.28	**(0.70)**	1.94	0.82
PBL2	0.20	**0.12**	0.39	0.68	**1.27**	**0.27**	**2.17**	0.42	3.08	**(−1.68)**	4.28	0.83	1.21	**0.30**	1.84	0.84
LSM1	0.23	0.14	0.43	0.68	1.54	0.71	2.71	0.49	2.83	**−1.08**	3.89	0.80	1.20	0.38	1.82	0.83
LSM2	**0.18**	**0.10**	**0.37**	0.71	**(1.89)**	**(1.00)**	**(3.45)**	**0.53**	**(4.11)**	**(−2.69)**	**(5.40)**	0.82	**(1.33)**	**−0.06**	1.91	0.83

Note: The values that differ from ref of more than 10% are marked in bold and put between parentheses if they show a lower skill than Ref

Ref in simulating Prcp, produces larger summer warm bias than Ref. Both MP schemes show an improvement in summer, but MP1 produces larger winter cold bias. As for PBL schemes, PBL1 improves T2 simulation in winter but worsens it in summer. It is warmer than Ref in both seasons, which is in accordance with previous studies (Kleczek *et al.* 2014; Jänicke *et al.*, 2017; Xu *et al.*, 2019). Xu *et al.* (2019) suggests that this is because PBL1 considers turbulent exchange induced by steep terrain when calculating the turbulent diffusion coefficient. Besides, our results show that the nonlocal PBL2 scheme is colder than both local schemes, which is in contrast with some previous studies concluding that nonlocal schemes generally produce higher temperature than local schemes (Hu *et al.*, 2010; Xie *et al.*, 2012; Kleczek *et al.*, 2014). This different behaviour might due to the unique PBL characteristics over HMA compared to the plain regions (Yang *et al.*, 2004; Zou *et al.*, 2005; Chen *et al.*, 2013), which indicates that PPSs behave very differently in different regions, and the choice of PPSs always depends on the specific purpose. For LSM schemes, again, LSM1 does not show significant improvement. Noah-MP (LSM2) was reported to have better performance than Noah LSM (Ref) in previous studies (Yang *et al.*, 2011; Gao *et al.*, 2017) and has been widely applied in numerical modelling studies in HMA (Collier and Immerzeel, 2015; Karki *et al.*, 2017; Norris *et al.*, 2017). However, our results show the opposite, that is, that Noah-MP is colder than Noah LSM. This might result from our relatively short spin-up time. All the studies mentioned above use a spin-up time longer than 2 weeks. Noah-MP needs longer spin-up time than Noah LSM to reach a climatological equilibrium state (Cai *et al.*, 2014; Barlage *et al.*, 2015; Gao *et al.*, 2015).

3.2 | Combination run and sensitivity to initial snow depth

The results from Section 3.1 indicate that there is a trade-off in model performance between Prcp and T2, as well as between January and July. No single PPS performs ideally in both seasons and for both quantities. For instance, PBL2 has a better skill in predicting Prcp, but it worsens the winter cold bias. PBL1 shows superior performance in terms of T2 during winter with the lowest MAE, MB and RMSE among all experiments, but it also features the highest warm bias in summer. Therefore, we used the TOPSIS method to find the optimal solution for each experiment set, giving the same weight to each criterion (MAE, MB, RMSE and r_s) for each quantity (Prcp and T2). Table 4 lists the closeness coefficient (CC) and TOPSIS ranking for each experiment set. We then used

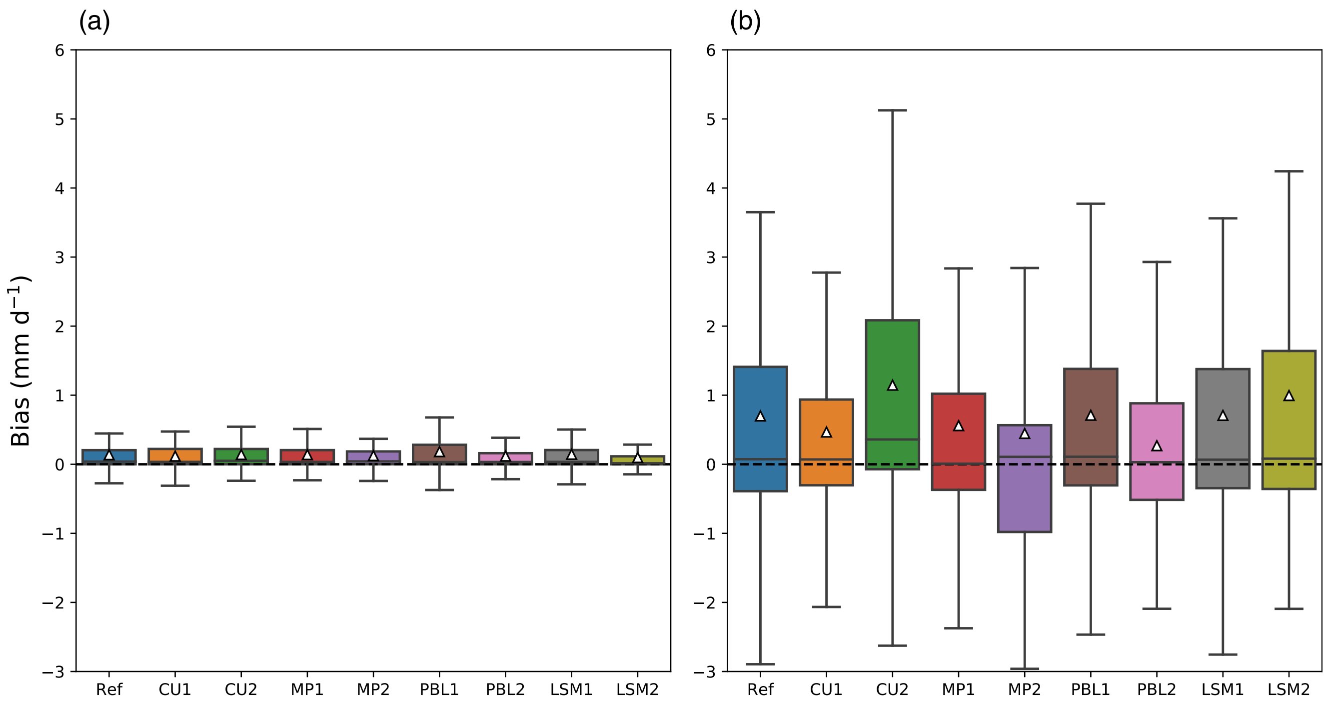

FIGURE 3 Box-whisker plot of Prcp bias in January (a) and July (b) for all the 103 GSOD stations (red and white points in Figure 1b). White triangles represent the mean value

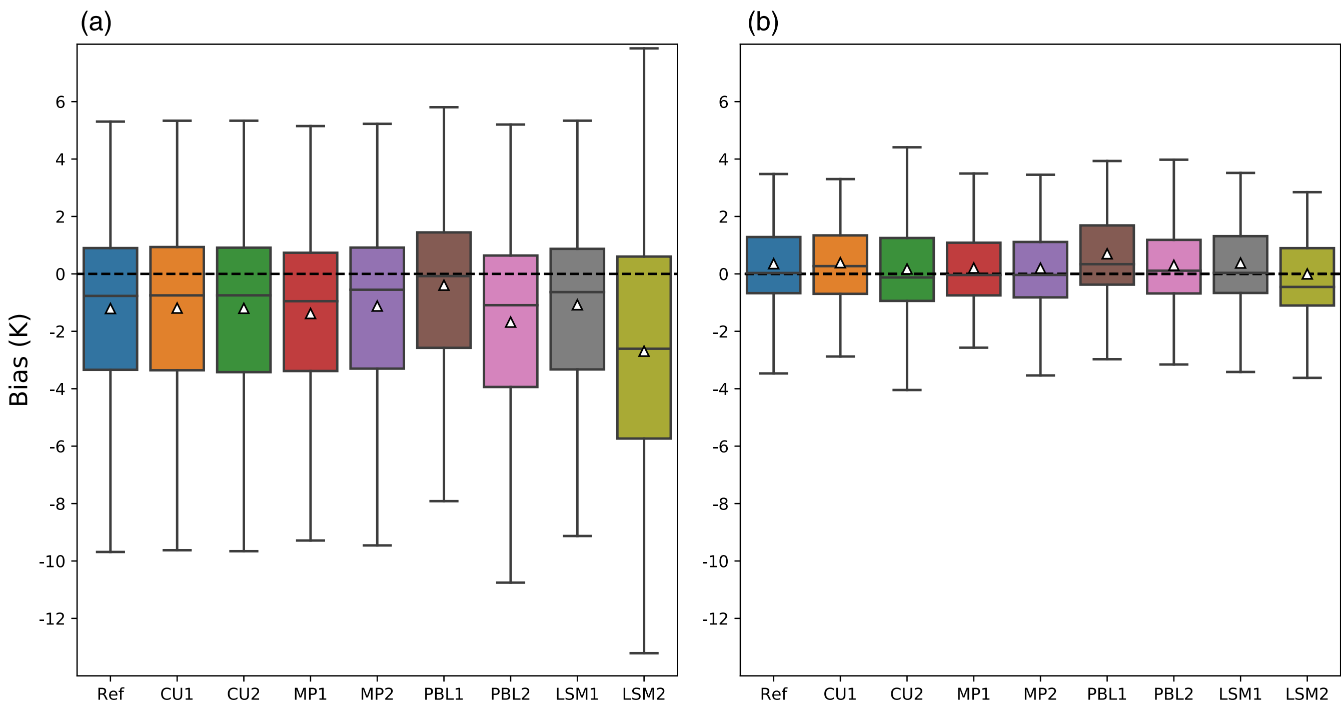

FIGURE 4 Box-whisker plot of T2 bias in January (a) and July (b) for all the 103 GSOD stations (red and white points in Figure 1b). White triangles represent the mean value

TABLE 4 Closeness coefficient (CC) and TOPSIS ranking for each experiment set

	Ref	CU1	CU2	Ref	MP1	MP2	Ref	PBL1	PBL2	Ref	LSM1	LSM2
CC	0.56	**0.61**	0.38	0.27	0.60	**0.88**	0.46	0.47	**0.53**	**0.51**	0.50	0.50
Ranking	2	**1**	3	3	2	**1**	3	2	**1**	**1**	3	2

Note: The best schemes under each set are marked in bold.

TABLE 5 Statistical scores for ref, COMB and COMB_S over 103 GSOD stations within d10km of the HAR v2 (red and white points in Figure 1b) for Prcp (mm·day^{-1}) and T2 (K)

	Prcp January				Prcp July				T2 January				T2 July			
	MAE	MB	RMSE	r_s	MAE	MB	RMSE	r_s	MAE	MB	RMSE	r_s	MAE	MB	RMSE	r_s
Ref	0.22	0.14	0.42	0.67	1.53	0.70	2.68	0.48	2.89	−1.21	3.99	0.80	1.20	0.35	1.81	0.83
COMB	**0.18**	**0.10**	**0.30**	0.67	**1.34**	**0.12**	2.58	0.44	2.96	**(−1.54)**	4.15	0.82	1.31	**0.29**	1.91	0.80
COMB_S	0.22	0.14	**0.35**	**(0.59)**	1.41	**0.22**	2.80	0.48	**1.65**	**0.17**	**2.68**	0.83	1.28	0.33	1.89	0.79

Note: The values that differ from ref of more than 10% are marked in bold and put between parentheses if they show a lower skill than Ref.

the first ranking schemes, namely Kain-Fritsch cumulus potential scheme (CU1), Morrison 2-moment scheme (MP2), Yonsei University scheme (PBL2) and Noah LSM (Ref) to conduct a combination run (COMB). Scores in Table 5 show that COMB inherits from these first-ranking schemes. COMB has a better skill in predicting Prcp than Ref because CU1, MP2 and PBL2 all perform better than Ref in Prcp simulation. The wet bias is significantly reduced in COMB, especially in summer. Figures 5a–d shows the bias of monthly Prcp at each GSOD stations for Ref and COMB. The stations with high bias are located along the foothills of the Himalayas, where the actual precipitation amount is also large (Figure S2). Compared to Ref, COMB produces lower Prcp in India and Pakistan. COMB captures Prcp over the TP better, especially for stations in Southern TP, where Ref exhibits a dry bias. However, COMB does not improve the simulation of winter T2 (Figures 6a–d, Table 5), which is mainly due to the larger winter cold bias induced by PBL2 (Table 3).

To validate the assumption that the large winter cold bias is partly induced by overestimation of snow depth in the initial conditions, we conducted a further simulation (COMB_S). The PPSs used in this experiment are the same as in COMB, but initial snow depth and snow water equivalent were corrected using the method described in Section 2.4.

After the correction, the prediction of winter T2 is largely improved with MB of −1.54 K in COMB and MB of 0.17 K in COMB_S (Table 5). The impact of snow correction on summer T2 is minor compared to winter. Comparing the station-wise bias scores (Figure 6c,e), the most affected stations are located in the TP and in Northern Pakistan, where snow depth differs the most between ERA5 and JRA-55 (Section 4.2). This snow correction approach leads to slightly higher Prcp in both months (Table 5), but COMB_S still performs better than Ref.

3.3 | Comparison of HAR v2 with HAR

COMB_S achieves a significantly better performance than Ref (Table 5). Therefore, the set-up of COMB_S was applied to the HAR v2. We switched to a newer version of WRF (V4.1) for the generation of the HAR v2 and firstly produced HAR v2 d10km data for the whole year of 2011. Figure S3 shows that the change of model version only has a minor impact on the output.

We compared daily Prcp from the HAR v2 with the HAR and GSOD data (Figure 7). Here, only 57 GSOD stations within the d10km of the HAR are used (red points in Figure 1b). The HAR shows an overall lower MB and is better in explaining the variance in observations than

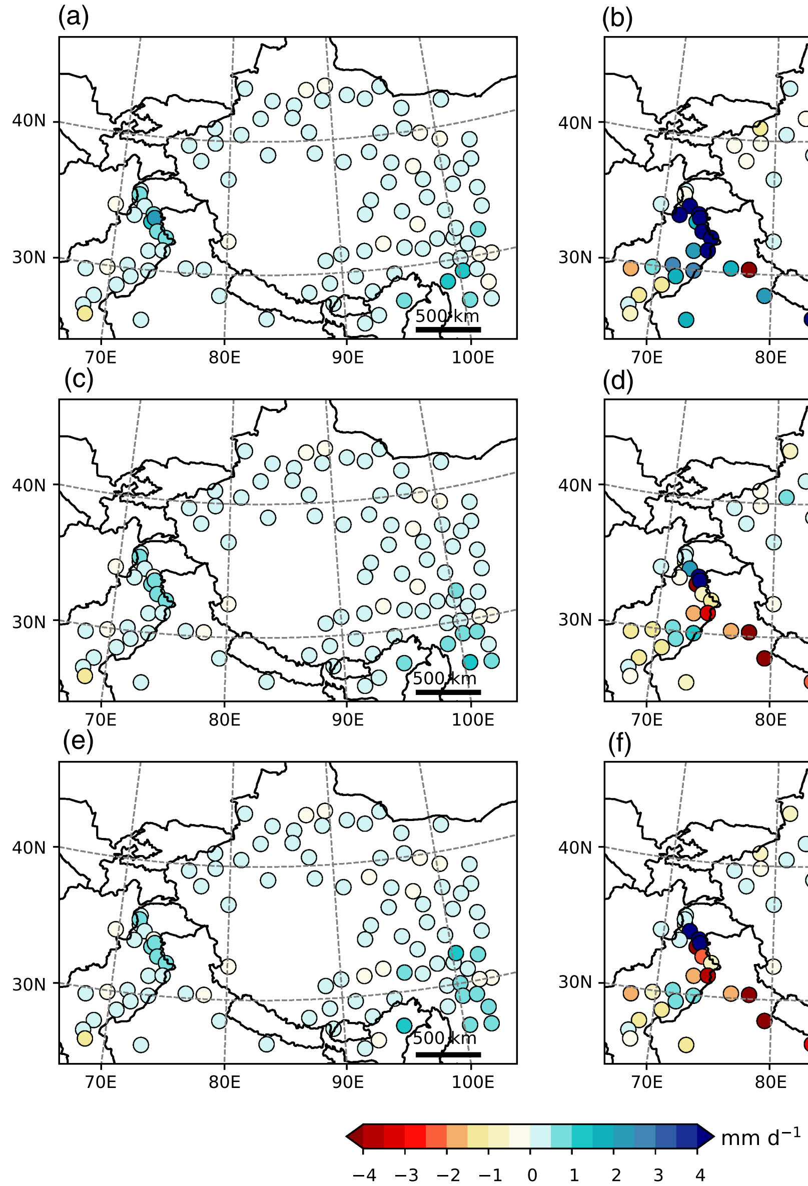

FIGURE 5 Bias scores of monthly mean Prcp for ref (a and b), COMB (c and d) and COMB_S (e and f) in January (a, c and e) and July (b, d and f) at all 103 GSOD stations

the HAR v2 (MB of 0.30 mm·day^{-1} and 0.36 mm·day^{-1}, r^2 of .61 and .57). The HAR v2 tends to produce more Prcp than the HAR in April and from September to December. However, this kind of comparison only demonstrates the ability of the two versions of the HAR to reproduce the overall Prcp amount. In addition, we utilized gridded Prcp products to examine their ability to reproduce the spatial distribution of Prcp. Figure 8 shows seasonal Prcp from the HAR, the HAR v2 and the TMPA in 2011. In winter, the HAR and the HAR v2 produce similar spatial patterns of Prcp with maximum Prcp over Pamir-Karakoram-western-Himalayas (PKwH) region. However, this maximum centre is not detected in the TMPA. In MAM, two Prcp maxima exist in both versions of the HAR: one in PKwH and one over the region near Brahmaputra River. Only the latter one is visible in the TMPA, but with lower Prcp amount. The HAR and the HAR v2 show the largest differences in JJA. The HAR v2 shares more spatial similarity with the TMPA, while the HAR produces higher Prcp amount in Indian, Pakistan and eastern TP. In SON, the HAR still shows higher Prcp amount in Indian. Same as in winter and spring, both versions of the HAR detect more Prcp in PKwH than the TMPA.

Daily T2 from the HAR and the HAR v2 are compared with each other and with GSOD (Figure 9). Both data sets reproduce T2 seasonality well (r^2 = .99). The HAR v2 simulates T2 better with MB of −0.58 K

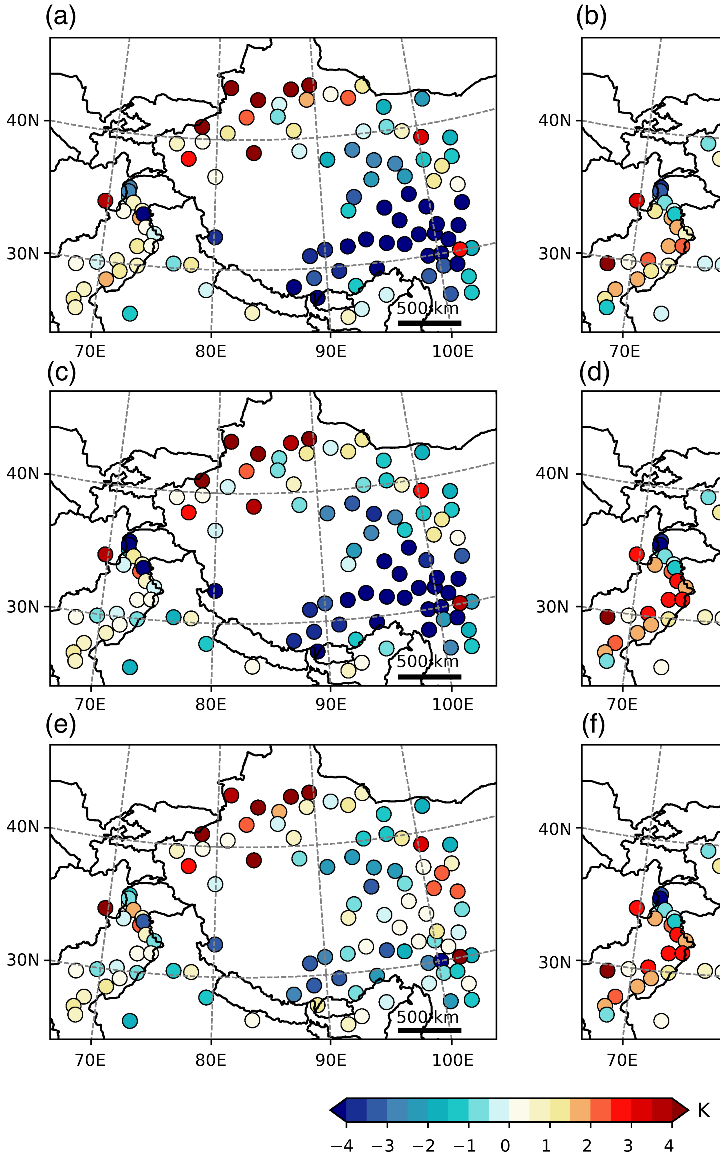

FIGURE 6 Bias scores of monthly mean T2 for ref (a and b), COMB (c and d) and COMB_S (e and f) in January (a, c and e) and July (b, d and f) at all 103 GSOD stations

compared to MB of −0.86 K for the HAR. The HAR has a cold bias in spring, autumn and winter but a warm bias in summer. The HAR v2 has the same pattern, but the magnitudes of these biases are smaller.

4 | DISCUSSION

4.1 | Sources of uncertainties

One should always keep in mind that the accuracy of observational data and the method used for validation could lead to uncertainty of the result. In this study, we use freely available observations from GSOD as the ground truth to calculate statistical metrics. Station data from this data set underwent quality control before being published (NOAA, 2019). However, some GSOD stations seem to be outliers, for example, the station in southeastern TP featuring winter warm bias, while all the surrounding stations show cold bias (Figure 6a,c,e). With regards to Prcp, it is well known that rain gauges undercatch Prcp under strong wind conditions (Duchon and Essenberg, 2001; WMO, 2008). We used T2 and Prcp from the nearest grid point to compare them with observations. With a resolution of 10 km, the distance between the actual location of stations and the associated grid point could reach a few kilometres, which leads to a large difference in elevation over complex terrain. A constant lapse rate of $-6.5\ \text{K·km}^{-1}$ is applied to correct the simulated T2 values. For Prcp, no correction is applied.

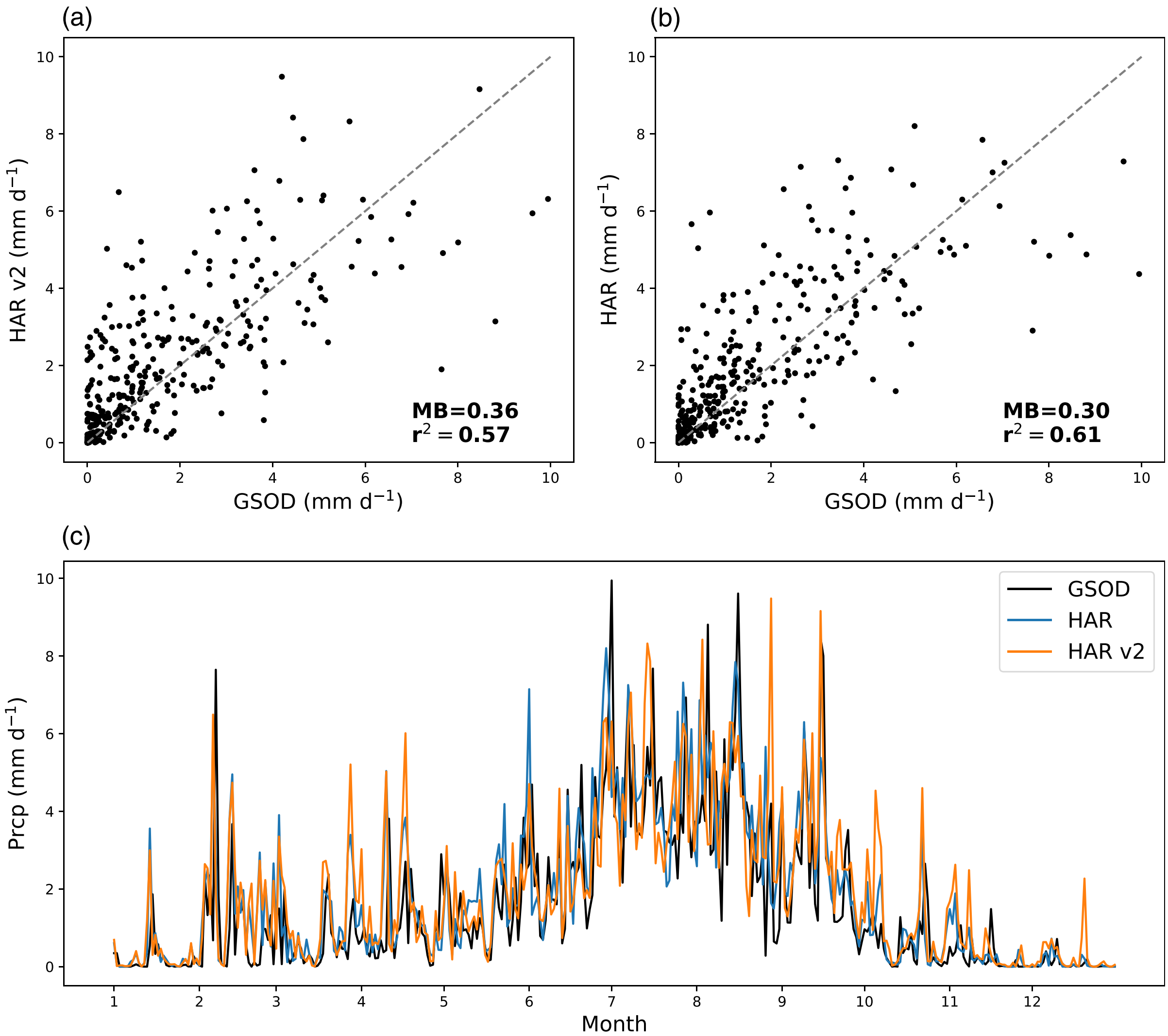

FIGURE 7 Comparison of daily Prcp from HAR, HAR v2 and GSOD averaged over 57 GSOD stations (red points in Figure 1b) in 2011. Scatter plots of HAR v2 (a) and HAR (b) with statistical scores MB in mm day^{-1} and r^2. Daily Prcp time series (c)

Therefore, the point-to-point comparison between modelled and observed Prcp always depicts discrepancies. Another source of uncertainty comes from the fact that we only conducted sensitivity experiments for the year 2011 due to the high computational costs.

Due to the sparse distribution of GSOD stations, we also compared the spatial distribution of seasonal Prcp from two versions of the HRA with the TMPA. The results (Figure 8) indicate that the TMPA fails to reproduce Prcp over PKwH in winter, spring and autumn. This region is dominated by mid-latitude westerlies and experiences its maximum Prcp amount in winter (Archer and Fowler, 2004; Bookhagen and Burbank, 2010; Curio and Scherer, 2016; Mölg *et al.*, 2018). The majority of Prcp in this region falls as snow (Maussion *et al.*, 2014). Satellite-based Prcp has a deficiency in detecting snowfall (Behrangi *et al.*, 2014; Immerzeel *et al.*, 2015). The Global Precipitation Measurement (GPM) mission, launched in February 2014, is a successor of TRMM and a next generation of global observations of precipitation from space. The onboard Dual-frequency Precipitation Radar is more sensitive than TRMM in detecting snowfall and light rainfall (NASA, 2019). GPM-based products are expected to be a better validation tool for modelled results after 2014.

4.2 | Snow correction approach

In this study, we propose a bias correction method for initial snow depth and snow water equivalent based on the concept of linear scaling approach. This approach has the advantage that only monthly climatological information is required. Figure 6 and Table 5 show that bias-

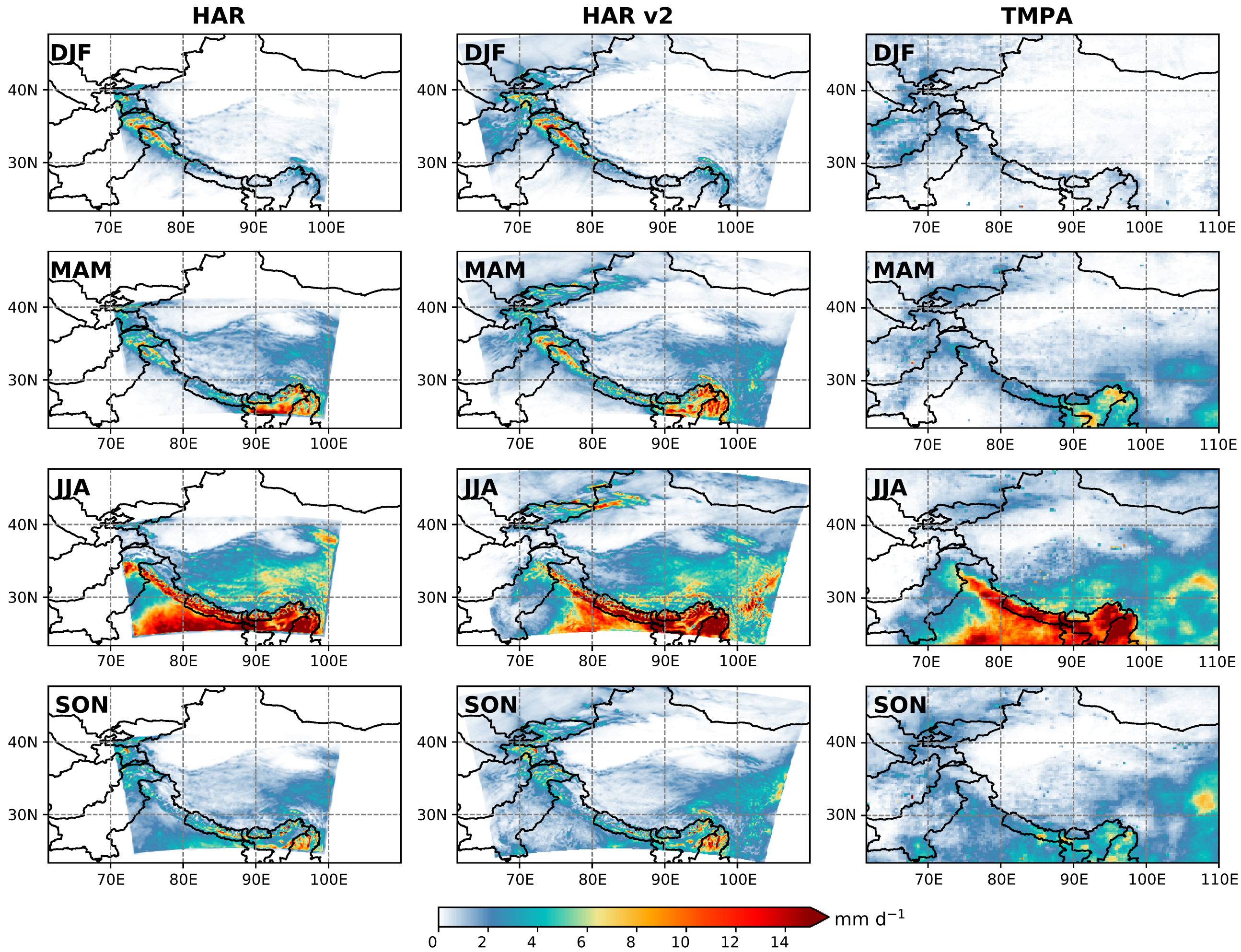

FIGURE 8 Seasonal mean Prcp from HAR, HAR v2 and TMPA in 2011

corrected initial conditions improve winter T2 simulations.

To further explore the influence of snow depth on T2, monthly mean difference of T2, surface albedo and snow cover fraction between COMB and COMB_S in January and July are presented in Figure 10. In January, regions with the highest increase of T2 in COMB_S are Southeastern TP, the Himalayas, and the Tian Shan. These regions also feature lower albedo and lower snow cover fraction. The same relationship can be found in July. However, the increase of T2 is much weaker in July than in January, which is in accordance with the MB scores in Table 5.

In the Noah LSM, surface albedo (α) is calculated as:

$$\alpha = \alpha_{\mathrm{sn,free}} + f_{\mathrm{sn}} \cdot (\alpha_{\mathrm{sn}} - \alpha_{\mathrm{sn,free}}) \tag{11}$$

where $\alpha_{\mathrm{sn,\ free}}$ and α_{sn} refer to snow-free albedo and snow albedo, respectively. f_{sn} is snow cover fraction. α is proportional to f_{sn}, which is defined as:

$$f_{\mathrm{sn}} = \begin{cases} 1 - e^{-a_s \frac{W}{W_{\mathrm{cr}}}} + \dfrac{W}{W_{\mathrm{cr}}} e^{-a_s}, W < W_{\mathrm{cr}} \\ 1, W \geq W_{\mathrm{cr}} \end{cases} \tag{12}$$

In Equation 12, W is snow depth in water equivalent; W_{cr} is a land-use-dependent threshold of snow depth, over which f_{sn} is set to 1; a_{s} is a distribution shape parameter. After snow correction, W decreases in most grid points in d10km (Figure 2), which leads to smaller f_{sn} according to Equation 12. Smaller f_{sn} results in lower surface albedo (Equation 11) and modifies the surface energy balance by reflecting less short-wave radiation, and thus, influences T2 in COMB_S. On the other hand, more absorption of short-wave radiation could lead to larger moist static energy (Meng *et al.*, 2014), which could be a possible reason of the enhanced precipitation in COMB_S (Table 5). Note that snow-related changes affect T2 and Prcp through multiple complex processes. Further work is needed to quantify the impact of these processes on T2 and Prcp.

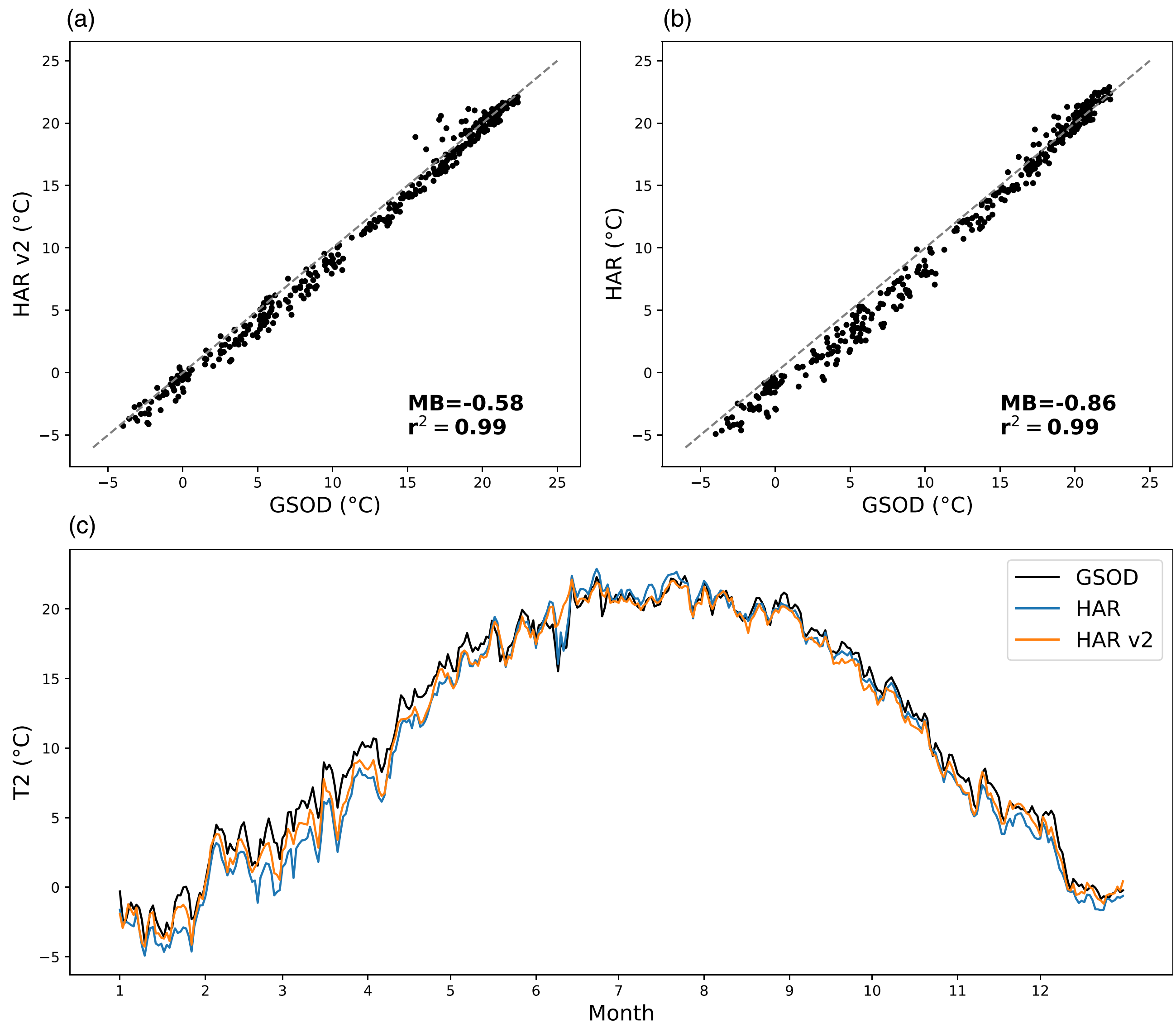

FIGURE 9 Comparison of daily T2 from HAR, HAR v2 and GSOD averaged over 57 GSOD stations (red points in Figure 1b) in 2011. Scatter plots of HAR v2 (a) and HAR (b) with statistical scores MB in K and r^2. Daily T2 time series (c)

The error sources of dynamical downscaling are twofold: deficiencies in model dynamics and physics, as well as inaccuracies in forcing data. After snow correction, the cold bias over TP is reduced but not eliminated. The distribution of T2 bias score in COMB_S (Figure 6e,f) share some similar pattern with the HAR (figure 2.4 in Maussion, 2014): cold bias over the TP but warm bias in Pakistan and Northwestern China. This implies that these patterns of biases are related to WRF itself, or the same errors exist in both ERA5 and FNL data sets. Cold bias over the TP is reported from many other WRF studies independent of the forcing data (ERA-interim: Gao *et al*., 2015; Karki *et al*., 2017; Bonekamp *et al*., 2018; Huang and Gao, 2018; FNL: Huang and Gao, 2018; Maussion, 2014; Zhou *et al*., 2018). Meng *et al*. (2018) pointed out that this is due to the overestimation of albedo over snow-covered areas in Noah LSM. After switching off albedo parameterization and replacing the albedo with MODIS time-varying albedo, their simulated T2 improved. According to Chen *et al*. (2014), the major weakness of LSM is snow processes, and Noah LSM needs a higher snow albedo to retain snow on ground since it only considers a single layer of snowpack.

5 | CONCLUSIONS

Validation of the reference experiment (Ref) indicates that, due to the changes in forcing data and domain configuration, the PPSs used in the HAR are no longer

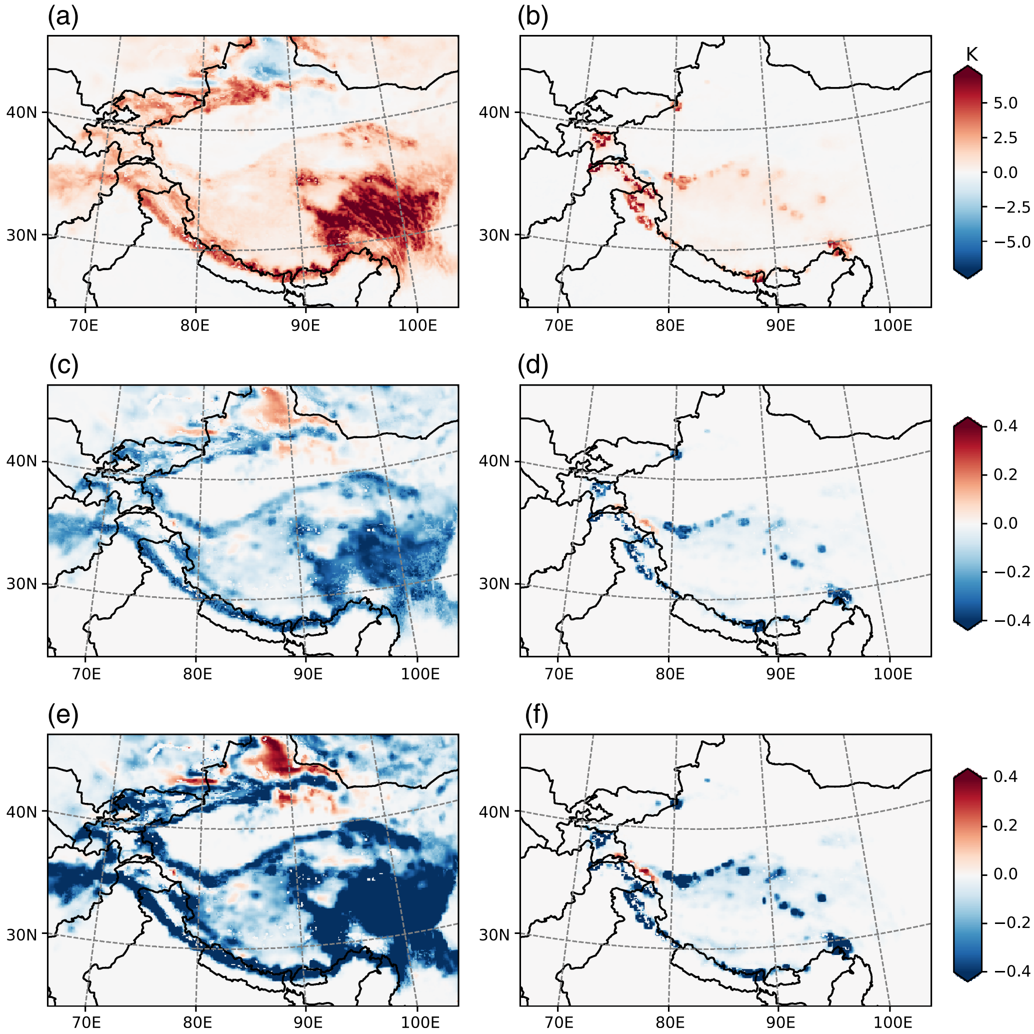

FIGURE 10 Difference of T2 (a and b), surface albedo (c and d) and snow cover fraction (e and f) between COMB_s and COMB in January (a, c and e) and July (b, d and f)

suitable for the HAR v2 since they produce large summer wet bias and winter cold bias. We examined model sensitivities to different PPSs and initial snow conditions to find an optimal set-up for the HAR v2. The results of PPS experiments reveal that no single PPS is optimally suited for both seasons (January and July) and both quantities (T2 and Prcp). Trade-offs always exist between quantities and between seasons.

Kain-Fritsch cumulus potential scheme, Morrison 2-moment scheme, Yonsei University scheme and Noah LSM are identified by the TOPSIS method as the best schemes. The combination run (COMB) inherits from these schemes and has a superior performance in Prcp but not in T2 when compared to Ref.

Overestimation of snow depth in ERA5 in most areas of d10km leads to higher snow cover fraction and higher surface albedo, which ultimately results in a cold bias. After correction of initial snow depth, cold bias is significantly reduced but not eliminated. This could imply a deficiency of Noah LSM. In future LSM development, improvement of representation of snow-related processes is needed.

After careful selection of PPSs and correction of initial snow depth, model performance is improved. The improved set-up was applied to produce 1 year of HAR v2 data for 2011. Compared to the old version, the HAR v2 generally produces slightly higher Prcp amounts, but the spatial distribution of seasonal Prcp matches better to observations. Both versions of the HAR show cold bias in spring, autumn and winter, but warm bias in summer. The HAR v2 has smaller magnitudes of these biases.

The HAR v2 is planned to have a temporal coverage of at least 30 years. Comprehensive validation against observations and comparison with the HAR are scheduled once data production is finished. The HAR v2 is developed within the framework of the "Climatic and Tectonic Natural Hazards in Central Asia (CaTeNA)" project to investigate the climatic triggering mechanism of landslides in Central Asia. However, application of this data set will not be limited to this research field. With high resolution and extended temporal coverage, the HAR v2 can contribute to a better understanding of climate-related processes in the remote and data-sparse HMA region.

ACKNOWLEDGMENTS

This work was supported by the German Federal Ministry of Education and Research (BMBF) under the framework of the "Climatic and Tectonic Natural Hazards in Central Asia (CaTeNA)" project (Grant Number FKZ 03G0878G). The authors acknowledge support by the Open Access Publication Fund of TU Berlin. The authors thank the two anonymous reviewers for their constructive comments and suggestions, which helped the authors to improve the manuscript.

ORCID

Xun Wang https://orcid.org/0000-0003-3294-747X
Marco Otto https://orcid.org/0000-0003-4464-2673
Dieter Scherer https://orcid.org/0000-0002-3670-0864

REFERENCES

Archer, D.R. and Fowler, H.J. (2004) Spatial and temporal variations in precipitation in the upper Indus Basin, global teleconnections and hydrological implications. *Hydrology and Earth System Sciences*, 8, 47–61.

Barlage, M., Tewari, M., Chen, F., Miguez-Macho, G., Yang, Z.L. and Niu, G.Y. (2015) The effect of groundwater interaction in north American regional climate simulations with WRF/Noah-MP. *Climatic Change*, 129, 485–498.

Behrangi, A., Tian, Y., Lambrigtsen, B.H. and Stephens, G.L. (2014) What does CloudSat reveal about global land precipitation detection by other spaceborne sensors? *Water Resources Research*, 50, 4893–4905.

Berg, L.K., Gustafson, W.I., Jr., Kassianov, E.I. and Deng, L. (2013) Evaluation of a modified scheme for shallow convection: implementation of CuP and case studies. *Monthly Weather Review*, 141, 134–147.

Bonekamp, P.N.J., Collier, E. and Immerzeel, W.W. (2018) The impact of spatial resolution, land use, and spinup time on resolving spatial precipitation patterns in the Himalayas. *Journal of Hydrometeorology*, 19, 1565–1581.

Bookhagen, B. and Burbank, D.W. (2010) Toward a complete Himalayan hydrological budget: spatiotemporal distribution of snowmelt and rainfall and their impact on river discharge. *Journal of Geophysical Research: Earth Surface*, 115, F03019.

Bougeault, P. and Lacarrere, P. (1989) Parameterization of orography-induced turbulence in a mesobeta—scale model. *Monthly Weather Review*, 117, 1872–1890.

Cai, X., Yang, Z.L., David, C.H., Niu, G.Y. and Rodell, M. (2014) Hydrological evaluation of the Noah-MP land surface model for the Mississippi River basin. *Journal of Geophysical Research: Atmospheres*, 119, 23–38.

Chai, T. and Draxler, R.R. (2014) Root mean square error (RMSE) or mean absolute error (MAE)?—arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7, 1247–1250.

Chen, F., Barlage, M., Tewari, M., Rasmussen, R., Jin, J., Lettenmaier, D., Livneh, B., Lin, C., Miguez-Macho, G., Niu, G. Y., Wen, L. and Yang, Z.L. (2014) Modeling seasonal snowpack evolution in the complex terrain and forested Colorado headwaters region: a model intercomparison study. *Journal of Geophysical Research: Atmospheres*, 119, 13–795.

Chen, S.H. and Sun, W.Y. (2002) A one-dimensional time dependent cloud model. *Journal of the Meteorological Society of Japan*, 80, 99–118.

Chen, X., Anel, J.A., Su, Z., de la Torre, L., Kelder, H., van Peet, J. and Ma, Y. (2013) The deep atmospheric boundary layer and its significance to the stratosphere and troposphere exchange over the Tibetan plateau. *PLoS One*, 8(2).e56909.

Collier, E. and Immerzeel, W.W. (2015) High-resolution modeling of atmospheric dynamics in the Nepalese Himalaya. *Journal of Geophysical Research: Atmospheres*, 120, 9882–9896.

Copernicus Climate Change Service (C3S) (2017) *ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. Copernicus Climate Change Service Climate Data Store (CDS)*. Available at https://cds.climate.copernicus.eu/cdsapp#!/home [Accessed 10th November 2018]

Curio, J. and Scherer, D. (2016) Seasonality and spatial variability of dynamic precipitation controls on the Tibetan plateau. *Earth System Dynamics*, 7, 767–782.

Curio, J., Maussion, F. and Scherer, D. (2015) A 12-year high-resolution climatology of atmospheric water transport over the Tibetan plateau. *Earth System Dynamics*, 6, 109–124.

Dai, F.C. and Lee, C.F. (2002) Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology*, 42, 213–228. https://doi.org/10.1016/S0169-555X(01)00087-3.

Duchon, C.E. and Essenberg, G.R. (2001) Comparative rainfall observations from pit and aboveground rain gauges with and without wind shields. *Water Resources Research*, 37, 3253–3263.

Dudhia, J. (1989) Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *Journal of the Atmospheric Sciences*, 46, 3077–3107.

Feser, F., Rockel, B., von Storch, H., Winterfeldt, J. and Zahn, M. (2011) Regional climate models add value to global model data: a review and selected examples. *Bulletin of the American Meteorological Society*, 92, 1181–1192.

Gao, Y., Xiao, L., Chen, D., Chen, F., Xu, J. and Xu, Y. (2017) Quantification of the relative role of land-surface processes and large-scale forcing in dynamic downscaling over the Tibetan plateau. *Climate Dynamics*, 48, 1705–1721.

Gao, Y., Xu, J. and Chen, D. (2015) Evaluation of WRF mesoscale climate simulations over the Tibetan plateau during 1979–2011. *Journal of Climate*, 28, 2823–2841.

Grell, G.A. (1993) Prognostic evaluation of assumptions used by cumulus parameterizations. *Monthly Weather Review*, 121, 764–787.

Grell, G.A. and Devenyi, D. (2002) A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophysical Research Letters*, 29, 38–31.

Grell, G.A. and Freitas, S.R. (2014) A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmospheric Chemistry and Physics*, 14, 5233–5250.

Hasson, S., Gerlitz, L., Schickhoff, U., Scholten, T. and Böhner, J. (2016) Recent climate change over High Asia. In: *Climate Change, Glacier Response, and Vegetation Dynamics in the Himalaya*. Cham, Switzerland: Springer, pp. 29–48. https://doi.org/10.1007/978-3-319-28977-9_2.

Hong, S.Y., Noh, Y. and Dudhia, J. (2006) A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, 134, 2318–2341.

Hong, Y., Adler, R. and Huffman, G. (2007) Use of satellite remote sensing data in the mapping of global landslide susceptibility. *Natural Hazards*, 43, 245–256.

Hu, X.M., Nielsen-Gammon, J.W. and Zhang, F. (2010) Evaluation of three planetary boundary layer schemes in the WRF model. *Journal of Applied Meteorology and Climatology*, 49, 1831–1844.

Huang, D. and Gao, S. (2018) Impact of different reanalysis data on WRF dynamical downscaling over China. *Atmospheric Research*, 200, 25–35.

Huffman, G.J. and Bolvin, D.T. (2015) Real-time TRMM multi-satellite precipitation analysis data set documentation. *NASA TechDoc*.

Huintjes, E., Sauter, T., Schröter, B., Maussion, F., Yang, W., Kropáček, J., Buchroithner, M., Scherer, D., Kang, S. and Schneider, C. (2015) Evaluation of a coupled snow and energy balance model for Zhadang glacier, Tibetan plateau, using glaciological measurements and time-lapse photography. *Arctic, Antarctic, and Alpine Research*, 47, 573–590.

Hwang, C.L. and Yoon, K.P. (1981) Multiple attribute decision making. In: *Methods and Applications: A State-of-the-Art Survey*, Vol. 186. Berlin, Germany: Springer.

ICIMOD. (2011) *Glacial Lakes and Glacial Lake Outburst Floods in Nepal, Technical Report*. Kathmandu, Nepal: ICIMOD.

Immerzeel, W.W., Wanders, N., Lutz, A.F., Shea, J.M. and Bierkens, M.F.P. (2015) Reconciling high-altitude precipitation in the upper Indus basin with glacier mass balances and runoff. *Hydrology and Earth System Sciences*, 19, 4673–4687.

Jänicke, B., Meier, F., Fenner, D., Fehrenbach, U., Holtmann, A. and Scherer, D. (2017) Urban–rural differences in near-surface air temperature as resolved by the Central Europe refined analysis (CER): sensitivity to planetary boundary layer schemes and urban canopy models. *International Journal of Climatology*, 37, 2063–2079.

Janjic, Z.I. (1994) The step-mountain eta coordinate model: further developments of the convection, viscous sublayer, and turbulence closure schemes. *Monthly Weather Review*, 122, 927–945.

Jena, P., Azad, S. and Rajeevan, M. (2015) Statistical selection of the optimum models in the CMIP5 dataset for climate change projections of Indian monsoon rainfall. *Climate*, 3, 858–875.

Jiménez, P.A., Dudhia, J., González-Rouco, J.F., Navarro, J., Montávez, J.P. and García-Bustamante, E. (2012) A revised scheme for the WRF surface layer formulation. *Monthly Weather Review*, 140, 898–918.

Kala, J., Andrys, J., Lyons, T.J., Foster, I.J. and Evans, B.J. (2015) Sensitivity of WRF to driving data and physics options on a seasonal time-scale for the southwest of Western Australia. *Climate Dynamics*, 44, 633–659.

Karki, R., Gerlitz, L., Schickhoff, U., Scholten, T. and Böhner, J. (2017) Quantifying the added value of convection-permitting climate simulations in complex terrain: a systematic evaluation of WRF over the Himalayas. *Earth System Dynamics*, 8, 507–528.

Kirschbaum, D., Adler, R., Adler, D., Peters-Lidard, C. and Huffman, G. (2012) Global distribution of extreme precipitation and high-impact landslides in 2010 relative to previous years. *Journal of Hydrometeorology*, 13, 1536–1551.

Kleczek, M.A., Steeneveld, G.J. and Holtslag, A.A. (2014) Evaluation of the weather research and forecasting mesoscale model for GABLS3: impact of boundary-layer schemes, boundary conditions and spin-up. *Boundary-Layer Meteorology*, 152, 213–243.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K. and Takahashi, K. (2015) The JRA-55 reanalysis: general specifications and basic characteristics. *Journal of the Meteorological Society of Japan*, 93, 5–48.

Kropáček, J., Neckel, N., Tyrna, B., Holzer, N., Hovden, A., Gourmelen, N., Schneider, C., Buchroithner, M. and Hochschild, V. (2015) Repeated glacial lake outburst flood threatening the oldest Buddhist monastery in North-Western Nepal. *Natural Hazards and Earth System Sciences*, 15, 2425–2437.

Leduc, M. and Laprise, R. (2009) Regional climate model sensitivity to domain size. *Climate Dynamics*, 32, 833–854.

Lenderink, G., Buishand, A. and Deursen, W.V. (2007) Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach. *Hydrology and Earth System Sciences*, 11, 1145–1159.

Leung, L.R., Qian, Y. and Bian, X. (2003) Hydroclimate of the western United States based on observations and regional climate simulation of 1981–2000. Part I: seasonal statistics. *Journal of Climate*, 16, 1892–1911.

Li, D., Bou-Zeid, E., Barlage, M., Chen, F. and Smith, J.A. (2013) Development and evaluation of a mosaic approach in the WRF-Noah framework. *Journal of Geophysical Research: Atmospheres*, 118, 11–918.

Li, D., Yang, K., Tang, W., Li, X., Zhou, X. and Guo, D. (2020) Characterizing precipitation in high altitudes of the western Tibetan plateau with a focus on major glacier areas. *International Journal of Climatology*.

Li, J., Liu, Z., Yao, Z. and Wang, R. (2019) Comprehensive assessment of coupled model Intercomparison project phase 5 global climate models using observed temperature and precipitation over mainland Southeast Asia. *International Journal of Climatology*, 39(10), 4139–4153.

Lo, J.C.F., Yang, Z.L. and Pielke, R.A., Sr. (2008) Assessment of three dynamical climate downscaling methods using the weather research and forecasting (WRF) model. *Journal of Geophysical Research: Atmospheres*, 113, D09112.

Maussion, F. (2014) *A new atmospheric dataset for High Asia: development, validation and applications in climatology and in glaciology*. PhD Thesis, Techinische Universität Berlin, Germany, http://doi.org/10.14279/depositonce-3979.

Maussion, F., Scherer, D., Finkelnburg, R., Richters, J., Yang, W. and Yao, T. (2011) WRF simulation of a precipitation event over the Tibetan plateau, China—an assessment using remote sensing and ground observations. *Hydrology and Earth System Sciences*, 15, 1795–1817.

Maussion, F., Scherer, D., Mölg, T., Collier, E., Curio, J. and Finkelnburg, R. (2014) Precipitation seasonality and variability over the Tibetan plateau as resolved by the high Asia reanalysis. *Journal of Climate*, 27, 1910–1927.

Meng, X.H., Evans, J.P. and McCabe, M.F. (2014) The influence of inter-annually varying albedo on regional climate and drought. *Climate Dynamics*, 42, 787–803.

Meng, X., Lyu, S., Zhang, T., Zhao, L., Li, Z., Han, B., Li, S., Ma, D., Chen, H., Ao, Y., Luo, S., Shen, Y., Guo, J. and Wen, L. (2018) Simulated cold bias being improved by using MODIS time-varying albedo in the Tibetan plateau in WRF model. *Environmental Research Letters*, 13, 044028.

Miguez-Macho, G., Stenchikov, G.L. and Robock, A. (2004) Spectral nudging to eliminate the effects of domain position and geometry in regional climate model simulations. *Journal of Geophysical Research: Atmospheres*, 109, D13104.

Mlawer, E.J., Taubman, S.J., Brown, P.D., Iacono, M.J. and Clough, S.A. (1997) Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research: Atmospheres*, 102, 16663–16682.

Mölg, N., Bolch, T., Rastner, P., Strozzi, T. and Paul, F. (2018) A consistent glacier inventory for Karakoram and Pamir derived from Landsat data: distribution of debris cover and mapping challenges. *Earth System Science Data*, 10, 1807–1827.

Mölg, T., Maussion, F. and Scherer, D. (2014) Mid-latitude westerlies as a driver of glacier variability in monsoonal high Asia. *Nature Climate Change*, 4, 68–73.

Morrison, H., Thompson, G. and Tatarskii, V. (2009) Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: comparison of one-and two-moment schemes. *Monthly Weather Review*, 137, 991–1007.

NASA (n.d.) *Precipitation measurement missions*. Available at https://pmm.nasa.gov/gpm/flight-project/dpr [Accessed 30th August 2019].

Niu, G.Y., Yang, Z.L., Mitchell, K.E., Chen, F., Ek, M.B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M. and Xia, Y. (2001) The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research: Atmospheres*, 116, D12109.

NOAA (2019) *Integrated Surface Database (ISD)*. Available at https://www.ncdc.noaa.gov/isd [Accessed 30th August 2019].

Norris, J., Carvalho, L.M., Jones, C., Cannon, F., Bookhagen, B., Palazzi, E. and Tahir, A.A. (2017) The spatiotemporal variability of precipitation over the Himalaya: evaluation of one-year WRF model simulation. *Climate Dynamics*, 49, 2179–2204.

Orr, A., Listowski, C., Couttet, M., Collier, E., Immerzeel, W., Deb, P. and Bannister, D. (2017) Sensitivity of simulated summer monsoonal precipitation in Langtang Valley, Himalaya, to cloud microphysics schemes in WRF. *Journal of Geophysical Research*, 122(12), 6298–6318.

Orsolini, Y., Wegmann, M., Dutra, E., Liu, B., Balsamo, G., Yang, K., de Rosnay, P., Zhu, C., Wang, W. and Senan, R. (2019) Evaluation of snow depth and snow-cover over the Tibetan plateau in global reanalyses using in-situ and satellite remote sensing observations. *The Cryosphere*, 13, 2221–2239.

Pritchard, D.M., Forsythe, N., Fowler, H.J., O'Donnell, G.M. and Li, X.F. (2019) Evaluation of upper Indus near-surface climate representation by WRF in the high Asia refined analysis. *Journal of Hydrometeorology*, 20, 467–487.

Qian, Y., Yan, H., Berg, L.K., Hagos, S., Feng, Z., Yang, B. and Huang, M. (2016) Assessing impacts of PBL and surface layer schemes in simulating the surface–atmosphere interactions and precipitation over the tropical ocean using observations from AMIE/DYNAMO. *Journal of Climate*, 29(22), 8191–8210.

Raju, K.S. and Kumar, D.N. (2014) Ranking general circulation models for India using TOPSIS. *Journal of Water and Climate Change*, 6, 288–299.

Roessner, S., Wetzel, H.U., Kaufmann, H. and Sarnagoev, A. (2005) Potential of satellite remote sensing and GIS for landslide hazard assessment in southern Kyrgyzstan (Central Asia). *Natural Hazards*, 35, 395–416.

Sikder, S. and Hossain, F. (2016) Assessment of the weather research and forecasting model generalized parameterization schemes for advancement of precipitation forecasting in monsoon-driven river basins. *Journal of Advances in Modeling Earth Systems*, 8, 1210–1228.

Skamarock, W.C. and Klemp, J.B. (2008) A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *Journal Computational Physics*, 227, 3465–3485.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., Wang, W., Powers, J. G., Duda, M. G., Barker, D. M. and Huang, X.-Y. (2019): A Description of the Advanced Research WRF Version 4. NCAR Tech. Note NCAR/TN-556+STR, 145 pp..

Stergiou, I., Tagaris, E. and Sotiropoulou, R.E.P. (2017) Sensitivity assessment of WRF parameterizations over Europe. *Multidisciplinary Digital Publishing Institute Proceedings*, 1, 119.

Tewari, M., Chen, F., Wang, W., Dudhia, J., LeMone, M.A., Mitchell, K., Ek, M., Gayno, G., Wegiel, J. and Cuenca, R.H. (2004) Implementation and verification of the unified NOAH land surface model in the WRF model. In: *20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction*, Vol. 1115. Seattle, WA: American Meteorological Society.

Thompson, G., Field, P.R., Rasmussen, R.M. and Hall, W.D. (2008) Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: implementation of a new snow parameterization. *Monthly Weather Review*, 136, 5095–5115.

Tomasi, E., Giovannini, L., Zardi, D. and de Franceschi, M. (2017) Optimization of Noah and Noah_MP WRF land surface schemes in snow-melting conditions over complex terrain. *Monthly Weather Review*, 145, 4727–4745.

Tzeng, G.H. and Huang, J.J. (2011) *Multiple Attribute Decision Making: Methods and Applications*. Boca Raton: Chapman and Hall/CRC.

Wilks, D.S. (2006) *Statistical Methods in the Atmospheric Sciences*, 2nd edition. Burlington: Academic press.

Willmott, C.J. and Matsuura, K. (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30, 79–82.

WMO (Ed.). (2008) Precipitation gauge errors and corrections. In: *Guide to Meteorological Instruments and Methods of Observation*, 7th edition. Geneva: WMO-8, I.6-6–I.6-7. World Meteorological Organisation.

Xie, B., Fung, J.C., Chan, A. and Lau, A. (2012) Evaluation of non-local and local planetary boundary layer schemes in the WRF model. *Journal of Geophysical Research: Atmospheres*, 117, D12103.

Xu, L., Liu, H., Du, Q. and Xu, X. (2019) The assessment of the planetary boundary layer schemes in WRF over the central Tibetan plateau. *Atmospheric Research*, 230, 104644.

Yang, K., Koike, T., Fujii, H., Tamura, T., Xu, X., Bian, L. and Zhou, M. (2004) The daytime evolution of the atmospheric boundary layer and convection over the Tibetan plateau: observations and simulations. *Journal of the Meteorological Society of Japan*, 82(6), 1777–1792.

Yang, Z.L., Niu, G.Y., Mitchell, K.E., Chen, F., Ek, M.B., Barlage, M., Longuevergne, L., Manning, K., Niyogi, D., Tewari, M. and Xia, Y. (2011) The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins. *Journal of Geophysical Research: Atmospheres*, 116, D12110.

Zhou, X., Yang, K. and Wang, Y. (2018) Implementation of a turbulent orographic form drag scheme in WRF and its application to the Tibetan plateau. *Climate Dynamics*, 50, 2443–2455.

Zou, H., Hu, Y., Li, D., Lü, S. and Ma, Y. (2005) Seasonal transition and its boundary layer characteristics in Anduo area of Tibetan plateau. *Progress in Natural Science*, 15(3), 239–245.

SUPPORTING INFORMATION

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

How to cite this article: Wang X, Tolksdorf V, Otto M, Scherer D. WRF-based dynamical downscaling of ERA5 reanalysis data for High Mountain Asia: Towards a new version of the High Asia Refined analysis. *Int J Climatol*. 2021;41: 743–762. https://doi.org/10.1002/joc.6686

Supplement to

WRF-based Dynamical Downscaling of ERA5 Reanalysis Data for High Mountain Asia: Towards a New Version of the High Asia Refined Analysis

Xun Wang[1], Vanessa Tolksdorf[1], Marco Otto[1], Dieter Scherer[1]

[1]Chair of Climatology, Institute of Ecology, Techinische Universität Berlin, Berlin, 12165, Germany

Correspondence to: Xun Wang, Techinische Universität Berlin, Institute of Ecology, Chair of Climatology, Rothenburgstraße 12, Berlin 12165, Germany. E-mail: xun.wang@tu-berlin.de

Table S1. Statistical scores for the reference experiment (Ref) and the HAR over 57 GSOD stations within d10km of the HAR (red points in Figure 1b) for Prcp (mm d^{-1}) and T2 (K). The values that differ from Ref of more than 10% are marked in bold and put between parentheses if they show a lower skill than Ref.

	Prcp January				Prcp July				T2 January				T2 July			
	MAE	MB	RMSE	r_s	MAE	MB	RMSE	r_s	MAE	MB	RMSE	r_s	MAE	MB	RMSE	r_s
Ref	0.24	0.19	0.47	0.59	2.21	1.25	3.48	0.34	2.82	-2.18	3.83	0.79	0.95	-0.14	1.22	0.95
HAR	**0.20**	**0.14**	**0.36**	**0.67**	**1.77**	**0.27**	**2.82**	**(0.15)**	2.67	**-1.72**	3.49	**(0.61)**	**(1.55)**	**0.08**	**(2.15)**	0.87

Sensitivity experiment: direct downscaling of ERA5 to 10 km resolution

The experiment is designed as follows:

- Simulation period: 2011-07-01
- Nesting strategy: ERA5 is directly downscaled to 10 km without d30km as the parent domain
- Other settings are the same as Ref

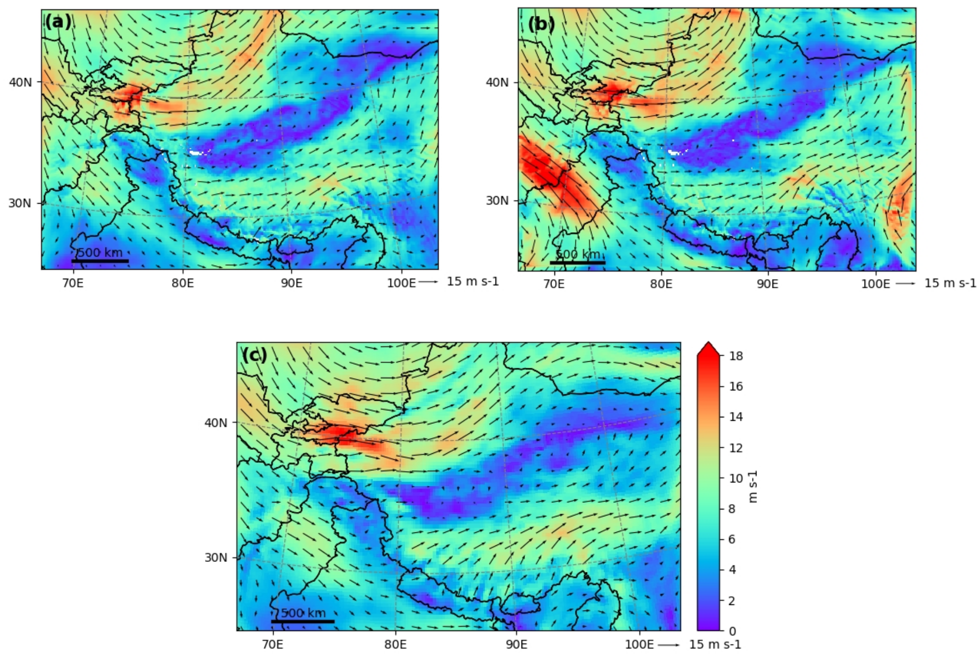

Figure S1. Comparison of 500 hPa wind vectors and wind speed (contour) from (a) two-way nesting approach; (b) direct downscaling approach; and (c) ERA5.

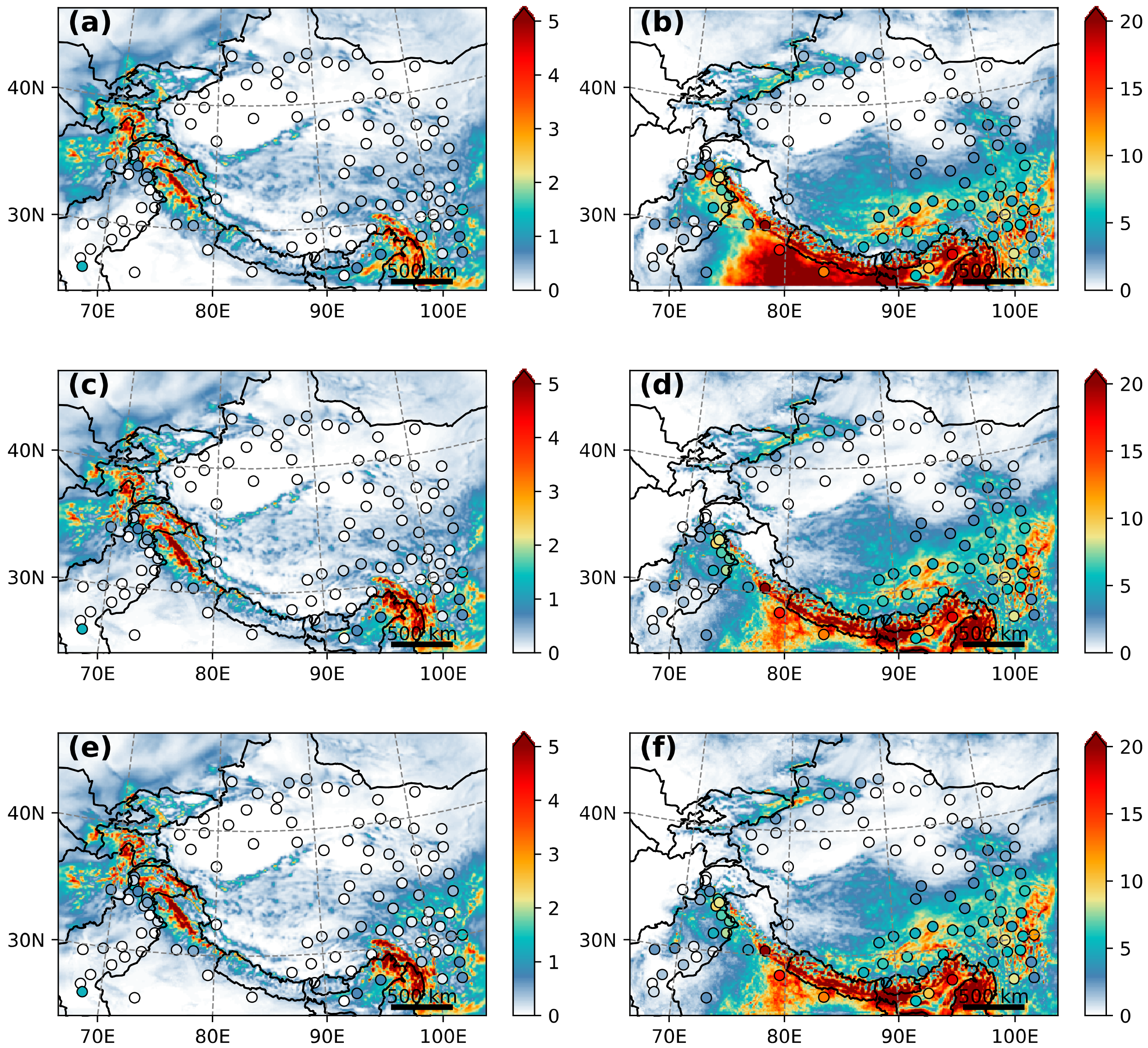

Figure S2. Monthly mean Prcp (mm d-1) for Ref (a, b), COMB (c, d) and COMB_S (e, f) in January (a, c, e) and July (b, d, f) with the in-situ Prcp observations from GSOD embedded as circles in each map. Note that, the scale of the color bars is different for January and July.

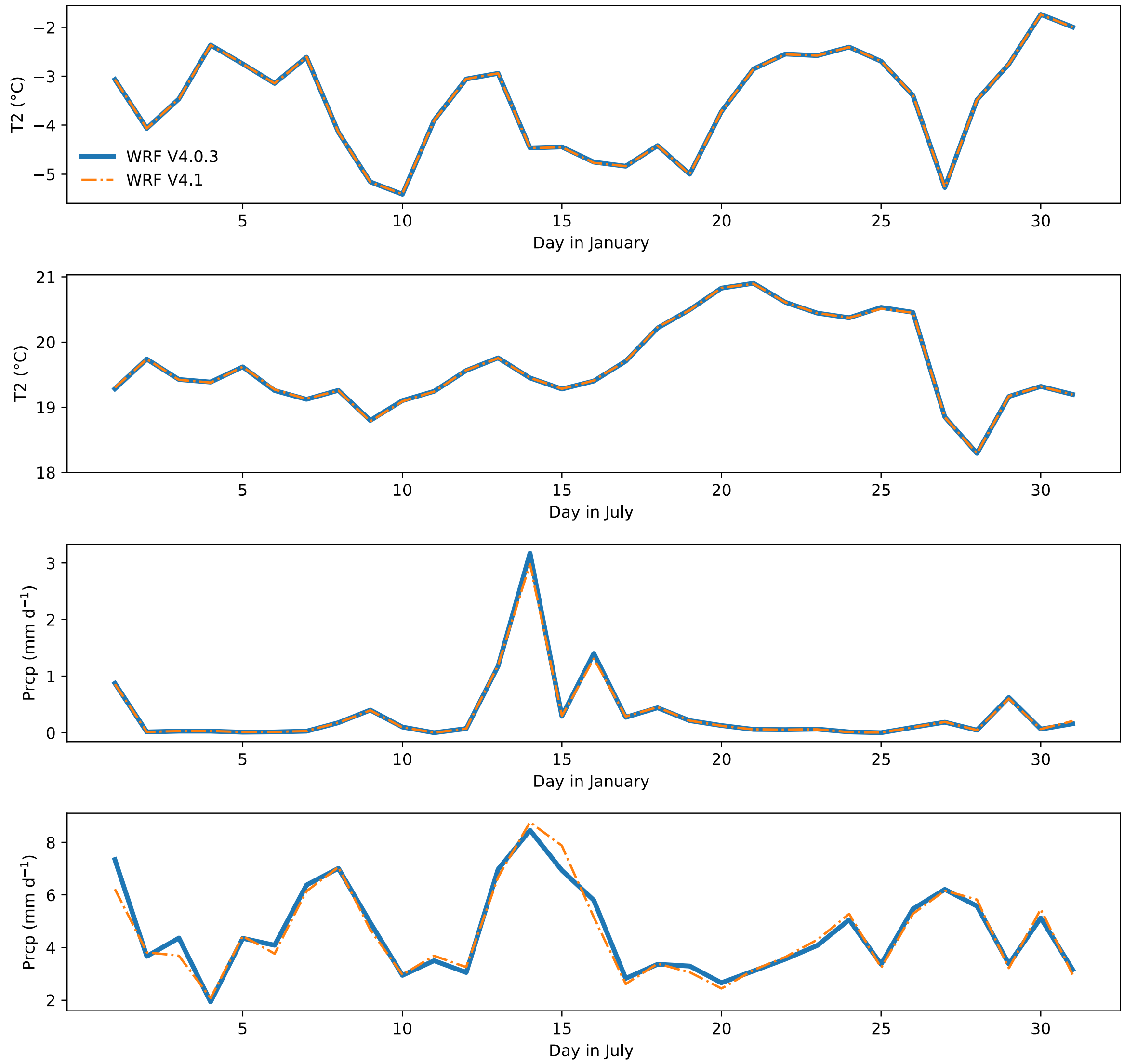

Figure S3. Impact of WRF versions on daily T2 and Prcp averaged over 103 GSOD stations (red and white points in Figure 1b) in January and July.

Paper II

Intercomparison of gridded precipitation datasets over a sub-region of Central Himalaya and the southwestern Tibetan Plateau

Hamm, A., Arndt, A., Kolbe, C., **Wang, X.**, Thies, B., Boyko, O., Reggiani, P., Scherer, D., Bendix, J. and Schneider, C., 2020: Intercomparison of Gridded Precipitation Datasets over a Sub-Region of the Central Himalaya and the Southwestern Tibetan Plateau. *Water*, 12(11), 3271, https://doi.org/10.3390/w12113271.

Status: Published.

Own contribution:

- numerical simulations
- data post-processing
- contribution to writing of manuscript and reviewing of text
- scientific discussion

water

MDPI

Article

Intercomparison of Gridded Precipitation Datasets over a Sub-Region of the Central Himalaya and the Southwestern Tibetan Plateau

Alexandra Hamm [1,2], Anselm Arndt [1,*], Christine Kolbe [3], Xun Wang [4], Boris Thies [3], Oleksiy Boyko [5], Paolo Reggiani [5], Dieter Scherer [4], Jörg Bendix [3] and Christoph Schneider [1]

1. Geography Department, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany; alexandra.hamm@natgeo.su.se (A.H.); christoph.schneider@geo.hu-berlin.de (C.S.)
2. Department of Physical Geography, Stockholm University, Svante Arrhenius väg 8, 11418 Stockholm, Sweden
3. Laboratory for Climatology and Remote Sensing, Department of Geography, Philipps-Universität Marburg, Deutschhausstrasse 12, 35032 Marburg, Germany; christine.kolbe@geo.uni-marburg.de (C.K.); thies@staff.uni-marburg.de (B.T.); bendix@geo.uni-marburg.de (J.B.)
4. Chair of Climatology, Technische Universität Berlin, Rothenburgstrasse 12, 12165 Berlin, Germany; xun.wang@tu-berlin.de (X.W.); dieter.scherer@tu-berlin.de (D.S.)
5. Department of Civil Engineering, University of Siegen, Paul-Bonatz-Str. 9-11, 57068 Siegen, Germany; oleksiy.boyko@uni-siegen.de (O.B.); paolo.reggiani@uni-siegen.de (P.R.)

* Correspondence: anselm.arndt@geo.hu-berlin.de

Received: 23 October 2020; Accepted: 16 November 2020; Published: 21 November 2020

Abstract: Precipitation is a central quantity of hydrometeorological research and applications. Especially in complex terrain, such as in High Mountain Asia (HMA), surface precipitation observations are scarce. Gridded precipitation products are one way to overcome the limitations of ground truth observations. They can provide datasets continuous in both space and time. However, there are many products available, which use various methods for data generation and lead to different precipitation values. In our study we compare nine different gridded precipitation products from different origins (ERA5, ERA5-Land, ERA-interim, HAR v2 10 km, HAR v2 2 km, JRA-55, MERRA-2, GPCC and PRETIP) over a subregion of the Central Himalaya and the Southwest Tibetan Plateau, from May to September 2017. Total spatially averaged precipitation over the study period ranged from 411 mm (GPCC) to 781 mm (ERA-Interim) with a mean value of 623 mm and a standard deviation of 132 mm. We found that the gridded products and the few observations, with few exceptions, are consistent among each other regarding precipitation variability and rough amount within the study area. It became obvious that higher grid resolution can resolve extreme precipitation much better, leading to overall lower mean precipitation spatially, but higher extreme precipitation events. We also found that generally high terrain complexity leads to larger differences in the amount of precipitation between products. Due to the considerable differences between products in space and time, we suggest carefully selecting the product used as input for any research application based on the type of application and specific research question. While coarse products such as ERA-Interim or ERA5 that cover long periods but have coarse grid resolution have previously shown to be able to capture long-term trends and help with identifying climate change features, this study suggests that more regional applications, such as glacier mass-balance modeling, require higher spatial resolution, as is reproduced, for example, in HAR v2 10 km.

Keywords: precipitation; reanalysis data; satellite retrieval; complex terrain; spatial resolution; temporal resolution; High Mountain Asia; Tibetan Plateau; third pole

1. Introduction

High Mountain Asia (HMA) is the major water source of large river systems, especially of the Yangtze, the Yellow, the Brahamputra, the Ganges and the Indus river. It forms the freshwater supply for billions of people in Asia who depend on it as a drinking and agriculture water supply or source for hydropower electricity, and it is among the most vulnerable water towers globally [1,2]. Hence, it is becoming increasingly important to monitor and model water availability as the climate is changing. The three main direct sources of water in HMA rivers are direct precipitation, snow melt and glacier runoff, all of which experience drastic changes due to increasing temperatures and altered precipitation patterns [3–6].

Observing precipitation constitutes a challenge, especially in complex terrain with harsh climatic conditions and limited access [7]. Precipitation measured with rain-gauge stations can provide information about spatial and temporal patterns, and they are therefore essential for monitoring and modeling. Direct observations at rain-gauge stations are (i) only available as point measurements; (ii) sparsely and unevenly distributed in space, especially in remote areas such as HMA; (iii) error-prone, especially for solid precipitation; and (iv) often discontinuous in time [8–14]. Further limitations arise when comparing different gauge stations among each other due to different instrumentation and site characteristics. A heated tipping bucket will give different results than a non-heated bucket, and vegetation types and changes over time can influence measured precipitation and possible interpretations about what has caused these changes [15].

To inform various research applications, such as hydrological models, precipitation data need to be continuous in both space and time. For this purpose, weather model-derived reanalysis datasets may provide spatially homogeneous gridded data. Gridded precipitation data can also be derived from interpolation of ground observations, which are subject to considerable uncertainties in data-scarce areas such as HMA [16]. Retrieving precipitation from satellites is another method for generating gridded data. Precipitation measurement missions such as the Tropical Rainfall Measuring Mission (TRMM) [17] and the Global Precipitation Measurement Mission (GPM) [18] were established to continuously observe precipitation from space.

The choice of dataset to use for hydrological modeling applications greatly impacts the results, as there are significant differences between both absolute and relative values among datasets [4,7,19–22]. It is an inherent feature of the research problem that it is not possible to ultimately determine whether any of the datasets provides the "true" value of precipitation. Nevertheless, it is possible to make an informed decision about the choice of dataset by knowing about the differences, limitations and similarities, and through validation against ground truth data. Depending on the study area, some datasets may outperform others.

A major issue with gridded precipitation in rugged terrain, such as HMA, is the accurate representation of a grid-mean value that represents the local variability of precipitation. The terrain heterogeneity and topographical features get smoothed out in coarse-grid resolution products. It has been shown that the comparison between observed and modeled elevation within a global climate model leads to a bias of up to 2 km in elevation over HMA with higher inaccuracies on the edges of the Tibetan Plateau, which shows the highest gradients in topography [23]. Besides the effect of altitude as such on the amount of precipitation, it can cause inaccuracies in spatial rainfall estimates due to local-scale dynamics of convective precipitation resulting from thermal slope breeze systems or orographically-induced precipitation.

The comparison of gridded data to actual measurements is problematic. Even though they are used in the majority of studies (e.g., [4,21,22]), ground observation stations are also not fully representative of the areas of the grid cells in which they are located. Usually, gauge stations are located in valley bottoms rather than on top of the mountains or on slopes. Further error sources of gauge station data are the undercatch due to wind drift, especially during snowfall, wetting and measurement inconsistencies [8,13,15,24]. However, as surface measurements are the only ground truth observations of precipitation, they are also used as a reference in this study.

The scope of this study is to compare the global reanalysis datasets ERA5 [25], ERA5-Land and ERA-Interim [25,26], the Japanese 55-year Reanalysis (JRA-55) [27] and the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) [28], the regional WRF-downscaled High Asia Refined analysis version 2–10 km domain (HAR v2 10 km) [29] and High Asia Refined analysis version 2–2 km domain (HAR v2 2 km) [29] gridded products, the station based precipitation dataset Global Precipitation Climatology Centre (GPCC) and the satellite derived precipitation product Precipitation REtrieval covering the TIbetan Plateau (PRETIP) [30,31]. Further information on spatial and temporal resolutions of the datasets and websites for data downloads are shown in Table 1. In a case study, we compared these datasets over a data-scarce sub-region covering each parts of the Tibetan Plateau (TiP), the Himalaya and the Himalaya foothills to the south during May to September 2017. To achieve a comprehensive intercomparison, we combined and extended different commonly used methods to inter-compare precipitation datasets and quantify differences based on terrain complexity. We finally compared gridded to rain-gauge data from the Chinese Ministry of Water Resources.

Comparable, longer-term comparisons across HMA have been carried out by e.g., Li et al., [20], who found that grid resolution plays a significant role in overall mean precipitation and local maximum precipitation, that observation-derived datasets are likely to underestimate precipitation due to their locations in the valley bottoms and that satellite products show high uncertainties, especially for solid precipitation. Similarly, Gao et al. [4] used precipitation indices to compare ERA-Interim reanalysis with WRF-downscaled products based on ERA-Interim and the community climate system model (CCSM) for the historical period and future projections over the Tibetan Plateau. They found that both ERA-Interim and CCSM greatly overestimate mean and extreme precipitation indices when compared to observation data. The dynamically downscaled products generally outperform their forcings in terms of absolute precipitation accuracy, and spatial and temporal patterns, indicating the importance of resolving small-scale processes. Similar conclusions were drawn by Huang and Gao [19], stating that ERA-Interim and final analysis data from the Global Forecasting System (GFS-FNL) datasets largely overestimate precipitation over the Tibetan Plateau (TiP). This wet bias is reduced in WRF-downscaled products. Further work by Yoon et al. [21] studied the terrestrial water budget over HMA, comparing different gridded precipitation data as boundary conditions for land surface models, including the older HAR (High Asia Refined analysis) version [32]. Mean estimates of precipitation were found to differ significantly between products, while the spatial patterns and seasonality were reasonably captured in all products. The first HAR version has also been evaluated by Pritchard et al. [33], who found that it is capable of representing precipitation in the Upper Indus Basin at multiple scales and matches ground observation data well. Furthermore, Wang and Zeng [7] used several predecessors of the current study over the TiP and found that the Global Land Data Assimilation Systems (GLADS) data has the overall best performance for precipitation when compared to station data over the 1992–2004 period. GLDAS is derived as a combination from surface observations and remote sensing. Additionally, Bai et al. [22] investigated different precipitation datasets over the Qinghai-Tibet Plateau, highlighting the importance of precipitation data in data-scarce regions and complex terrain such as the TiP. In their study satellite products, blended satellite and gauge station measurements, and climate modeling data, such as the HAR dataset, have been compared. They conclude that extreme precipitation is generally overestimated, while light precipitation (less than 1 mm day^{-1}) was mostly underestimated by most products.

In our study, we complement those earlier studies by including the new and even higher spatially resolved HAR v2 10 km and HAR v2 2 km datasets, and by applying additional ways of comparing different gridded precipitation datasets. We emphasize that differences between datasets must be discussed based on season, precipitation type and spatial context. With a set of selected analysis methods, our aim was to address the following key research questions: (1) How similar are the various gridded precipitation datasets? (2) What is the effect of terrain complexity on variations in precipitation between products?

2. Data and Methods

In order to address the proposed question, we compiled a set of methods to compare the datasets. Similarities and differences are mostly related to grid-cell based values and how the various products represent precipitation at the same location and the same time or period. In this section, we present the study region, the datasets used for the intercomparison and the methods applied to address similarities and differences.

2.1. Study Area and Period

The study area encompasses parts of the TiP, the Himalayas and the Himalaya foothills (Figure 1). It stretches from 81° E to 88° E and from 28° N to 32° N (about 230,000 km^2). We chose this study area to include different topographic features, and to represent the transition from the central parts of the Himalayas to the Tibetan Plateau and the Transhimalaya. From southwest to northeast, the first part represents the low-lying southern slopes of the Himalaya, followed by the extreme relief of the Himalayas, and the less complex TiP terrain. The study period was set from May to September 2017, which is the first year in which PRETIP precipitation can be considered. Further, the period covers a full Indian Summer Monsoon season, which exhibits the most interesting features in precipitation for any kind of research application in the study area. The 2017 monsoon season was also unobtrusive in the amount and length of the monsoon precipitation, making it a suitable study period. The choice of the study area was further motivated in the course of follow-up research by Kropáček et al. [34] dealing with glacier lake outburst floods in the Limi Valley originating from the small Halji glacier in northwestern Nepal, which is located within the boundaries of the present study area (close to the west-station in Figure 1).

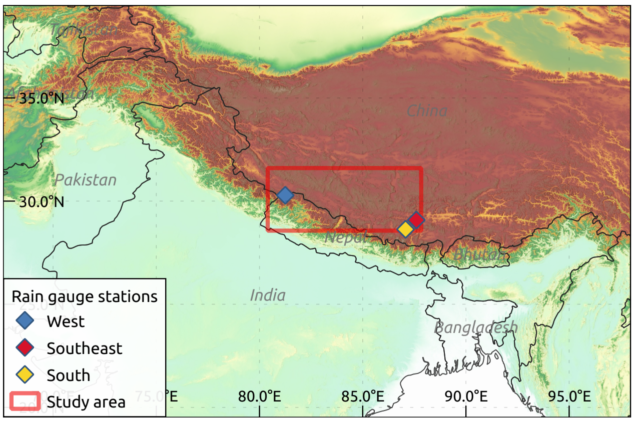

Figure 1. Overview of the study area and the 3 rain gauge stations located within the boundaries of the area.

2.2. Data

The datasets used in this study and their respective properties are listed in Table 1. For comparison purposes, all datasets were aggregated to daily sums. As with other precipitation datasets that do not cover either the study period or study area, we have excluded the Aphrodite dataset [35] from the analysis in this study, which is often used in precipitation comparisons in Asia.

Table 1. Overview datasets.

Dataset	Temporal Resolution	Spatial Resolution (Approx.)	Temporal Coverage	Spatial Coverage
ERA5 [36] [1]	1 h	30 km	1979—near real time	global
ERA5-Land [25] [2]	1 h	9 km	1981—near real time	global
ERA-Interim [26] [3]	6 h	80 km	1979–August 2019	global
HAR v2 10 km [29] [4]	1 h	10 km	2004–2018	HMA only
HAR v2 2 km [29]	1 h	2 km	April–October 2017	study area
JRA-55 [27] [5]	1 h	55 km	1958—near real time	global
MERRA-2 [28] [6]	1 h	55×69 km	1980—near real time	global
PRETIP [30,31] [7]	30 min	4 km	May 2017–September 2017	TiP
GPCC [37] [8]	d	111 km	January 2009–present	global

[1] https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5, [2] https://www.ecmwf.int/en/era5-land, [3] https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim, [4] http://www.klima.tu-berlin.de/HARv2, [5] https://climatedataguide.ucar.edu/climate-data/jra-55, [6] https://climatedataguide.ucar.edu/climate-data/nasa-merra, [7] https://doi.org/10.5678/LCRS/DAT.395, [8] https://climatedataguide.ucar.edu/climate-data/gpcp-monthly-global-precipitation-climatology-project.

European Centre for Medium-Range Weather Forecasts datasets. In the framework of their so-called reanalysis project the European Centre for Medium-Range Weather Forecasts (ECMWF) offers different atmospheric reanalyses. With the exception of their global coverage they differ in spatial and temporal resolution, in temporal coverage (cf. Table 1) and in the applied parameterizations. In the present study, we used the three latest products ERA-Interim, ERA5 and ERA5-Land. Please note that ERA5-Land uses the same atmospheric forcing as ERA5, interpolating the data to a higher grid resolution (see ERA5-Land documentation (https://confluence.ecmwf.int/display/CKB/ERA5-Land%3A+data+documentation#ERA5Land:datadocumentation-LandSurfaceModel)). Therefore, it was not expected to see considerable differences between ERA5 and ERA5-Land. The gridded output variables have been downloaded from the Copernicus Climate Change Service (C3S) Climate Date Store.

High Asia Refined analysis version 2. The High Asia Refined analysis version 2 (HAR v2) is an atmospheric dataset generated by dynamical downscaling of ERA5 reanalysis data. The regional climate model used for this purpose is the Weather Research and Forecasting model version 4.1 (WRF V4.1, [38]). In contrast to traditional regional climate simulations, WRF is re-initialized daily and integrated over 36 h with the first 12 h discarded as spin-up time. The HAR v2 provides meteorological fields at 10 km grid spacing and hourly temporal resolution. The 10 km domain covers the whole TiP and the surrounding mountains. The HAR v2 is described in detail by Wang et al. [29]. The dataset currently covers the period from 2004 to 2018 and will be both extended back to 1979 and updated continuously into the future. To investigate the influence of horizontal grid spacing on precipitation simulation, ERA5 has also been downscaled to 2 km grid spacing using WRF V4.1 for the study area from April 2017 to October 2017 (hereinafter HAR v2 2 km). The model setup for HAR v2 2 km was the same as HAR v2 10 km, except that no cumulus parameterization scheme was used for HAR v2 2 km and cumulus convection was thus explicitly resolved.

Precipitation REtrieval covering the TIbetan Plateau. PRETIP is a new satellite-based precipitation retrieval dataset for the TiP and originates from a feasibility study, which aimed at the combination of the brightness temperatures from the geostationary satellites Insat-3D and Elektro-L2 for precipitation retrieval [39,40]. PRETIP was trained using a random forest approach. The reference for the model training is GPM (Global Precipitation Measurement Mission) IMERG (Integrated Multi-satellite Retrievals for GPM) from which only the rain gauge calibrated microwave precipitation data are used [41]. Gauge calibrated microwave precipitation is the most reliable precipitation estimate from space thus far [18,42,43]. The temporal coverage is restricted to May–September 2017 due to the limited availability of Elektro-L2. PRETIP has the same temporal resolution as IMERG, which is 30 min, and is available in both 11 and 4 km resolutions. This increase in resolution from 11 to 4 km constitutes the advantage of PRETIP over IMERG. The spatial coverage is confined by the Tibetan Plateau and areas above 2500 m a.s.l., which does only partly cover the study area (c.f. Figure 5). Further, PRETIP is limited by the availability of microwave data, which are not available for every single 30-min timestep. Scenes for which no microwave based precipitation but satellite data (Insat-3D, Elektro-L2) are available were modeled using a daily model, which was built from the microwave based precipitation available on that day. However, due to the lack of availability of Insat-3D and Elektro-L2 at some time slots, some data gaps exist. Therefore, the daily product only contains the available timesteps. The number of available scenes per day is illustrated in Figure A1 in the Appendix A. For further details about PRETIP please refer to Kolbe et al. [30,31] .

Japanese 55-year Reanalysis. JRA-55 is the second reanalysis project carried out by The Japan Meteorological Agency [27]. Observations used in JRA-55 consist of those used in ERA-40 [44] and an additional array of observations listed in the former paper. The product utilizes a four dimensional variance analysis (4D-VAR) for data assimilation. The spatial resolution is $0.56^{\circ} \times 0.56^{\circ}$ and it covers the period from 1958 to near real-time. We obtained the dataset through The Data Support Section facilities at the National Center of Atmospheric Research, and for purposes of the paper, accumulated 6-hourly precipitation values to daily sums.

Modern-Era Retrospective analysis for Research and Applications, Version 2. MERRA-2 is the second version of the Modern-Era Retrospective analysis for Research and Applications produced by NASA's Global Modeling and Assimilation Office. It replaces its predecessor, MERRA, by including additional observations and updates to the Goddard Earth Observing System model and analysis scheme. It has been available in 1-hourly temporal resolution and $0.5^{\circ} \times 0.625^{\circ}$ spatial resolution in near real-time since 1980.

Global Precipitation Climatology Centre. The GPCC First Guess Daily Product is a global gridded daily precipitation estimate based on station data. The measurements undergo automatic quality control, and are interpolated between grid cells using an ordinary block kriging [37]. The spatial resolution of the grid is 1° latitude by 1° longitude and the dataset is available from January 2009 until near real-time. Within our study area, a total of three gauge stations are used to derive daily precipitation.

Ground observations. For a ground validation of the precipitation products we resorted to the collection of precipitation data provided by the Chinese Ministry of Water Resources and collected by the hydrometerological service of Tibet. The amount of precipitation was measured by tipping bucket rain gauges installed according to World Meteorological Organization standards over the period 2007–2015. The network, albeit sparse given the size of the area, provides the only set of ground observations available to assess the gridded precipitation datasets. The stations of network used in this study are shown in Figure 1.

2.3. Methods

2.3.1. Correlation Coefficient

To compare the different precipitation products, we used the non-parametric Spearman's rank correlation coefficient, R, which describes how similar the spatial pattern of precipitation is within the compared grids on a daily or multi-daily basis. Due to the different spatial resolutions, for each pair of products, we aggregated the higher resolution product to match the grid resolution of the lower resolved product within each comparison. Similarities between various generations from the same source (ERA products) and different spatial resolutions of the same product (HAR v2 products) can help to assess variations resulting from diverse methodologies and parameterizations in the generations of these datasets. We used different temporal aggregation intervals to assess whether the timing of precipitation events is different within the products and whether multi-day-sums increase their similarities. Correlations were only derived for grid cells with valid values in both datasets.

2.3.2. Comparison to Station Data

To obtain an approximation of ground truth precipitation, we utilized three rain-gauge stations within our study area that provide daily precipitation sums. We compared their cumulative sums over the study period to the cumulative sum of the respective grid cell in the precipitation products. We extracted the elevation of each station from the Advanced Land Observing Satellite (ALOS) Digital elevation model (DEM), provided by ALOS World 3D—30 m (AW3D30) of the Japanese Aerospace Exploration Agency (https://www.eorc.jaxa.jp/ALOS/en/aw3d30/index.htm). The stations' elevations were then compared to the modeled elevation of the grid cell for the reanalysis and WRF-downscaled products, and to the mean elevation of the grid cell for PRETIP and GPCC (derived from ALOS, Table 2). These comparisons provide insights into the possible reasons for differences between ground-based weather station observations and gridded reanalysis or satellite data, because the generation of several products relies on the topography, and thus the resolution of the underlying digital elevation model.

Table 2. Modeled elevation (m a.s.l) of used grid cell.

Dataset	West	Southeast	South
Rain gauge	4134 *	4320 *	4476 *
ERA5	4824	4995	4944
ERA5-Land	4415	4359	4507
ERA-Interim	3573	4856	4919
HAR v2 10 km	4448	4682	4615
HAR v2 2 km	4151	4505	4465
JRA-55	3810	4887	4887
MERRA-2	4007	3512	2989
PRETIP	4243 *	4234 *	4467 *
GPCC	4903 *	4907 *	4907 *

* derived from ALOS.

2.3.3. Climdex

Climate indices are usually used to quantify how climate has changed over long periods, how it differs in space or to identify and track climate extremes (e.g., [45]). In this study, we used a set of climdex indices to compare the different precipitation datasets similarly to Gao et al. [4]. The indices used in this study are R1, R10, R20, Rx1, Rx5 and PTOT. They were calculated for every grid cell and summarized for the different products. An overview over the different indices and their definitions is given in Table 3.

Table 3. Selection of climdex indices used in this study for intercomparison between different precipitation (P) products.

Index	Definition	Unit
R1	number of wet days ($P > 1$ mm)	days
R10	number of wet days with $P > 10$ mm	days
R20	number of wet days with $P > 20$ mm	days
Rx1	maximum 1-day precipitation	mm
Rx5	maximum 5-day precipitation	mm
PTOT	total precipitation	mm

2.3.4. Terrain Complexity

There are various options to geometrically and statistically define terrain complexity [46]. In this study, we assessed the influence of terrain complexity on the differences between the precipitation datasets on the basis of the ALOS DEM, as illustrated in Figure 2. Two levels of complexity are defined by the standard deviation (SD) of elevation from the high resolution ALOS-grid cells within single grid cells according to the product with the lowest resolution (GPCC). Complexity is defined as low or high based on the percentiles of SD of grid cells. For high complexity, we set a threshold at the 75% percentile (Q3) of SD among all grid cells. This means that 25% of the grid-cells above this threshold are classified as "high complexity." The remaining 75% of the grid-cells represent "low complexity." For each product, we calculated the mean difference between the products with regard to terrain complexity in order to derive its potentially varying influence on rainfall calculation. In order to compare products with different spatial resolutions, we resampled all products to the coarsest common denominator grid (GPCC, 111 km, 24 grid cells).

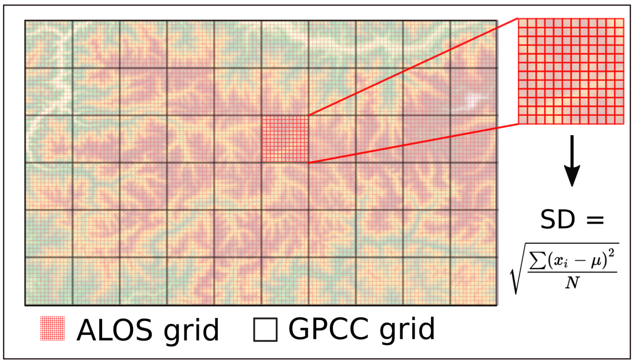

Figure 2. Schematic overview of the method applied to derive terrain complexity. Black lines represent the grid of the lowest resolved precipitation product (GPCC), red lines represent the grid of the ALOS digital elevation model (DEM). The topography in the background is an example topography. In the equation to calculate the DEM standard deviation (SD) in each GPCC grid cell, x_i stands for the values within the ALOS DEM cell, μ for the overall mean and N for the number of ALOS DEM grid cells within each GPCC grid cell.

3. Results

3.1. Statistical Analysis

In this section, we describe and visualize the datasets used for comparison and the results of the statistical analysis.

To illustrate how the different precipitation products compare within the study region and period, we provide the cumulative sum of precipitation from May to September 2017 (Figure 3), the sum of precipitation for each month within the study period (Figure 4), and a spatial plot with per-pixel sums over the study period (Figure 5).

Overall, the per-pixel sum (cf. Figure 3) is between 600 and 800 mm for all ECMWF products, the WRF-downscaled HAR products and JRA-55. MERRA-2 and GPCC only show 400 to 500 mm of precipitation, which results in a difference up to 100% between the datasets. Despite the missing lower-lying areas (<2500 m a.s.l.) and the fact that the daily values are built only from available satellite scenes, PRETIP amounted to 525 mm for the period between May and September 2017, which falls within the range of the other datasets.

Monthly sums (Figure 4) show that all products have their maximum precipitation in July and August, while September has the lowest values. The relative variability between datasets is greatest in the pre-monsoon season (May), while the agreement is best between most datasets in July to August (except for MERRA-2, PRETIP and GPCC). Other than for PRETIP, for which no valid values in the southwestern corner of the study area exist due to the elevation below 2500 m, the other datasets generally show highest precipitation sums in the southwest along the foothill of the Himalayas, and lowest values occur along the transition from the Himalayas to the TiP. The Himalaya range generally shows the highest spatial heterogeneity as long as the spatial resolution is sufficient to depict these small-scale changes (Figure 5). In general, it can be seen that only the HAR v2 datasets and in parts the ERA5 products are able to resolve orographic precipitation, while the resolution of the other products only gives grid values based on averages in the area. Surprisingly, the satellite product PRETIP, which has the second highest grid resolution (4 km) is not able to capture small-scale patterns of topographically-induced precipitation.

The correlation on a daily basis for each combination of datasets is given in Table 4. The highest correlation was achieved between ERA5 and ERA5-Land, with $R = 1$, while lowest correlation was found between the reanalysis product MERRA-2 and the satellite dataset PRETIP ($R = 0.33$). In general, the correlation between the ERA products and the ERA-derived products (HAR v2 10 km and HAR v2 2 km) is quite high ($R > 0.66$), suggesting that their precipitation values depicting the most probable range (cumulative values of 600–800 mm, c.f. Figure 3). The fact that they are not identical, however, shows that there are also considerable differences between the datasets, which is most likely the effect of different representations of precipitation processes at different scales and the different representations of cumulus convection in the models.

Temporally aggregating precipitation over a 5-day window generally increases the correlation (Table 5). The highest correlation can still be found between ERA5 and ERA5-Land ($R = 1$), but the lowest correlation can now be found between the observation-based product GPCC and the satellite product PRETIP ($R = 0.56$). In general, PRETIP shows the overall smallest correlation to all other products. With a mean of 0.63 and generally similar values regarding the comparison to the other datasets, PRETIP appears to have the largest differences in overall grid-based precipitation. Further aggregation of precipitation over ten days and entire months did not significantly increase correlations, indicating that most differences in the timing of precipitation between products are covered within a 5-day period (see Tables A1 and A2).

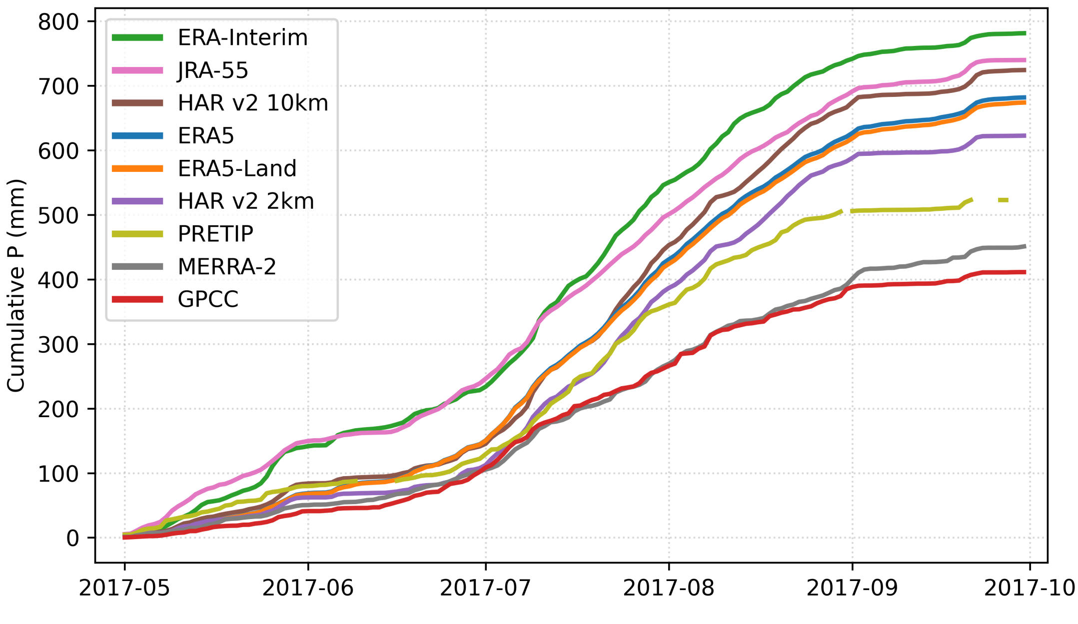

Figure 3. Spatial mean cumulative sum of precipitation throughout the study period.

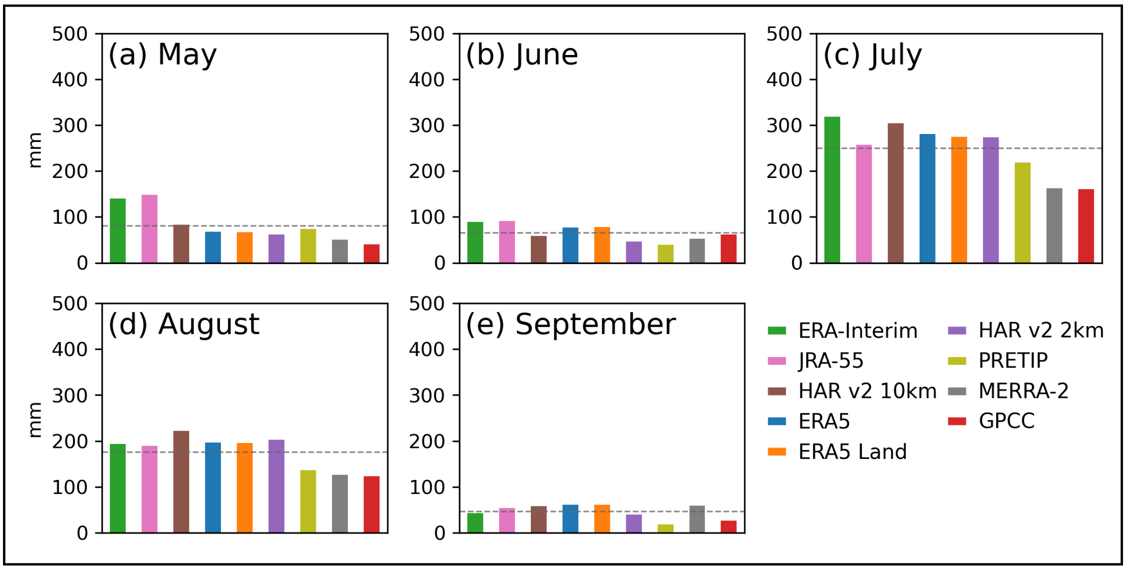

Figure 4. Spatial average monthly sum of precipitation during the study period. The gray dashed line represents the mean precipitation in each month over all datasets.

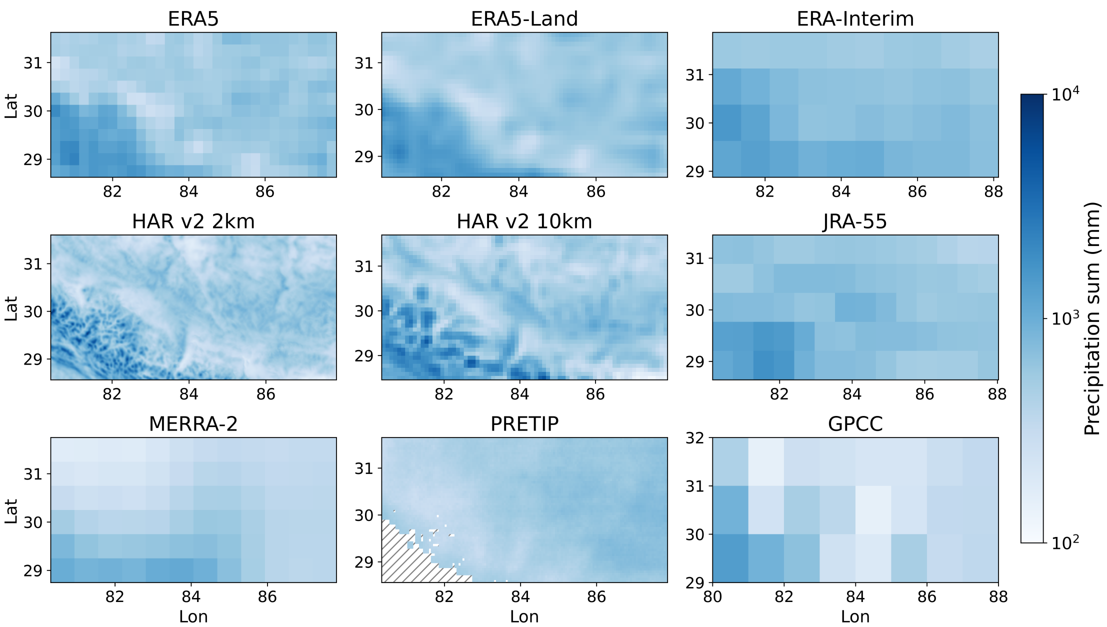

Figure 5. Spatial log-scaled per-grid-cell sum over the study period for each of the precipitation products. Sums were only calculated for valid values, which excluded the south-western corner in the PRETIP product (hatched area) and individual grid cells lower than 2500 m.a.s.l.

Table 4. Correlation coefficient **R** for all datasets in mm/day. The five highest correlations are highlighted with bold font. All correlations are statistically significant at the 99% confidence interval.

Dataset	ERA5	ERA-Interim	ERA5-Land	HAR v2 2 km	HAR v2 10 km	JRA55	MERRA2	PRETIP
ERA-Interim	0.72							
ERA5-Land	**1.00**	0.72						
HAR v2 2 km	**0.74**	0.67	0.67					
HAR v2 10 km	**0.74**	0.68	**0.74**	**0.77**				
JRA55	0.61	0.66	0.64	0.61	0.60			
MERRA2	0.50	0.48	0.53	0.48	0.48	0.44		
PRETIP	0.47	0.51	0.44	0.34	0.40	0.45	0.33	
GPCC	0.55	0.49	0.55	0.54	0.51	0.48	0.55	0.35

Table 5. Correlation coefficient **R** for all datasets in **mm/5days**. The five highest correlations are highlighted with bold font. All correlations are statistically significant at the 99% confidence interval.

Dataset	ERA5	ERA-Interim	ERA5-Land	HAR v2 2 km	HAR v2 10 km	JRA55	MERRA2	PRETIP
ERA-Interim	0.82							
ERA5-Land	**1.00**	0.82						
HAR v2 2 km	**0.85**	0.79	0.82					
HAR v2 10 km	**0.84**	0.80	**0.84**	**0.87**				
JRA55	0.69	0.76	0.71	0.72	0.70			
MERRA2	0.72	0.69	0.74	0.70	0.71	0.63		
PRETIP	0.64	0.66	0.63	0.59	0.61	0.63	0.59	
GPCC	0.77	0.68	0.78	0.73	0.73	0.67	0.74	0.56

3.2. Comparison with Rain Gauge Data

Daily values from rain gauge stations and grid values from the daily precipitation products are cumulatively summed up over the study period as illustrated in Figure 6. In general, the station data shows significantly lower values than most of the gridded products. Exceptions can be seen at the south station, where both HAR products show lower cumulative sums than the observations at the station. At the southeast station, the observed values are almost identical to the grid values of MERRA-2 and GPCC, while both HAR products only show slightly more precipitation by the end of the study period. A similar trend can be seen in the south station, where the before mentioned products represent the observations best. The other products generally show more precipitation than what is observed at these stations, up to four times as much. The west station is located in a generally dry valley, which receives, on average, less than 200 mm of annual precipitation [47]. This can be seen by the total cumulative precipitation observed at the station of only 64.6 mm. The closest gridded values are again MERRA-2, GPCC and HAR v2 2 km with about 250 mm. While both HAR products show very similar values at the south and southeast station, they are fairly different at the west station with the 10 km resolution product showing almost twice as much precipitation as the 2 km product. ERA-Interim, on the other hand, greatly overestimates precipitation in this grid-cell by 24 times as much precipitation as observed by the station. In general, the timing of precipitation is better represented between station and gridded product than the actual amount. Most products agree on the majority of precipitation falling between June and August and little precipitation from August until the end of the study period. However, the absolute differences between observed and gridded precipitation are, in parts, substantial.

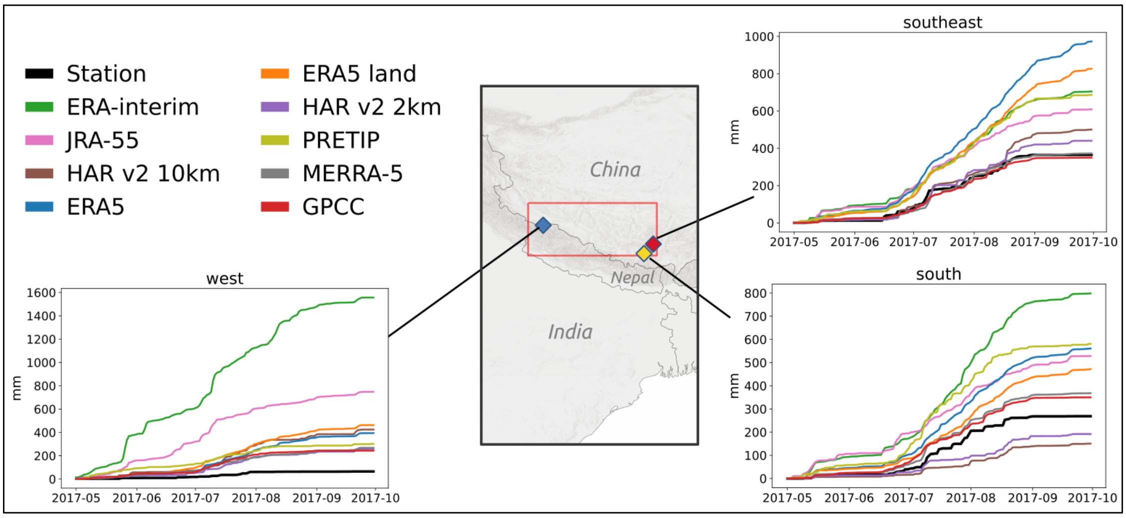

Figure 6. Cumulative sum of daily precipitation throughout the study period for the station data (black line) and the gridded precipitation products (colored lines).

3.3. Terrain Complexity

The magnitude of difference in precipitation with respect to terrain complexity is given in Figure 7. Overall, it can be seen that the difference in precipitation is consistently higher in complex terrain (red dots, SD > Q3), than in less complex terrain (blue squares, SD $\leq$ Q3). The biggest difference in precipitation can be seen between PRETIP and HAR v2 10 km with 3.9 mm d^{-1} followed by PRETIP and MERRA-2 with 3.7 mm d^{-1}, and PRETIP and ERA-Interim with 3.6 mm d^{-1}. Visually, the differences based on terrain complexity can be distributed in different groups: (i) overall low differences and small variation between high and low complexity pixels (e.g., HAR v2 10 km and ERA5-Land), (ii) overall higher differences, but small variation between high and low complexity pixels (e.g., GPCC and JRA-55) and (iii) differences spread out greatly between low and high complexity (e.g., PRETIP and HAR v2 10 km). The lowest mean difference can be seen between ERA5 and ERA5-Land with only 0.2 mm d^{-1}, which further affirms that the forcing in ERA5-Land is the same as in ERA5 and that interpolation is done linearly. The second-lowest mean difference can be seen between HAR v2 2 km and HAR v2 10 km with 0.9 mm d^{-1}. The overall mean difference between products (yellow diamond) is between 1 and 2.5 mm d^{-1}, with the highest value between GPCC and ERA-Interim, the two products with the coarsest grid resolutions. Overall, for low complexity terrain, most precipitation differences are between 0 and 2 mm d^{-1} while high complexity differences mostly range between 1.5 and 4 mm d^{-1}.

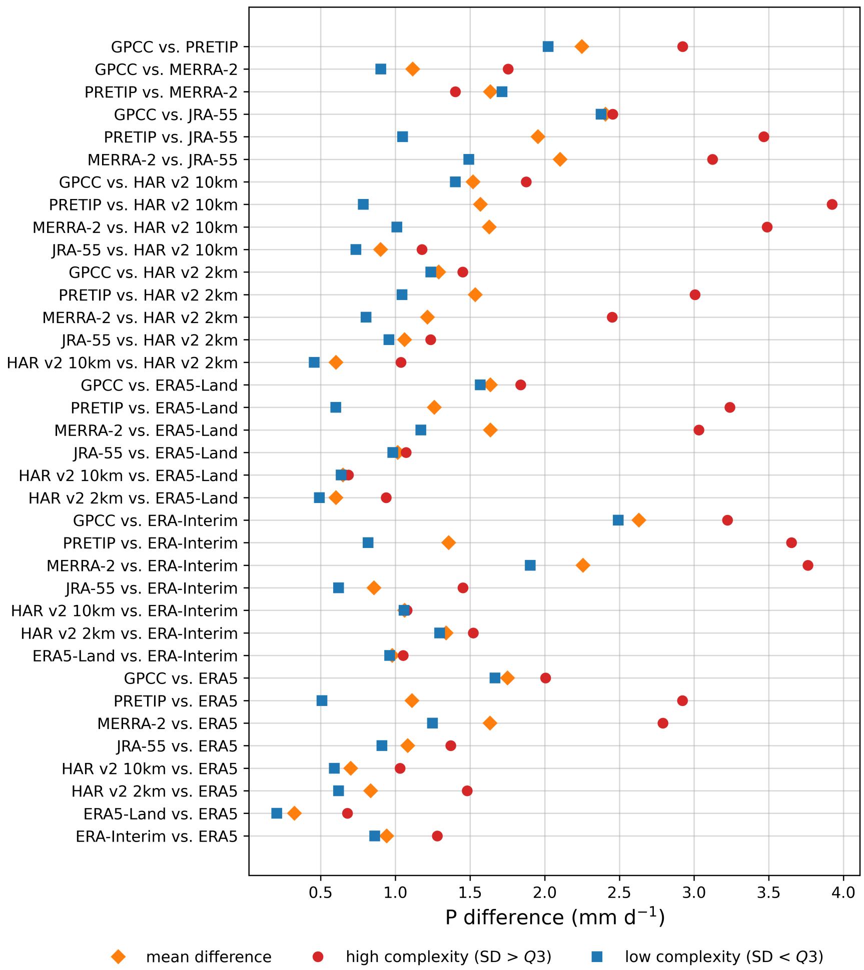

Figure 7. Absolute precipitation difference (mm day^{-1}) based on terrain complexity aligned with the coarsest grid (GPCC). Complexity is described as high (SD > Q3) or low (SD $\leq$ Q3) standard deviation of ALOS-DEM elevation within a single grid cell of the common grid. Blue rectangles represent low terrain complexity, red dots indicate high terrain complexity and the yellow diamonds depict the mean difference.

3.4. Climdex Indices

With the climdex indices, we aim at quantifying precipitation extremes for each product and compare the spatial mean. Figure 8 shows boxplot charts for each index where every value represents a single grid cell within each product. In this representation, grid resolutions were not aggregated in order to capture the full range of grid values in each product. To be able to compare the products universally, we additionally compiled an equivalent representation of the same indices but with the same coarse grid resolutions. The resulting illustration can be found in the appendix (Figure A2a). Similar overall values were found in both versions but maximum values are considerably smaller due to the spatial aggregation. In order to allow for a more straightforward comparison of original and spatially aggregated climdex data, in Figure A2b we include the data behind Figure 8 but with the scaling as used in Figure A2a. The following presentation and discussion of the results will focus on the climdex indices based on the original spatial resolution of each precipitation product as presented in Figure 8. In general, it can be seen that the higher the spatial resolution, the larger is the data range between all grid cells (except for PRETIP).

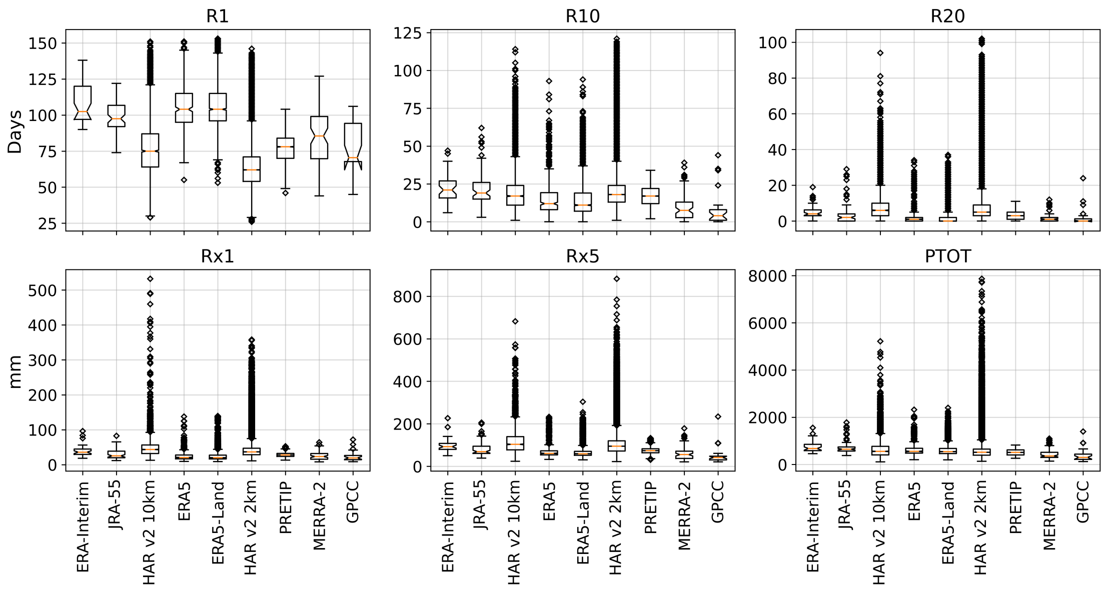

Figure 8. Visualization of the selected climdex indices R1, R10, R20, Rx1, Rx5 and PTOT as boxplots (for descriptions, see Table 3). Each box contains all grid cell values within the precipitation product. Boxes range from the 1st to 3rd quartile; the yellow line denotes the median; and whiskers indicate 1.5 fold interquartile ranges from the upper to lower boundaries. Values outside this range are displayed as black dots. Please note that the different products have different spatial resolutions.

Data points for days with more than 10 and 20 mm of precipitation (R10 and R20), show that the higher resolved products (HAR v2 10 km and HAR v2 2 km) return the overall highest values while they have much lower mean values and lower maximum values for the general wet-day count (R1). This implies that individual grid cells in higher resolved products (e.g., HAR v2 2 km) can experience more extreme precipitation events in multiple grid cells than coarser products (e.g., ERA-Interim). Higher overall median values in the extreme precipitation indices (R10 and R20) in the higher resolved products further imply the resolution of locally confined heavy precipitation events. Continuing with the extreme event indices (Rx1 and Rx5), it can be seen that the highest values can also be found within the higher resolved products, followed by a decreasing trend with decreasing grid resolution. However, it needs to be mentioned that in contrast to HAR v2 2 km, in which convective systems are explicitly resolved, HAR v2 10 km uses a cumulus parameterization scheme, which has some uncertainties and can, in rare occasions, lead to extremely high values, such as more than 500 mm in one day, which can not be found in any of the other products. The third-highest amount of precipitation in a single day

(Rx1) can be found within ERA5 and ERA5-Land with about 140 mm. Over a 5-day period (Rx5), the maximum values increase to 700 and more than 800 mm in the HAR v2 10 km and HAR v2 2 km products, respectively, while ERA5 and ERA5-Land range between 200 and 300 mm.

Total precipitation in a single grid cell is highest in HAR v2 2 km with a maximum of 7865 mm in a single grid cell. It is followed by HAR v2 10 km with 5217 mm and ERA5 with 2317 mm. MERRA-2 has the lowest maximum total precipitation with only 1100 mm over the study period. Comparing the two products with similar spatial resolution, ERA5-Land and HAR v2 10 km difference between linear interpolation and WRF-downscaling become obvious. While the grid-cell with the maximum PTOT in ERA5-Land amounts to 2400 mm, the maximum in HAR v2 2 km amounts to 7865 mm, which is more than three times as much compared to ERA5-Land within the 5 month-period.

Figure A2 reveals that the outstanding maximum values of the two HAR v2 datasets in Rx5 and PTOT are mainly a consequence of higher spatial resolution. As soon as spatial resolution is equalized by spatial aggregation, the maximum values are very much different, and the two HAR datasets do not show extra-ordinary values. In fact, in the spatially aggregated version (Figure A2a) the GPCC dataset shows the highest maximum value of Rx5, indicating that interpolation of station measurements to larger areas may negatively impact hydrological modeling.

Notably, despite its second-highest grid resolution, the satellite product PRETIP shows the smallest variation regarding precipitation rates within grid cells and few outliers in all indices, which relate to the overall more homogeneous distribution of precipitation throughout the study area in this product (c.f. Figure 5).

4. Discussion

Despite the short period of analysis presented, it is possible to discover substantial similarities and differences between the different gridded precipitation products over the study area. As observed in Figure 3, the study area is influenced by the Indian Summer Monsoon which becomes visible in the increase of precipitation during July and August and its withdrawal starting in September. Most products show a good agreement within the monsoon season, except for PRETIP, GPCC and MERRA-2. In addition, the area is also affected by the westerlies, which becomes visible in the pre-monsoon season (May). The inconsistency between JRA-55 and ERA-Interim and all other datasets might originate from different parameterizations for westerly-driven mostly solid precipitation. Combined, it appears that ERA5, ERA5-Land, HAR v2 10 km, HAR v2 2 km and for the most part PRETIP consistently match both the pre-monsoon and monsoon precipitation, while the remaining datasets have limitations in either one of those two periods.

Based on the correlation between datasets, it became obvious that some are more similar than others. ERA5-Land and ERA5 are essentially identical when aggregating ERA5-Land to ERA5 resolution. This is to be expected, as ERA5 is using ERA5 atmospheric forcing to derive land-surface parameters. Hence, it should be noted that ERA5-Land does not add any value regarding orographically-induced precipitation over ERA5 when using atmospheric data. While all the ERA products and the ERA5-derived HAR products generally are very similar, the satellite product PRETIP exhibits the lowest correlations, even after aggregating precipitation over multiple days. Considering the spatial patterns of PRETIP precipitation, it is no surprise that the correlations are low. While the other products show a spatially decreasing trend in precipitation from southwest to northeast with a highly variable region in the Himalaya mountain range, PRETIP exhibits a much more homogeneous distribution throughout the study area. It even shows lower values for the Himalaya mountain range than the area covering the TiP. This is a result of the averaging character of the random forest algorithm which is smoothing for more extreme (low and high) precipitation and tends toward average precipitation rates. In future developments, the training should be either separated for convective and stratiform precipitation, or another machine learning algorithm that better captures meteorological extremes should be developed [48,49]. On the other hand, the similarities between the ERA products, the HAR products and to some extent JRA-55 lead to the conclusion that these products display the most likely range of precipitation in this study. The differences between those modeled datasets can be attributed to differences in model dynamics. This is in line with Zhang and Li [50], who found that differences in moisture advection parameterizations greatly change precipitation patterns on steep slopes. It is not possible to ultimately say how well these products match the "true" precipitation. However, the few observations that are available suggest that the above mentioned products are the ones with precipitation amounts being the closest to actual precipitation amounts.

The comparison with rain gauge station data revealed that both HAR v2 datasets have the best matches with the ground observations. For the south and southeast stations, they also show very similar values, though they are more different at the west station. Here, the gauge station is located in a very localized, dry area, making local processes even more important. While these processes seem to be better represented in HAR v2 2 km, the 10 km grid seems to catch precipitation that might be outside the confined dry area. Considering ERA-Interim in this comparison, it becomes obvious that the extremely coarse grid resolution must be covering areas with higher precipitation outside the dry valley the station is located in. The elevation comparison between modeled elevation of ERA-Interim and the DEM-extracted elevation of the rain gauge station (Table 2) shows that the station is located higher (4134 m a.s.l) than the modeled elevation of the ERA-Interim grid cell (3573 m a.s.l.). However, even though it is to be expected that stations located in low-lying areas would exhibit less precipitation than higher-lying areas, ERA-Interim shows much higher values than the gauge station, which emphasizes the limitations of trying to explain precipitation discrepancies by solely considering altitude as the determining factor. The good match between GPCC and the ground observations can be attributed to the fact that GPCC synthesizes station-based data and interpolates between them. Hence, it is to be expected that GPCC scores high correlations with surface observations in grid cells with observations, making it useful for individual grid-cells. However, the heavily interpolated values in between distant station data are subject to extreme uncertainties as no topograhical and regional features can be captured. Generally, rain gauge stations are often located in valley bottoms and easily accessible areas. Precipitation at the adjacent mountain peak or on its slopes might be higher, which can be represented by the modeled data, but not by rain-gauge station observations.

With the six climate indices (climdex) we found that the products with the highest grid resolution exhibited the highest number of days with heavy precipitation (R10 and R20) and the largest amount of precipitation in a single day and five consecutive days (Rx1 and Rx5). On the other hand, the mean values of the wet-day count (R1) were much smaller, which is an improvement compared to ERA-Interim precipitation in particular. According to Gao et al. [4], ERA-Interim tends to overestimate precipitation on average, especially in the frequency of precipitation events. With the mean values of R1 in both HAR datasets in our case study being much lower than those in ERA-Interim, they seem to better represent the distribution of precipitation. The same feature, albeit lower in magnitude, can be observed between ERA-Interim and ERA5, indicating an improvement of precipitation representation between the two generations of ECMWF-reanalysis products in this specific case study. Overall, extreme precipitation events can occur in multiple grid cells within the higher-resolved HAR datasets. However, the cumulus parameterization in HAR v2 10 km seems to produce extremely high values of more than 500 mm in a single day, which does not happen in the 2 km grid version of the product. This finding is in accordance with Ou et al. [51], in which high-resolution WRF experiments with and without cumulus convection scheme were conducted at a gray-zone grid spacing of 9 km. They found that the experiment without a cumulus scheme generally outperforms the experiments with cumulus schemes in terms of the mean total precipitation, and the diurnal cycles of precipitation amount and frequency. The total precipitation (PTOT) for all products shows that the maximum amount of a single grid cell can vary between less than 2000 mm up to almost 8000 mm. It became obvious that this cumulative difference in precipitation over only five months will strongly impact on the results of research applications if either one or the other product is chosen for the specific location.

Overall, our findings in terms of spatial resolution are in line with other studies, suggesting that higher grid resolution is needed to accurately represent terrain-induced precipitation patterns [20]. In this study, only the HAR datasets and partly the ERA5 datasets were able to represent large orographic complexity. However, an increase in spatial resolution does not always yield higher accuracy in complex terrain, as can be seen within the PRETIP product, which is much more homogeneous than some of the lower resolved products. On the other hand, the coarsest product GPCC might perform much better in areas where individual grid cells contain measurements, while the interpolated cells in between are subject to high uncertainties. Further, GPCC has a high probability of underestimating precipitation due to the locations of ground observation stations being in valleys rather than on slopes or mountain summit areas.

The role of terrain complexity was assessed with the help of a digital elevation model. We found that all datasets displayed higher differences in precipitation when the terrain complexity (ALOS standard deviation) was larger than Q3, except for one pair (PRETIP and MERRA-2). Based on the grouping of the pairs depending on their relationship between mean difference and precipitation, for the difference between high and low complexity terrain four main clusters can be derived (Figure 9). While cluster I includes most of the similar datasets, such as ERA and HAR datasets due to their overall similarity, cluster II comprises mostly comparisons with the coarsely resolved GPCC product. The greater overall mean difference between GPCC and the other products is most likely a result of the heavily interpolated values for grid cells without measurements. However, terrain complexity does not seem to have a significant additional impact on the differences. Cluster IIIa and IIIb are mostly dominated by comparisons with PRETIP and MERRA-2. While the differences with PRETIP are attributable to the averaging nature of the random forest approach and the resulting smoothing in complex terrain, the comparisons with MERRA-2 canot be interpreted in a straight forward way. All comparisons with MERRA-2, except for the comparison between PRETIP and MERRA-2, are grouped within cluster III, which leads to the conclusion that precipitation in terrain with high complexity within MERRA-2 seems to be weaker compared to most other products. The inverse behavior of the pair PRETIP and MERRA-2 in terms of precipitation in complex terrain vs. less complex terrain is probably attributable to the fact that this pair has the lowest overall correlation for daily values and hence has the largest differences in all grid cells, independently of topography.

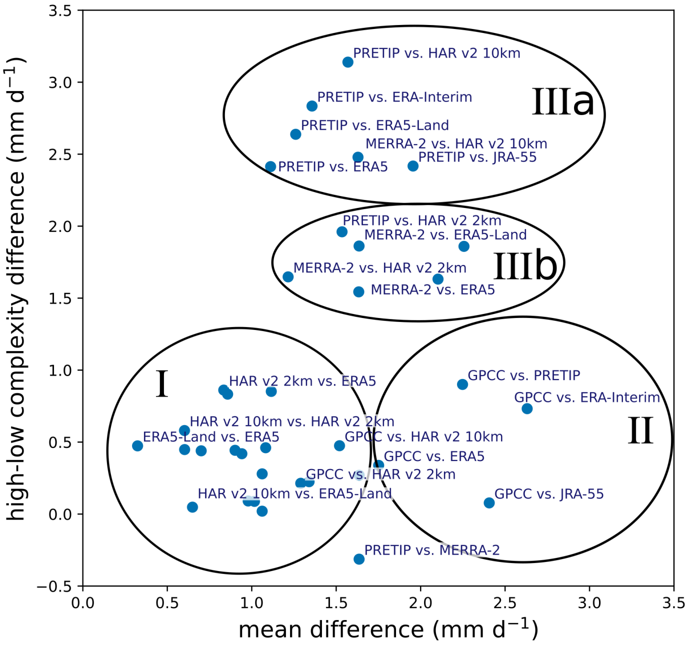

Figure 9. Visualization of precipitation differences between each two precipitation products based on the relationship between mean difference (yellow diamonds in Figure 7) and the difference between high (red dots in Figure 7) and low (blue squares in Figure 7) complexity precipitation. The groups describe: (I) low mean difference and low difference between high and low terrain complexity, (II) high mean difference but low difference with respect to terrain complexity and (III) medium overall difference but large variation depending on terrain complexity. Only some labels of all pairs as listed in Figure 7 are displayed.

5. Conclusions

This study presents the intercomparison of nine differently generated gridded precipitation products from a study area in HMA from May to September 2017. Precipitation as boundary condition for any research application can greatly influence the outcome and respective interpretation. In order to be able to understand and predict the future behavior of a system, it is necessary to apply tools, such as modeling, which require a certain spatial and temporal coverage of their input data. This is particularly challenging for remote regions with complex terrain, such as in HMA. Making an informed decision about the boundary conditions used for the respective applications is key to achieving reliable predictions and can be a difficult endeavor. In this study, we highlighted the similarities and differences of spatially and temporally continuous gridded precipitation data from various sources over one full monsoon period that can be used as boundary conditions for longer-term applications, such as climate-change assessments, runoff-calculations, glacier mass balance modeling and hydropower-applications, among others. While a product with coarse grid resolution such as ERA-Interim might be able to reproduce seasonal patterns and long-term climate trends [4], glacier modeling applications might require much higher grid resolution as for example in HAR v2 2 km, which resolves processes related to local topography much better than products based on coarser grids. However, the HAR v2 2 km product has high computational demands due to its high resolution dynamical downscaling. It is only available for distinctive study regions and periods where it is of high value to analyze the effects of grid resolution and topography. The HAR v2 10 km, on the other hand, shows very good matches with observational data and is available for a longer periods and the entire HMA. It shows slight limitations compared to the 2 km version originating from the cumulus parameterization, which can overestimate precipitation falling in a single day. Nonetheless, HAR v2 10 km is the only product (together with HAR v2 2 km) that is able to resolve topographic precipitation features (c.f. Figure 5). Similarly, gauge station data might not be representative of the wider areas due to their typical locations in areas of low-complexity terrain. Hence, products derived from station data such as GPCC might underestimate areal precipitation, especially if there are only one or two stations within a grid cell, as is usually the case in HMA. Higher grid resolution, as in PRETIP, on the other hand, might also not improve precipitation estimates, as this satellite-based product is limited to the averaging within the random-forest methodology. We therefore suggest to not only rely on a single dataset in any application but to elaborate on the potential influences of different datasets in comparison. We suggest selecting a precipitation dataset based on one's application and requirements. For example, if data are needed for multi-decadal hydro-meteorological or hydro-climatological research applications, ERA5 is currently the best choice. When HAR v2 10 km becomes available for longer periods it will replace ERA5 in this position. If precipitation in complex terrain at high spatial resolution is to be investigated, HAR v2 2 km would be the optimally applicable dataset, which might still require bias correction for local applications. HAR v2 10 km and ERA5 might be employed over larger study areas or extended study periods. Similarly, glacio-hydrological studies, which usually expand over small areas, require high spatial resolution to accurately represent the prevailing accumulation patterns of the area. For studies focusing on the broader precipitation patterns under consideration of terrain complexity, most ERA products, the HAR products and JRA-55 have shown to be very similar. PRETIP offers a great opportunity for near-real time applications, such as flood forecasting, as the satellite data can be available within hours after the passage of the satellite, whereas reanalysis products are only available after several weeks.

Overall, in this study we elaborate and conclude on the following:

(1) How similar are the different gridded precipitation datasets? Depending on the origins and generation of the datasets, some datasets are very similar (e.g., HAR v2 2 km and HAR v2 10 km; ERA5 and ERA5-Land), while other datasets show larger discrepancies (e.g., Merra and GPCC). Despite some data gaps, the satellite product (PRETIP) falls within the range of cumulative precipitation and shows similar trends to other products. When comparing the grid values to station data, we conclude that spatial resolution plays a significant role and that gauge measurements likely

exhibit a dry bias due to their locations on valley floors or other areas of low terrain complexity. However, most products represent the timing and patterns of precipitation events well.

(2) What is the effect of terrain complexity on variations in precipitation between products? Terrain complexity increases the difference of precipitation between products. In complex terrain, the difference within daily precipitation can be up to $4\,\mathrm{mm\,d^{-1}}$, whereas it is generally below $2\,\mathrm{mm\,d^{-1}}$ in more homogeneous landscapes. Overall, the differences in precipitation derived from the analysis based on terrain complexity enables one to draw conclusions on how well some products work for studies focusing on complex terrain. For instance, it is possible to use the ERA5-Land dataset rather than the HAR v2 10 km dataset, if the latter is not available. Locally, the differences can still be large, but the overall precipitation estimates over a wider area are consistent between both datasets.

Author Contributions: Conceptualization, A.H., A.A. and C.S.; methodology, A.H.; software, A.H.; validation, C.S., D.S., J.B., B.T. and P.R.; formal analysis, A.H. and A.A.; investigation, A.H., A.A.; data curation, P.R., O.B., X.W., A.A. and C.K.; interpretation of results, D.S., A.H., C.K., X.W., A.A., J.B., C.S. and P.R.; writing—original draft preparation, A.H.; writing—review and editing, A.H., A.A., C.K., X.W., J.B., D.S., C.S., P.R. and O.B.; visualization, A.H.; supervision, C.S., D.S., J.B. and B.T.; project administration, C.S., D.S. and J.B.; funding acquisition, C.S., D.S., J.B. and P.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the German Research Foundation's (DFG) research grant "Precipitation patterns, snow and glacier response in High Asia and their variability on sub-decadal time scales." The program consisted of the subprojects "snow cover and glacier energy and mass balance variability" (prime-SG, SCHN680/13-1), sub-projects: "Remote Sensing of precipitation" (prime-RS, BE1780/46-1 and TH1531/6-1), and "prime-HYD—High Mountain Asian HYDrological variability" (prime-HYD, RE3834/4-1). We also acknowledge support from the German Federal Ministry of Education and Research (BMBF) within the project "Climatic and Tectonic Natural Hazards in Central Asia" (CaTeNA, FKZ 03G0878G).

Acknowledgments: We greatly thank the Chinese Ministry of Water Resources and Zhiyu Liu for sharing the Chinese rain-gauge observation data for the non-profit research study within our project. We would also like to thank the three anonymous reviewers and the editors, who substantially helped to improve the study. Further, we acknowledge support by the German Research Foundation (DFG) and the Open Access Publication Fund of Humboldt-Universität zu Berlin.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study, in the collection, analyses or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

List of Acronyms

ALOS	Advanced Land Observing Satellite.
DEM	Digital elevation model.
ECMWF	European Centre for Medium-Range Weather Forecasts.
ERA5	ERA5.
ERA5-Land	ERA5-Land.
ERA-Interim	ERA-Interim.
GPCC	Global Precipitation Climatology Centre.
HAR v2	High Asia Refined analysis version 2.
HAR v2 2 km	High Asia Refined analysis version 2–2 km domain.
HAR v2 10 km	High Asia Refined analysis version 2–10 km domain.
HMA	High Mountain Asia.
JRA-55	Japanese 55-year Reanalysis.
MERRA-2	Modern-Era Retrospective analysis for Research and Applications, Version 2.
PRETIP	Precipitation REtrieval covering the TIbetan Plateau.
TiP	Tibetan Plateau.

Appendix A

Table A1. Correlation coefficient **R** for all datasets in **mm/10days**. All correlations are statistically significant at the 99% confidence interval.

Dataset	ERA5	ERA-Interim	ERA5-Land	HAR v2 2 km	HAR v2 10 km	JRA55	MERRA2	PRETIP
ERA-Interim	0.84							
ERA5-Land	1.00	0.84						
HAR v2 2 km	0.88	0.83	0.85					
HAR v2 10 km	0.86	0.84	0.86	0.89				
JRA55	0.72	0.80	0.74	0.75	0.73			
MERRA2	0.76	0.73	0.78	0.74	0.75	0.67		
PRETIP	0.72	0.74	0.72	0.71	0.71	0.72	0.64	
GPCC	0.80	0.73	0.81	0.75	0.75	0.69	0.76	0.58

Table A2. Correlation coefficient **R** for all datasets in **mm/month**. All correlations are statistically significant at the 99% confidence interval.

Dataset	ERA5	ERA-Interim	ERA5-Land	HAR v2 2 km	HAR v2 10 km	JRA55	MERRA2	PRETIP
ERA-Interim	0.83							
ERA5-Land	1.00	0.83						
HAR v2 2 km	0.89	0.86	0.88					
HAR v2 10 km	0.87	0.87	0.87	0.92				
JRA55	0.71	0.86	0.73	0.78	0.76			
MERRA2	0.68	0.71	0.69	0.68	0.69	0.64		
PRETIP	0.71	0.75	0.74	0.76	0.72	0.66	0.51	
GPCC	0.74	0.73	0.74	0.74	0.76	0.67	0.68	0.41

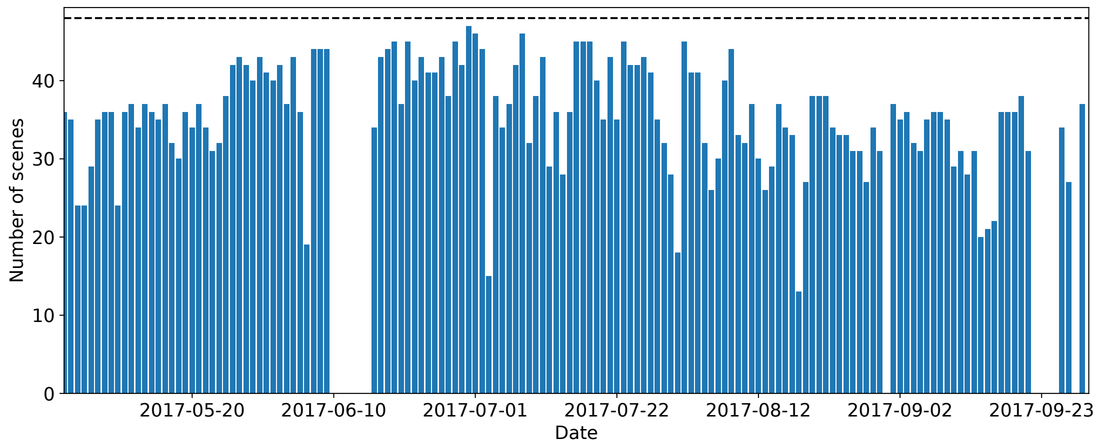

Figure A1. Amount of available PRETIP scenes per day. The maximum value is 48 (2 scenes per hour) and marked with the black dotted line. On average, 32.6 scenes per day are available.

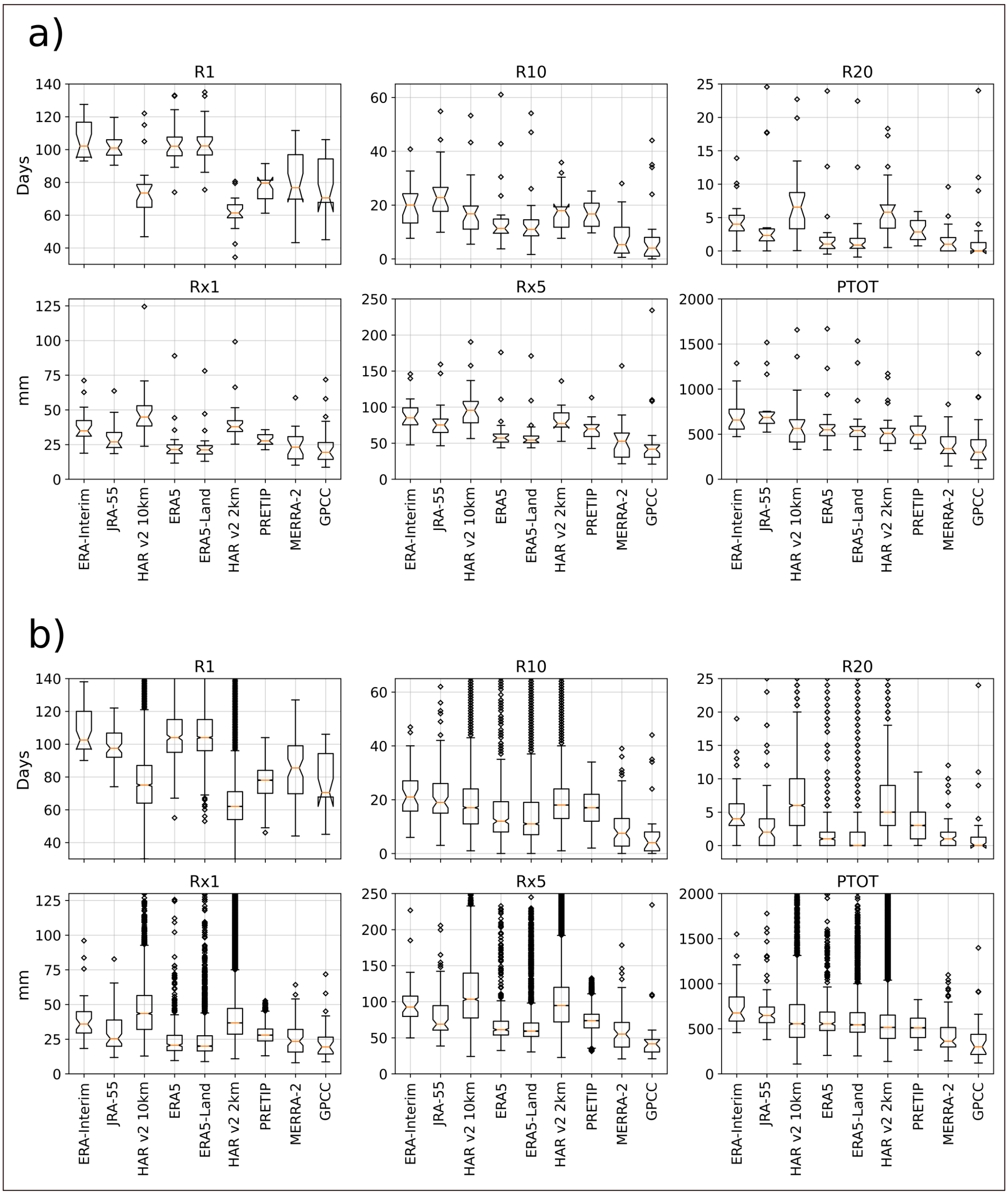

Figure A2. Visualization of the selected climdex indices R1, R10, R20, Rx1, Rx5 and PTOT as boxplot charts equivalent to Figure 8 (for description see Table 3). (**a**) depicts resulting values after resampling every product to the grid resolution of the lowest resolved product. (**b**) shows the same boxplot charts as Figure 8, but with the y-axis limits adjusted to the range in (a) to allow for direct comparison between both versions.

References

1. Immerzeel, W.W.; Lutz, A.F.; Andrade, M.; Bahl, A.; Biemans, H.; Bolch, T.; Hyde, S.; Brumby, S.; Davies, B.J.; Elmore, A.C.; et al. Importance and vulnerability of the world's water towers. *Nature* **2020**, *577*, 364–369. [CrossRef]
2. Immerzeel, W.W.; van Beek, L.P.H.; Bierkens, M.F.P. Climate Change Will Affect the Asian Water Towers. *Science* **2010**, *328*, 1382–1385. [CrossRef] [PubMed]
3. Wang, S.; Zhang, M.; Wang, B.; Sun, M.; Li, X. Recent changes in daily extremes of temperature and precipitation over the western Tibetan Plateau, 1973–2011. *Quat. Int.* **2013**, *313–314*, 110–117. [CrossRef]

4. Gao, Y.; Xiao, L.; Chen, D.; Xu, J.; Zhang, H. Comparison between past and future extreme precipitations simulated by global and regional climate models over the Tibetan Plateau. *Int. J. Climatol.* **2018**, *38*, 1285–1297. [CrossRef]
5. Shean, D.E.; Bhushan, S.; Montesano, P.; Rounce, D.R.; Arendt, A.; Osmanoglu, B. A Systematic, Regional Assessment of High Mountain Asia Glacier Mass Balance. *Front. Earth Sci.* **2020**, *7*, 363. [CrossRef]
6. Rounce, D.R.; Hock, R.; Shean, D.E. Glacier Mass Change in High Mountain Asia Through 2100 Using the Open-Source Python Glacier Evolution Model (PyGEM). *Front. Earth Sci.* **2020**, *7*, 331. [CrossRef]
7. Wang, A.; Zeng, X. Evaluation of multireanalysis products with in situ observations over the Tibetan Plateau. *J. Geophys. Res. Atmos.* **2012**, *117*. [CrossRef]
8. Allerup, P.; Madsen, H. Accuracy of Point Precipitation Measurements. *Hydrol. Res.* **1980**, *11*, 57–70, [CrossRef]
9. Villarini, G.; Mandapaka, P.V.; Krajewski, W.F.; Moore, R.J. Rainfall and sampling uncertainties: A rain gauge perspective. *J. Geophys. Res. Atmos.* **2008**. [CrossRef]
10. New, M.; Todd, M.; Hulme, M.; Jones, P. Precipitation measurements and trends in the twentieth century. *Int. J. Climatol.* **2001**. [CrossRef]
11. Sevruk, B. Adjustment of tipping-bucket precipitation gauge measurements. *Atmos. Res.* **1996**, *42*, 237–246. [CrossRef]
12. Immerzeel, W.W.; Wanders, N.; Lutz, A.F.; Shea, J.M.; Bierkens, M.F.P. Reconciling high altitude precipitation in the upper Indus Basin with glacier mass balances and runoff. *Hydrol. Earth Syst. Sci. Discuss.* **2015**, *12*, 4755–4784. [CrossRef]
13. Rasmussen, R.; Baker, B.; Kochendorfer, J.; Meyers, T.; Landolt, S.; Fischer, A.P.; Black, J.; Thériault, J.M.; Kucera, P.; Gochis, D.; et al. How Well Are We Measuring Snow: The NOAA/FAA/NCAR Winter Precipitation Test Bed. *Bull. Am. Meteorol. Soc.* **2012**, *93*, 811–829. [CrossRef]
14. Rollenbeck, R.; Bendix, J. Rainfall distribution in the Andes of southern Ecuador derived from blending weather radar data and meteorological field observations. *Atmos. Res.* **2011**, *99*, 277–289. [CrossRef]
15. Sevruk, B. Point precipitation measurements: Why are they not corrected? In *Water for the Future: Hydrology in Perspective*; IAHS Press: Wallingford, UK, 1987.
16. Rudolf, B.; Schneider, U. Calculation of gridded precipitation data for the global land-surface using in-situ gauge observations. In Proceedings of the Second Workshop of the International Precipitation Working Group 2014, Monterey, CA, USA, 17 February 2014; pp. 231–247.
17. Huffman, G.J.; Bolvin, D.T.; Nelkin, E.J.; Wolff, D.B.; Adler, R.F.; Gu, G.; Hong, Y.; Bowman, K.P.; Stocker, E.F. The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales. *J. Hydrometeorol.* **2007**, *8*, 38–55. [CrossRef]
18. Hou, A.Y.; Kakar, R.K.; Neeck, S.; Azarbarzin, A.A.; Kummerow, C.D.; Kojima, M.; Oki, R.; Nakamura, K.; Iguchi, T. The Global Precipitation Measurement Mission. *Bull. Am. Meteorol. Soc.* **2014**, *95*, 701–722. [CrossRef]
19. Huang, D.; Gao, S. Impact of different reanalysis data on WRF dynamical downscaling over China. *Atmos. Res.* **2018**, *200*, 25–35.
20. Li, D.; Yang, K.; Tang, W.; Li, X.; Zhou, X.; Guo, D. Characterizing precipitation in high altitudes of the western Tibetan plateau with a focus on major glacier areas. *Int. J. Climatol.* **2020**, joc.6509. [CrossRef]
21. Yoon, Y.; Kumar, S.V.; Forman, B.A.; Zaitchik, B.F.; Kwon, Y.; Qian, Y.; Rupper, S.; Maggioni, V.; Houser, P.; Kirschbaum, D.; et al. Evaluating the Uncertainty of Terrestrial Water Budget Components Over High Mountain Asia. *Front. Earth Sci.* **2019**, *7*, 120. [CrossRef]
22. Bai, L.; Wen, Y.; Shi, C.; Yang, Y.; Zhang, F.; Wu, J.; Gu, J.; Pan, Y.; Sun, S.; Meng, J. Which Precipitation Product Works Best in the Qinghai-Tibet Plateau, Multi-Source Blended Data, Global/Regional Reanalysis Data, or Satellite Retrieved Precipitation Data? *Remote Sens.* **2020**, *12*, 683. [CrossRef]
23. Jain, S.; Mishra, S.K.; Salunke, P.; Sahany, S. Importance of the resolution of surface topography vis-à-vis atmospheric and surface processes in the simulation of the climate of Himalaya-Tibet highland. *Clim. Dyn.* **2019**, *52*, 4735–4748. [CrossRef]
24. Chvíla, B.; Ondras, M.; Sevruk, B. The wind-induced loss of precipitation measurement of small time intervals as recorded in the field. In Proceedings of the WMO/CIMO Technical Conference, Bratislava, Slovakia, 23–25 September 2002.

25. Service, C.C.C. C3S ERA5-Land Reanalysis. 2019. Available online: https://cds.climate.copernicus.eu/cdsapp#!/home (accessed on 20 November 2020).
26. Dee, D.P.; Uppala, S.M.; Simmons, A.J.; Berrisford, P.; Poli, P.; Kobayashi, S.; Andrae, U.; Balmaseda, M.A.; Balsamo, G.; Bauer, P.; et al. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **2011**, *137*, 553–597. [CrossRef]
27. Kobayashi, S.; Ota, Y.; Harada, Y.; Ebita, A.; Moriya, M.; Onoda, H.; Onogi, K.; Kamahori, H.; Kobayashi, C.; Endo, H.; et al. The JRA-55 Reanalysis: General Specifications and Basic Characteristics. *J. Meteorol. Soc. Jpn. Ser. II* **2015**, *93*, 5–48. [CrossRef]
28. Gelaro, R.; McCarty, W.; Suárez, M.J.; Todling, R.; Molod, A.; Takacs, L.; Randles, C.A.; Darmenov, A.; Bosilovich, M.G.; Reichle, R.; et al. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *J. Clim.* **2017**, *30*, 5419–5454. [CrossRef] [PubMed]
29. Wang, X.; Tolksdorf, V.; Otto, M.; Scherer, D. WRF-based Dynamical Downscaling of ERA5 Reanalysis Data for High Mountain Asia: Towards a New Version of the High Asia Refined Analysis. *Int. J. Climatol.* **2020**. [CrossRef]
30. Kolbe, C.; Thies, B.; Egli, S.; Lehnert, L.; Schulz, H.; Bendix, J. Precipitation Retrieval over the Tibetan Plateau from the Geostationary Orbit—Part 1: Precipitation Area Delineation with Elektro-L2 and Insat-3D. *Remote Sens.* **2019**, *11*, 2302. [CrossRef]
31. Kolbe, C.; Thies, B.; Turini, N.; Liu, Z.; Bendix, J. Precipitation retrieval over the Tibetan Plateau from the geostationary Orbit—Part 2: Precipitation rates with Elektro-L2 and Insat-3D. *Remote Sens.* **2020**, *12*, 2114. [CrossRef]
32. Maussion, F.; Scherer, D.; Mölg, T.; Collier, E.; Curio, J.; Finkelnburg, R. Precipitation Seasonality and Variability over the Tibetan Plateau as Resolved by the High Asia Reanalysis. *J. Clim.* **2014**, *27*, 1910–1927. [CrossRef]
33. Pritchard, D.M.W.; Forsythe, N.; Fowler, H.J.; O'Donnell, G.M.; Li, X.F. Evaluation of Upper Indus Near-Surface Climate Representation by WRF in the High Asia Refined Analysis. *J. Hydrometeorol.* **2019**, *20*, 467–487. [CrossRef]
34. Kropáček, J.; Neckel, N.; Tyrna, B.; Holzer, N.; Hovden, A.; Gourmelen, N.; Schneider, C.; Buchroithner, M.; Hochschild, V. Repeated glacial lake outburst flood threatening the oldest Buddhist monastery in north-western Nepal. *Nat. Hazards Earth Syst. Sci.* **2015**, *15*, 2425–2437. [CrossRef]
35. Yatagai, A.; Kamiguchi, K.; Arakawa, O.; Hamada, A.; Yasutomi, N.; Kitoh, A. APHRODITE: Constructing a Long-Term Daily Gridded Precipitation Dataset for Asia Based on a Dense Network of Rain Gauges. *Bull. Am. Meteorol. Soc.* **2012**, *93*, 1401–1415, [CrossRef]
36. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, qj.3803. [CrossRef]
37. Schamm, K.; Ziese, M.; Becker, A.; Finger, P.; Meyer-Christoffer, A.; Schneider, U.; Schröder, M.; Stender, P. Global gridded precipitation over land: A description of the new GPCC First Guess Daily product. *Earth Syst. Sci. Data* **2014**, *6*, 49–60. [CrossRef]
38. Skamarock, W.; Klemp, J.; Dudhia, J.; Gill, D.; Zhiquan, L.; Berner, J.; Wang, W.; Powers, J.; Duda, M.G.; Barker, D.M.; et al. *A Description of the Advanced Research WRF Model Version 4 NCAR Technical Note*; Technical Report; NCAR-UCAR: Boulder, CO, USA, 2019. [CrossRef]
39. National Satellite Meteorological Centre. *Insat-3D Data Products Catalog*; Technical Report; 2014. Available online: http://satellite.imd.gov.in/dynamic/INSAT3D_Catalog.pdf (accessed on 20 November 2020).
40. Zak, A. Zenit delivers Elektro-L2. 2016. Available online: http://www.russianspaceweb.com/elektro-l2.html (accessed on 20 November 2020).
41. Huffman, G.J.; Bolvin, D.T.; Nelkin, E.J. Integrated Multi-satellitE Retrievals for GPM (IMERG) Technical Documentation. *NASA/GSFC Code* **2018**, *612*, 47. [CrossRef]
42. Tang, G.; Ma, Y.; Long, D.; Zhong, L.; Hong, Y. Evaluation of GPM Day-1 IMERG and TMPA Version-7 legacy products over Mainland China at multiple spatiotemporal scales. *J. Hydrol.* **2016**, *533*, 152–167. [CrossRef]
43. Kirschbaum, D.B.; Huffman, G.J.; Adler, R.F.; Braun, S.; Garrett, K.; Jones, E.; McNally, A.; Skofronick-Jackson, G.; Stocker, E.; Wu, H.; et al. NASA's Remotely Sensed Precipitation: A Reservoir for Applications Users. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 1169–1184, [CrossRef]
44. Uppala, S.M.; Kållberg, P.; Simmons, A.; Andrae, U.; Bechtold, V.D.C.; Fiorino, M.; Gibson, J.; Haseler, J.; Hernandez, A.; Kelly, G.; et al. The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.* **2005**, *131*, 2961–3012.

45. Zhang, X.; Alexander, L.; Hegerl, G.C.; Jones, P.; Tank, A.K.; Peterson, T.C.; Trewin, B.; Zwiers, F.W. Indices for monitoring changes in extremes based on daily temperature and precipitation data: Indices for monitoring changes in extremes. *Wiley Interdiscip. Rev. Clim. Chang.* **2011**, *2*, 851–870. [CrossRef]
46. Lu, H.; Liu, X.; Bian, L. Terrain complexity: definition, index, and DEM resolution. In *Geoinformatics 2007: Geospatial Information Science*; Chen, J., Pu, Y., Eds.; International Society for Optics and Photonics, SPIE: Nanjing, China, 2007; Volume 6753, pp. 753–763. [CrossRef]
47. Miehe, G.; Winiger, M.; Böhner, J.; Yili, Z. The Climatic Diagram Map of High Asia: Purpose and Concepts (Klimadiagramm-Karte von Hochasien. Konzept und Anwendung). *Erdkunde* **2001**, *55*, 94–97.
48. Pham, Q.B.; Yang, T.C.; Kuo, C.M.; Tseng, H.W.; Yu, P.S. Combing random forest and least square support vector regression for improving extreme rainfall downscaling. *Water* **2019**, *11*. [CrossRef]
49. Nashwan, M.S.; Shahid, S. Symmetrical uncertainty and random forest for the evaluation of gridded precipitation and temperature data. *Atmos. Res.* **2019**, *230*, 104632. [CrossRef]
50. Zhang, Y.; Li, J. Impact of moisture divergence on systematic errors in precipitation around the Tibetan Plateau in a general circulation model. *Clim. Dyn.* **2016**, *47*, 2923–2934. [CrossRef]
51. Ou, T.; Chen, D.; Chen, X.; Lin, C.; Yang, K.; Lai, H.W.; Zhang, F. Simulation of summer precipitation diurnal cycles over the Tibetan Plateau at the gray-zone grid spacing for cumulus parameterization. *Clim. Dyn.* **2020**, *54*, 1–15.

Paper III

Atmospheric triggering conditions and climatic disposition of landslides in Kyrgyzstan and Tajikistan at the beginning of the 21st century

Wang, X., Otto, M. and Scherer, D., 2021: Atmospheric triggering conditions and climatic disposition of landslides in Kyrgyzstan and Tajikistan at the beginning of the 21st century. *Nat. Hazards Earth Syst. Sci.*, 21, 2125–2144, https://doi.org/10.5194/nhess-21-2125-2021.

Status: Published.

Own contribution:

– design of the study (with co-authors)

– numerical simulations

– data preparation

– data analysis and visualization

– writing and revision of manuscript

Nat. Hazards Earth Syst. Sci., 21, 2125–2144, 2021
https://doi.org/10.5194/nhess-21-2125-2021

Natural Hazards and Earth System Sciences
Open Access

Atmospheric triggering conditions and climatic disposition of landslides in Kyrgyzstan and Tajikistan at the beginning of the 21st century

Xun Wang, Marco Otto, and Dieter Scherer

Chair of Climatology, Technische Universität Berlin, Berlin, 12165, Germany

Correspondence: Xun Wang (xun.wang@tu-berlin.de)

Received: 18 December 2020 – Discussion started: 6 January 2021
Revised: 11 June 2021 – Accepted: 11 June 2021 – Published: 13 July 2021

Abstract. Landslide is a major natural hazard in Kyrgyzstan and Tajikistan. Knowledge about atmospheric triggering conditions and climatic disposition of landslides in Kyrgyzstan and Tajikistan is limited even though this topic has already been investigated thoroughly in other parts of the world. In this study, the newly developed, high-resolution High Asia Refined analysis version 2 (HAR v2) data set generated by dynamical downscaling was combined with historical landslide inventories to analyze the atmospheric conditions that initialized landslides in Kyrgyzstan and Tajikistan. The results indicate the crucial role of snowmelt in landslide-triggering processes since it contributes to the initialization of 40 % of landslide events. Objective thresholds for rainfall, snowmelt, and the sum of rainfall and snowmelt (rainfall + snowmelt) were defined. Thresholds defined by rainfall + snowmelt have the best predictive performance. Mean intensity, peak intensity, and the accumulated amount of rainfall + snowmelt events show similar predictive performance. Using the entire period of rainfall + snowmelt events results in better predictive performance than just considering the period up to landslide occurrence. Mean annual exceedance maps were derived from defined regional thresholds for rainfall + snowmelt. Mean annual exceedance maps depict climatic disposition and have added value in landslide susceptibility mapping. The results reported in this study highlight the potential of dynamical downscaling products generated by regional climate models in landslide prediction.

1 Introduction

Landslide is one of the most severe natural hazards in Kyrgyzstan and Tajikistan. More than 300 big landslides occurred in Kyrgyzstan from 1993 to 2010, causing 256 fatalities and direct economic losses of USD 2.5 million per year (Torgoev et al., 2012). Under global warming, wildfires, glacial retreat, and permafrost degradation are much more likely to enhance slope instabilities in mountainous areas (Froude and Petley, 2018; Palmer, 2020), making these regions, including Kyrgyzstan and Tajikistan, more vulnerable to climate change. The occurrence of landslides depends on disposition and triggering events. Disposition refers to the general settings that make slopes prone to failure without actually initiating it, such as slope gradient and aspect, geology, vegetation cover, climate, etc. (Dai et al., 2002). Common triggers for landslides are extreme and prolonged rainfall, rapid snowmelt, and earthquakes (Wieczorek, 1996).

The majority of landslide research in Kyrgyzstan and Tajikistan focused on characterizing landslide susceptibility, i.e., “where” landslides are prone to occur (e.g., Braun et al., 2015; Saponaro et al., 2015; Havenith et al., 2015b), and how to improve the landslide susceptibility models (Ozturk et al., 2020; Barbosa et al., 2021). But little attention is paid to the atmospheric triggering conditions, and our knowledge of “when” landslides are likely to occur is limited in this region. In addition, most landslide susceptibility studies only took non-climatic factors into account or simply applied annual precipitation as a climatic factor. According to Segoni et al. (2018), no rainfall threshold for landslide triggering has been defined for Kyrgyzstan and Tajikistan yet even though this topic has already been thoroughly investigated in other parts of the world with high landslide susceptibility (e.g., Berti et al., 2012; Gariano et al., 2015; Giannecchini et al., 2016; Leonarduzzi et al., 2017). The reasons are twofold. Firstly, although landslide inventories have been developed in this region, e.g., the Tien Shan Geohazards Database (Havenith et al., 2015a, b) and the multi-temporal landslide inventory from Behling and Roessner (2020), there is a lack of landslide inventories with the exact date of landslide occurrence. Given the highly dynamic nature of weather phenomena, at least a daily timestamp of landslide records is required to investigate weather conditions that trigger landslides. Secondly, there is a lack of atmospheric data. The number of in situ observation stations in Kyrgyzstan and Tajikistan decreased sharply in the 1990s due to reduced funding. There are currently eight stations in Kyrgyzstan and 26 stations in Tajikistan available from Global Surface Summary of the Day (GSOD), which is a publicly available data set. These numbers are already significantly below the recommendation of the World Meteorological Organization even for flat areas (Ilyasov et al., 2013). Despite the sparse distribution, most GSOD stations are located in low-lying valleys and are not fully representative of the area.

Rainfall is the most common trigger of landslides all over the world (Wieczorek, 1996). Over snow-covered regions, snowmelt is recognized as another common trigger of shallow landslides and debris flows (Wieczorek, 1996; Mostbauer et al., 2018). In Kyrgyzstan and Tajikistan, more than half of the annual precipitation falls in the form of snow. Snow cover duration over high mountain ranges in the Tien Shan and the Pamir is more than $200\,\mathrm{d\,yr^{-1}}$ (Dietz et al., 2014). A large amount of water stored in snowpacks is released during the melting season. Snowmelt is another important source of water infiltrating into the soil that increases slope instability. Thus, in Kyrgyzstan and Tajikistan, snowmelt might also play a role in landslide triggering besides rainfall. But snowmelt is not as easy to be observed as rainfall and might often be neglected as a landslide trigger, especially when co-occurring with rainfall.

There are two main approaches to assess rainfall thresholds for landslide triggering. The first approach is physically based and requires detailed lithological, morphological, and geotechnical information of each landslide event (Guzzetti et al., 2007). Unfortunately, this level of detail is usually restricted to small areas and is not available for the whole of Kyrgyzstan and Tajikistan. The second one is the empirical approach based on historical landslide and rainfall data. The majority of studies applying this approach relied on rain gauge data to analyze rainfall thresholds (e.g., Berti et al., 2012; Khan et al., 2012; Bui et al., 2013). However, rain gauge data are point measurements that cannot capture the large spatial heterogeneity of rainfall, especially over complex terrain. Gridded products can provide continuous data in both space and time and can be used in detecting atmospheric triggering conditions of landslides.

We aim to analyze the atmospheric triggering conditions of landslides and generate climatic disposition maps that contain information on these triggering conditions in Kyrgyzstan and Tajikistan. For this purpose, we combined freely available gridded atmospheric data with historical landslide events. Atmospheric triggers for each landslide event were determined by the co-occurrence of landslide and weather events. Properties (mean intensity, peak intensity, accumulated amount) of landslide-triggering events (LTEs) and non-landslide-triggering events (NLTEs) were compared. Objective thresholds of these properties for different atmospheric triggers (rainfall, snowmelt, and the sum of rainfall and snowmelt) were defined so that they can best separate the atmospheric conditions that resulted and did not result in landslides. Finally, we applied the thresholds with the best predictive performance to generate maps of mean annual exceedance. In this way, we can transform the weather-scale triggering conditions into climate-scale dispositions (hereafter referred to as "climatic disposition").

The objective of this study is threefold: (1) investigate the role of snowmelt in landslide-triggering processes; (2) find appropriate quantities of atmospheric triggers for assessing landslide hazards; and (3) characterize climatic disposition in terms of rainfall and snowmelt over Kyrgyzstan and Tajikistan.

The paper is organized as follows: we describe the data and methods used in this study in the following section. Results are presented in Sect. 3 and discussed in Sect. 4. Conclusions are drawn in Sect. 5.

2 Data and method

2.1 Data

2.1.1 Landslide catalog

Landslide events used in this study come from two sources: the Global Landslide Catalog (GLC) (Kirschbaum et al., 2010, 2015) and the Global Fatal Landslide Database (GFLD) (Froude and Petley, 2018). The GLC has been compiled by NASA since 2007 and contains all types of mass movements triggered mostly by rainfall. The sources of the GLC are mainly media reports, disaster databases, and scientific reports. The GFLD only includes landslide events that caused fatalities and is obtained from media reports. It currently covers the period from 2004 to 2017. These two landslide inventories were chosen because, to the best of our knowledge, they are the only ones with the exact landslide dates available for the study region.

We selected landslide events triggered by atmospheric factors in Kyrgyzstan and Tajikistan from 2007 to 2018 from the GLC and 2004 to 2017 from the GFLD. Then we merged these two data sets and deleted duplicate events that occurred on the same day and came from the same source link, re-

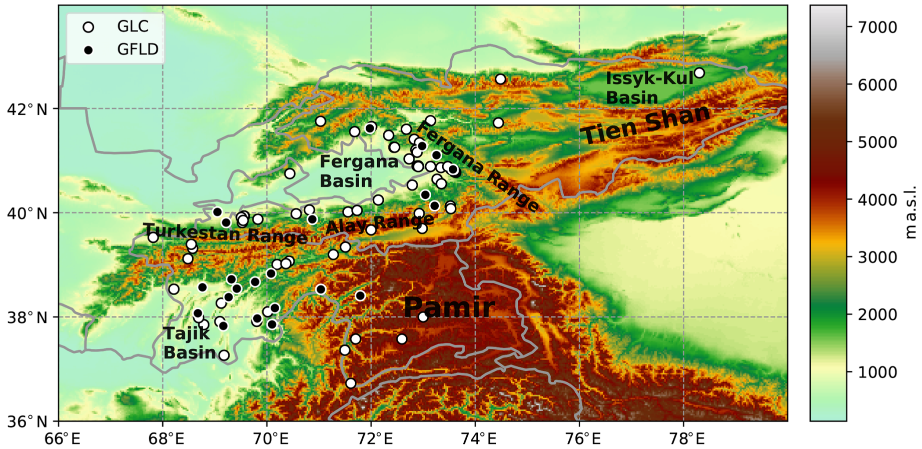

Figure 1. Landslide events from 2004 to 2018 extracted from the GLC (white points) and the GFLD (black points). Background contour is topography from digital elevation model (DEM) data from Shuttle Radar Topographic Mission (SRTM).

sulting in 96 landslide events for Kyrgyzstan and Tajikistan from 2004 to 2018 (Fig. 1).

2.1.2 Atmospheric data

Rainfall and snowmelt data are extracted from the High Asia Refined analysis version 2 (HAR v2). The HAR v2 is a newly developed regional atmospheric data set. It was generated by dynamical downscaling of the ERA5 reanalysis data using the Weather Research and Forecasting (WRF) model. It provides atmospheric data with high resolution and accuracy over High Mountain Asia (Hamm et al., 2020; Wang et al., 2021). Detailed modeling strategies of the HAR v2 are described in Wang et al. (2021). The HAR v2 has a grid spacing of 10 km and is available in hourly, daily, monthly, and yearly aggregations. Daily products were used in this study to determine the climatic trigger of each landslide event (Sect. 2.2.1) and to define thresholds for landslide triggering (Sect. 2.2.2). Rainfall was calculated as the difference between total precipitation and snowfall. Snowmelt is not a standard output of the WRF and was calculated using the surface energy balance (SEB). The SEB in the HAR v2 is resolved by the Noah land surface model (LSM) (Tewari et al., 2004):

$$H_\mathrm{m} = R_\mathrm{n} - H_\mathrm{s} - H_\mathrm{l} - H_\mathrm{g}, \tag{1}$$

where R_n, H_s, H_l, and H_g are net radiation, sensible heat flux, latent heat flux, and ground heat flux (in $\mathrm{W\,m^{-2}}$), respectively. These four variables are directly available in the HAR v2. H_m is the heat flux for melting and refreezing (in $\mathrm{W\,m^{-2}}$). $H_\mathrm{m} > 0$ indicates melting process, while $H_\mathrm{m} < 0$ refers to refreezing process. When $H_\mathrm{m} > 0$, snowmelt h_m ($\mathrm{kg\,m^{-2}\,s^{-1}}$) is calculated as

$$h_\mathrm{m} = \frac{H_\mathrm{m}}{\lambda_\mathrm{m}}, \tag{2}$$

where λ_m is the latent heat of fusion. When the calculated h_m is greater than snow water equivalent, then h_m is set to be equal to snow water equivalent.

2.2 Methods

2.2.1 Determine the atmospheric trigger of landslide events

The atmospheric trigger of a landslide event is determined by the co-occurrence of the landslide event with rainfall and snowmelt event. If a landslide event only occurred within or 1 d after a rainfall (snowmelt) event, then this landslide event is defined as rainfall (snowmelt) triggered. If there are both a rainfall event and a snowmelt event on the day or 1 d before the landslide occurrence day, then the atmospheric trigger of this landslide event is mixed.

To define a rainfall (snowmelt) event, the daily time series of rainfall (snowmelt) were extracted from the grid cells where landslides occurred. For each time series, an independent rainfall (snowmelt) event is defined as a series of consecutive days in which more than $0.2\,\mathrm{mm\,d^{-1}}$ of rainfall (snowmelt) is simulated. The value of $0.2\,\mathrm{mm\,d^{-1}}$ is chosen because it is the traditional precision of daily precipitation measurement (Jarraud, 2008) and can be applied to separate dry and wet conditions (Rodwell et al., 2010).

2.2.2 Threshold model for atmospheric triggers

The threshold model developed in this study contains three steps: (1) define LTEs and NLTEs; (2) define the thresholds for rainfall, snowmelt, and the sum of rainfall and snowmelt (hereafter referred to as rainfall + snowmelt) based on maximizing the predictive performance using 2 × 2 contingency tables; and (3) validate and assess the uncertainties of the defined thresholds. The methods for the first two steps were adopted from Leonarduzzi et al. (2017). Only the landslide events for which the atmospheric triggers could be determined were used for threshold modeling.

The first step is to define LTEs and NLTEs for rainfall, snowmelt, and rainfall + snowmelt. Here, we take rainfall as an example to describe the procedure. First, the method used in Sect. 2.2.1 is applied to define rainfall events for each time series extracted from grid cells where landslides occurred. Next, if a landslide event occurred during or 1 d after a rainfall event, then this rainfall event is classified as a landslide-triggering event. Given the uncertainty in timestamps of landslide events, the day after is also considered as a temporal relaxation. Otherwise, if a rainfall event is not associated with any landslide events, it is classified as a non-landslide-triggering event. For each rainfall event, we calculated three event properties: mean intensity I_{mean}, maximum intensity I_{max}, and the accumulated amount of rainfall for the entire event Q. For triggering events, we also calculated these three properties by only considering the period up to the day of the landslide occurrence (hereafter referred to as UTL, meaning up-to-landslide). Note that not all the landslide events co-occurred with a rainfall event. For these events, we set I_{mean}, I_{max}, and Q to zero. The same procedure for defining LTEs and NLTEs was conducted for snowmelt and rainfall + snowmelt as well.

The second step is to define thresholds of rainfall, snowmelt, and rainfall + snowmelt for entire events and UTL events using I_{mean}, I_{max}, and Q. No single threshold can perfectly separate LTEs from NLTEs since their distributions overlap. We applied 2×2 contingency tables to select the threshold that yields the best predictive performance. Using a certain threshold as a binary classifier, LTEs and NLTEs were categorized into true positive (TP), true negative (TN), false positive (FP), and false negative (FN). The Peirce skill score (PSS) (Hanssen and Kuipers, 1965) was applied as the measure of the predictive performance because it is trail-independent, which means it is unbiased even when the numbers of LTEs and NLTEs are not equally presented (Woodcock, 1976). The PSS is also known as the Hanssen–Kuiper skill score and the true skill statistic. It is calculated as the difference between hit rate (HR) and false alarm rate (FAR):

$$\mathrm{PSS} = \mathrm{HR} - \mathrm{FAR}, \tag{3}$$

$$\mathrm{HR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{4}$$

$$\mathrm{FAR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}. \tag{5}$$

We chose the threshold that maximizes the PSS. We also computed the Euclidean distance (d) to the optimal point ($\mathrm{HR} = 1$, $\mathrm{FAR} = 0$), which is another commonly used skill score in this application (e.g., Gariano et al., 2015; Piciullo et al., 2017; Postance et al., 2018; Zhuo et al., 2019). Additionally, the receiver operating characteristic (ROC) curve was used to determine the general predictive power of a certain predictor by calculating the area under the ROC curve (AUC) (Fawcett, 2006).

The last step is to validate the threshold model and assess uncertainty. For the calibration of thresholds, all landslide event samples were utilized, and corresponding statistic measures were calculated; i.e., the threshold model was trained and tested on the same data set. To test the model's predictive ability on an unseen data set, we performed k-fold cross-validation. Landslide events were randomly split into k folds with $k = 8$. Then for each unique fold, the fold was taken as the testing set, and the remaining $k - 1$ folds were taken as the training set. Mean values of thresholds, the corresponding statistic measures, and their uncertainties represented by standard deviations were reported.

2.2.3 Mean annual exceedance

Mean annual exceedance ($\overline{N}_{\mathrm{th}}$) is calculated for each HAR v2 grid cell. It is defined as the number of events that exceed a certain threshold over a certain period (N_{th}) divided by the total number of years (N_{a}):

$$\overline{N}_{\mathrm{th}} = \frac{N_{\mathrm{th}}}{N_{\mathrm{a}}}. \tag{6}$$

The unit of $\overline{N}_{\mathrm{th}}$ is the number of events per year. Mean annual exceedance transforms weather-scale triggering conditions to climate-scale disposition. It depicts where landslides are likely to occur from the climatic aspect.

3 Results

3.1 The role of snowmelt in landslide triggering

Figure 2 shows the climatology of seasonal rainfall, snowmelt, and rainfall + snowmelt resolved by the HAR v2. We define seasons as commonly done in meteorology, spanning 3 months each: winter (December–February, DJF), spring (March–May, MAM), summer (June–August, JJA), and autumn (September–November, SON). A high amount of rainfall concentrates in the western foothill of the Fergana Range, the northern foothill of the Turkestan Range, and the Tajik Basin in spring and shifts northeastwards into the Tien Shan in summer. Snowmelt occurs in spring over most high elevated areas. In summer, while most regions are snowmelt-free, the Pamir plateau still experiences a high amount of continuous snowmelt, which is in line with the results of Dietz et al. (2014) using remote sensing data.

Atmospheric triggers for each landslide event are determined using the method described in Sect. 2.2.1, and the results are shown in Fig. 3. Table A1 lists all 96 events and the climatic triggers detected by the HAR v2. Figure A1 shows the temporal process of rainfall and snowmelt for selected landslide cases. Nine landslide events did not occur within any rainfall event, snowmelt event, or rainfall + snowmelt

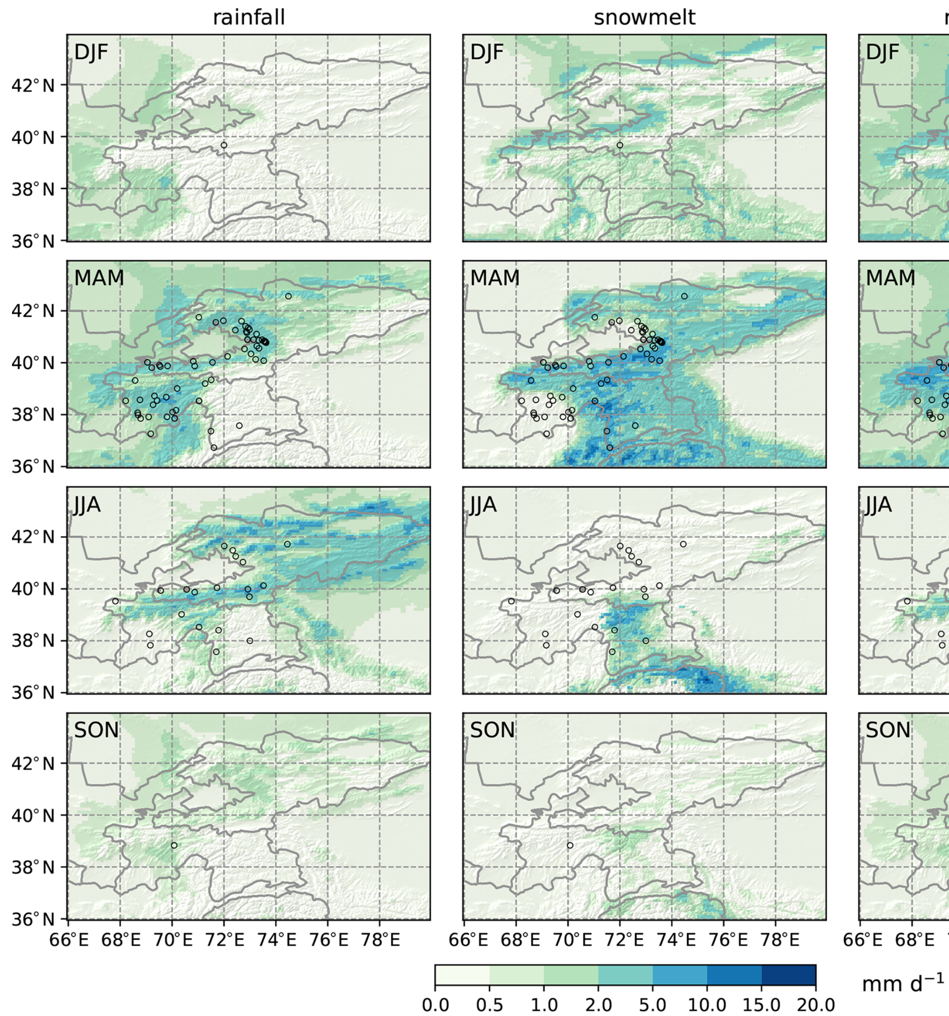

Figure 2. Seasonal rainfall, snowmelt, and rainfall + snowmelt from the HAR v2 from 2004 to 2018. Black circles: seasonal landslide events from the GLC and the GFLD. Topographic shading is based on DEM data from SRTM. DJF: December–February; MAM: March–May; JJA: June–August; and SON: September–November.

event. This mismatch between landslide information and weather information stems from the uncertainties in landslide locations and timing, as well as the uncertainties from rainfall and snowmelt simulated in the HAR v2 (detailed discussion in Sect. 4.1). These nine events are referred to as "not detected" (white points in Fig. 3) and are excluded. The remaining 87 landslide events were used for further analysis. Landslide events that were only triggered by rainfall mainly cluster in Tajik Basin and the northeastern rim of the Fergana Basin, where the contribution of rainfall to the annual sum of rainfall and snowmelt is high (Fig. 3).

The annual cycles of rainfall, snowmelt, and rainfall + snowmelt are compared with monthly landslide occurrences in Fig. 4. The study region experiences a peak of landslide activity in April and May, which corresponds with the peak of rainfall + snowmelt. While rainfall is the dominant trigger of landslides, snowmelt contributes to the triggering of 40 % of landslide events (35 out of 87). A total of 29 % of landslide events (25 out of 87) are attributed to the combined effect of rainfall and snowmelt. Most snowmelt-contributing events occurred in April when snowmelt amount is the highest. March and June have almost the same amount of rainfall + snowmelt. However, there are more landslide occurrences in June. This could be the result of soil still being frozen in March which stabilizes the slope. As shown in Fig. 4a, both soil temperature at the top soil layer (0–0.1 m) and air temperature at 2 m are still below zero in March.

3.2 Thresholds of atmospheric triggers for landslides in Kyrgyzstan and Tajikistan

Statistics of different properties of LTEs and NLTEs for rainfall, snowmelt, and rainfall + snowmelt are presented in Fig. 5 in the form of empirical cumulative distribution function (eCDF). Rainfall and snowmelt have a high percentage

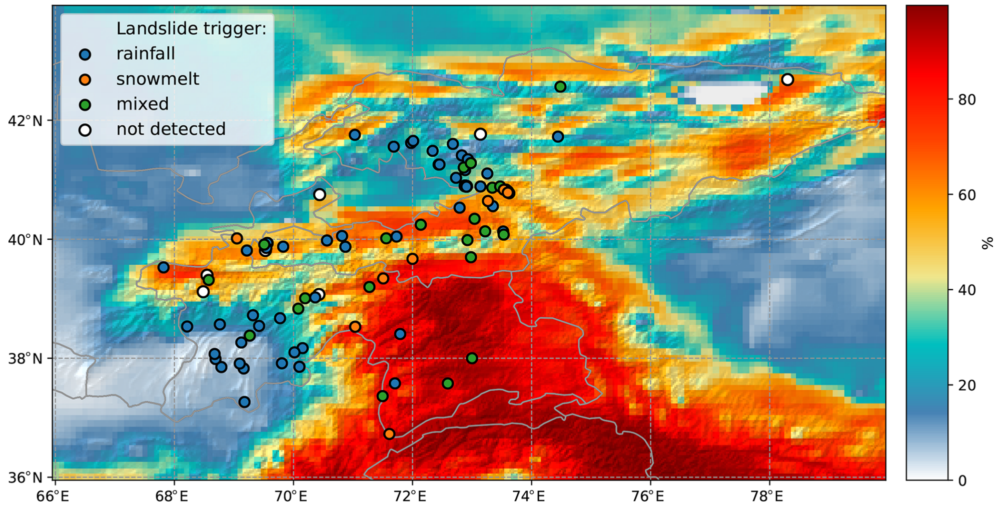

Figure 3. Contribution (%) of snowmelt to the annual sum of rainfall and snowmelt (background contour) and atmospheric triggers of 96 landslide events extracted from the GLC and the GFLD (points). Topographic shading is based on DEM data from SRTM.

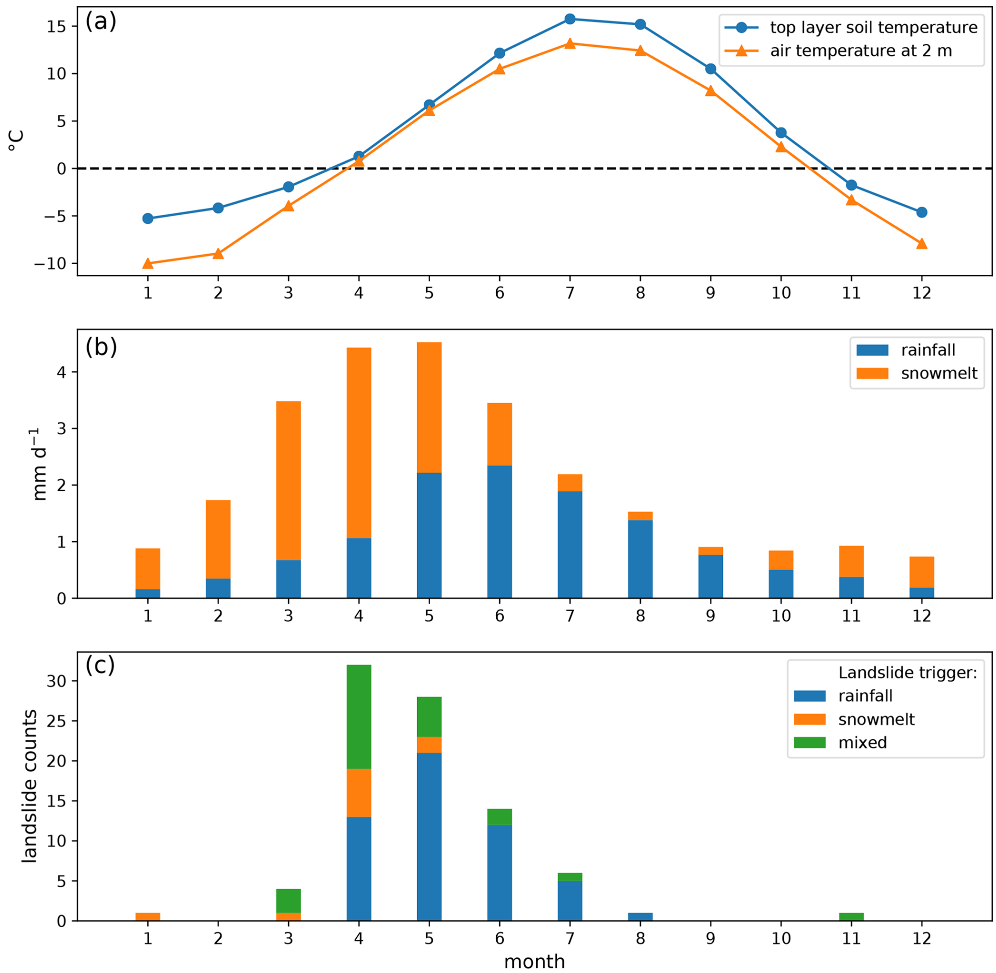

Figure 4. (a) Mean monthly soil temperature at the top soil layer (0–0.1 m) and air temperature at 2 m averaged over Kyrgyzstan and Tajikistan extracted from the HAR v2; **(b)** mean monthly rainfall and snowmelt averaged over Kyrgyzstan and Tajikistan extracted from the HAR v2; and **(c)** mean monthly landslide occurrences in Kyrgyzstan and Tajikistan from 2004 to 2018.

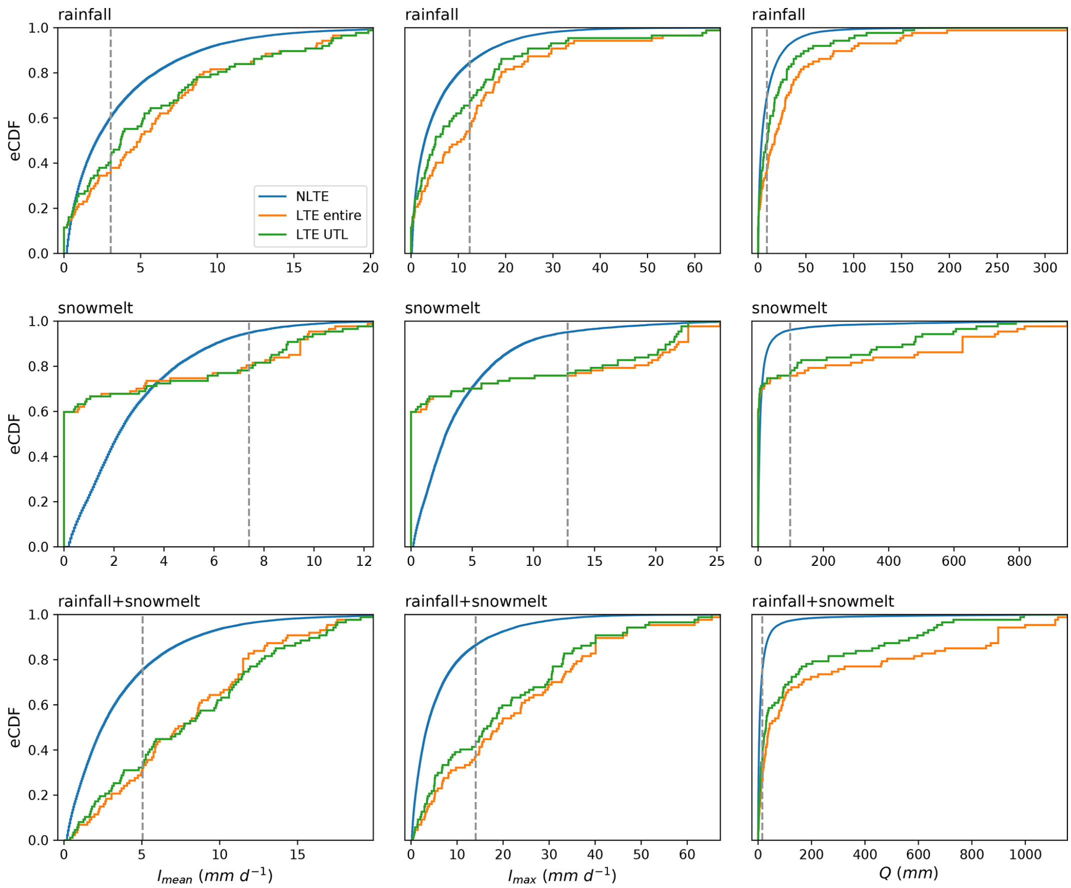

Figure 5. The eCDF curves of I_{mean}, I_{max}, Q of NLTE, landslide-triggering entire event (LTE entire), and landslide-triggering up-to-landslide event (LTE UTL) for rainfall, snowmelt, and rainfall + snowmelt during the period of 2004–2018. Dashed grey lines represent the thresholds for UTL events defined in Table 1.

of events with $I_{\mathrm{mean}} = 0$, $I_{\mathrm{max}} = 0$, and $Q = 0$. This is because, for landslide events that cannot be detected by only rainfall (orange points in Fig. 3), I_{mean}, I_{max}, and Q of rainfall for these events were all set to zero. The same procedure was conducted for events that cannot be detected by only snowmelt (blue points in Fig. 3). It can be seen in Fig. 5 that LTEs for both entire events and UTL events have stronger I_{mean} and I_{max}, as well as larger Q, compared to NLTEs. Moreover, snowmelt events have much higher Q but lower I_{mean} and I_{max} than rainfall events, indicating that snowmelt events are in general prolonged and not as intense as rainfall events. Overall, the HAR v2 combined with landslide inventories from the GLC and the GFLD can distinguish LTEs from NLTEs well and has potential in landslide threshold modeling.

We calibrated thresholds of I_{mean}, I_{max}, and Q using rainfall, snowmelt, and rainfall + snowmelt as predictors. The procedure was conducted for both entire events and UTL events. Predictive performance is better when using the entire period than just using the UTL period (Table 1), which was also concluded by Leonarduzzi et al. (2017). One of the reasons is that by considering a longer period, I_{mean}, I_{max}, and especially Q of LTEs generally increase, making it easier to distinguish LTEs from NLTEs. This can also be seen from the eCDFs in Fig. 5. In the eCDF space, the threshold defined by maximizing PSS is the point on the x axis where the vertical distance between the LTE curve and the NLTE curve is the largest. The eCDFs of UTL events are closer to the NLTE curve than eCDFs of the entire events. Therefore, the maximum PSSs of UTL events are smaller (Fig. 5).The better performance by considering the entire period could also indicate that there exists some uncertainty of landslide timing reported in the GLC and the GFLD. It can be seen from Table 1 that rainfall + snowmelt has the best predictive performance for both entire events and UTL events. The predictive performance indicated by d, PSS, and AUC of the three event properties (I_{mean}, I_{max}, and Q) are quite similar, but using I_{max} as a predictor leads to a lower FAR but also a lower HR when compared with Q and I_{mean}.

K-fold cross-validation results for entire events and UTL events are presented in Tables A2 and A3. Cross-validation reduces the sample size and makes the results more sensitive to outliers. The validation results are in line with the conclusions drawn by calibration: (1) among all predictors, rain-

Table 1. Calibrated thresholds of mean intensity I_{mean} (mm d^{-1}), maximum intensity I_{max} (mm d^{-1}), and accumulated amount Q (mm) for entire events and UTL events of rainfall, snowmelt, and the sum of rainfall and snowmelt (rainfall + snowmelt), as well as corresponding performance statistics.

Predictor	Property	Threshold	HR	FAR	d	PSS	AUC
Rainfall (entire event)	I_{mean}	3.60	0.62	0.35	0.51	0.27	0.62
	I_{max}	11.20	0.49	0.18	0.54	0.32	0.65
	Q	16.95	0.52	0.18	0.52	0.34	0.67
Snowmelt (entire event)	I_{mean}	7.05	0.23	0.06	0.77	0.17	0.31
	I_{max}	13.45	0.24	0.04	0.76	0.20	0.32
	Q	119.60	0.24	0.03	0.76	0.21	0.33
Rainfall + snowmelt (entire event)	I_{mean}	4.95	0.71	0.25	0.38	0.46	0.78
	I_{max}	12.80	0.67	0.15	0.37	0.51	0.81
	Q	17.15	0.74	0.23	0.35	0.50	0.81
Rainfall (UTL event)	I_{mean}	3.05	0.60	0.40	0.57	0.20	0.59
	I_{max}	12.40	0.34	0.16	0.67	0.19	0.58
	Q	9.25	0.52	0.31	0.57	0.21	0.59
Snowmelt (UTL event)	I_{mean}	7.40	0.22	0.05	0.78	0.17	0.31
	I_{max}	12.80	0.24	0.05	0.76	0.19	0.32
	Q	98.30	0.24	0.04	0.76	0.20	0.32
Rainfall + snowmelt (UTL event)	I_{mean}	5.05	0.68	0.25	0.41	0.43	0.76
	I_{max}	14.05	0.59	0.14	0.44	0.45	0.77
	Q	15.65	0.66	0.25	0.43	0.40	0.76

fall + snowmelt has the best predictive performance for both entire events and UTL events; (2) predictive performance is better when using the entire period than just using the UTL period; and (3) predictive performance of I_{mean}, I_{max}, and Q for rainfall + snowmelt are quite similar, but I_{max} has a lower FAR and also a lower HR.

3.3 Mean annual exceedance

Using the thresholds defined in Sect. 3.2 for rainfall + snowmelt UTL events, Fig. 6 presents the annual number of rainfall + snowmelt events that exceed the thresholds of I_{mean} =5.05 mm d^{-1}, I_{max} =14.05 mm d^{-1}, and Q = 15.65 mm (hereafter referred to as $I_{mean,th}$, $I_{max,th}$, and Q_{th}). Here, only the results for UTL events are presented since the defined thresholds of entire events and UTL events for rainfall + snowmelt are very similar and only deviate within 10 %, although their predictive performance is different (Table 1).

Locations with higher mean annual exceedance over $I_{max,th}$ indicate a higher chance of having rainfall + snowmelt events with high intensity, such as the Fergana Range and the northeastern Tajik Basin. These two regions have a high contribution of rainfall to annual rainfall + snowmelt (Fig. 3), and rainfall events tend to have stronger intensity than snowmelt events (Fig. 5). Locations with high mean annual exceedance over Q_{th} but low exceedance over $I_{max,th}$, including the Pamir Plateau and the Tien Shan, indicate that prolonged events instead of short and intense events are more frequent. The mean annual exceedance maps of Q_{th} and $I_{mean,th}$ correspond better with the landslide occurrences since they encompass both extreme events and prolonged events. Landslide events reported from the GLC and the GFLD are generally located in areas with high exceedance over Q_{th} and $I_{mean,th}$. However, the mean annual exceedance maps of Q_{th} and $I_{mean,th}$ also have more areas with false alarms, i.e., areas with high mean annual exceedance but no landslide occurrence. In remote areas, such as the Tien Shan, high false alarms could be due to the fact that landslides extracted from media reports are generally underreported in remote regions. This is discussed in detail in Sect. 4.1. In contrast, the mean annual exceedance map of $I_{max,th}$ misses more landslide events but has less false alarm area when compared to the exceedance maps of Q_{th} and $I_{mean,th}$.

4 Discussion

4.1 Sources of uncertainty

The uncertainty of the results depends on the accuracy of the data and the method applied to analyze the data. Our approach is purely empirical-based, which allows us to investigate broader areas without knowing the detailed surface characteristics of each landslide event. However, slope instability

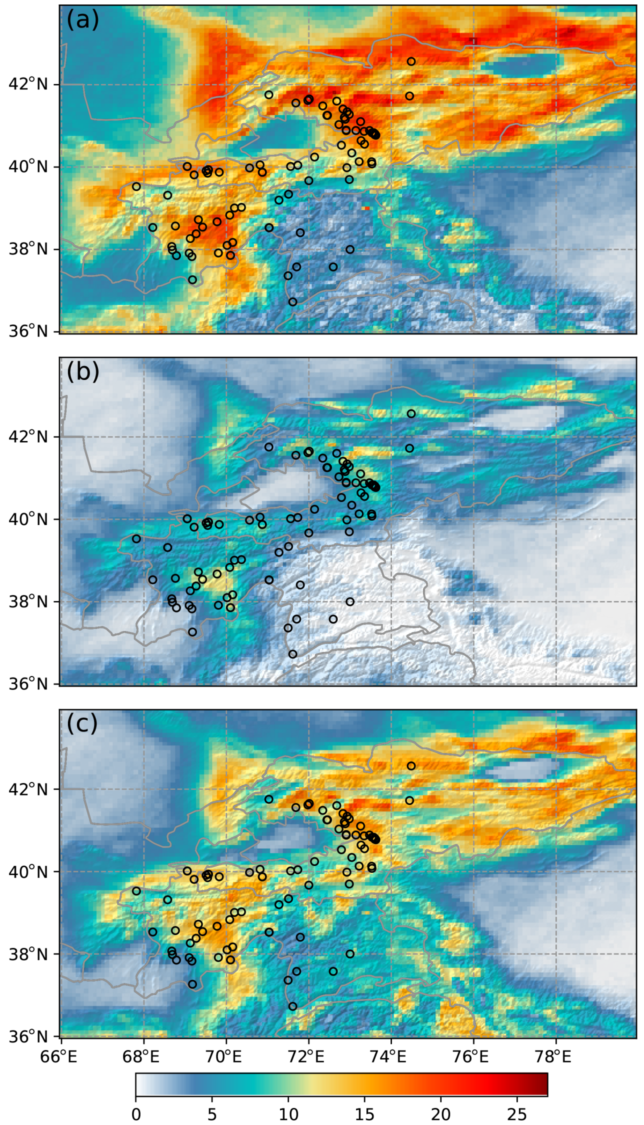

Figure 6. Mean annual exceedance (number of events per year) of **(a)** $I_{\text{mean}} = 5.05\,\text{mm}\,\text{d}^{-1}$, **(b)** $I_{\text{max}} = 14.05\,\text{mm}\,\text{d}^{-1}$, and **(c)** $Q = 15.65\,\text{mm}$ for the rainfall + snowmelt UTL events. Black circles: landslide events from the GLC and the GFLD. Topographic shading is based on DEM data from SRTM.

often results from numerous factors. The interaction between non-climatic characteristics and atmospheric triggers is also responsible for the initiation of landslides (Berti et al., 2012; Jia et al., 2020), which can not be captured by empirical methods. This is the reason why not all rainfall + snowmelt events that exceed $I_{\text{mean,th}}$, $I_{\text{max,th}}$, and Q_{th} triggered landslides (Fig. 6) even though the number of landslides is underestimated.

Uncertainty in landslide inventories and atmospheric data is a very common issue in studies investigating thresholds for landslide triggering. These two sources of uncertainty have been comprehensively discussed and quantified (e.g., Nikolopoulos et al., 2014, 2015; Marra et al., 2016, 2017; Marra, 2019; Rossi et al., 2017; Peres et al., 2018). Uncertainty in these two data sources generally results in an underestimation of rainfall thresholds, leading to a higher false alarm rate (Nikolopoulos et al., 2014, 2015; Marra et al., 2016; Peres et al., 2018). In the following subsections, we discuss the uncertainty stemming from the landslide inventories (the GLC and the GFLD) and rainfall and snowmelt in the HAR v2.

4.1.1 Uncertainty of landslide inventories

Uncertainties of the GLC and the GFLD are comprehensively discussed in Kirschbaum et al. (2010), Kirschbaum et al. (2015), and Froude and Petley (2018). The first major problem of these two data sets is that they underestimate the total number of landslides. This is because these two data sets' primary sources are media reports, which are biased towards events with human casualties (Carrara et al., 2003). The second issue is that the spatial distribution of landslides is biased towards populated areas. In our study area, landslide events also tend to cluster in areas with high population density, e.g., the eastern rim of the Fergana Basin and the Tajik Basin. Landslide number over remote areas is much more likely to be underreported. In addition, there is large uncertainty in landslide location because most media reports do not contain the exact location where landslides were initiated but rather just the name of the village, road, or city affected by landslides. An example in our case is the landslide event in the Issyk-Kul Basin (Fig. 1), the location of which is in a flat area, and the location accuracy provided by the GLC is "exact". This landslide event's initial zone must be different from the reported location and somewhere nearby with slopes. We also failed to determine the climatic trigger of this landslide event using the HAR v2. Last but not least, landslide timing was also reported with a certain degree of uncertainty. Although it is more typical that a landslide was reported after its actual occurrence (positive errors), negative errors are also possible depending on the interpretation of historical landslide information by an analyst (Peres et al., 2018). Our results show that using the entire weather event period leads to a better predictive performance than just using the UTL period (Table 1). This could be an indication of negative errors in the landslide timing.

Despite these known limitations, the GLC and the GFLD still provide the lower boundary of landslide number and are proven to be valuable in global and regional landslide studies. For example, the GLC has been successfully applied to detect the initiation of rainfall-induced landslides globally (Jia et al., 2020), to investigate the spatiotemporal distribution of potential landslide-triggering factors (Stanley et al., 2020), to explore the synoptic-scale precursors of landslides (Hunt and Dimri, 2021), and to evaluate the Global Landslide Hazard Assessment Model (Kirschbaum and Stanley, 2018). Although the landslide number is known to be incomplete, our results show that they can still present the seasonal distribution of landslide occurrence reasonably well (Fig. 4). This was also concluded by Kirschbaum et al. (2015), who stated that the reason for the unbiased seasonal distribution of landslide occurrence is that the compilation method depends on media alerts, which are consistent throughout the year. Additionally, even though location uncertainty exists, we could determine atmospheric triggers of 91 % of landslide events (87 out of 96). The reason could be that landslide-triggering rainfall and snowmelt events generally have a large spatial extend (Leonarduzzi et al., 2017).

4.1.2 Uncertainty of atmospheric data

Extracting weather data that can represent the exact weather conditions at landslide sites is always a challenge in studies investigating rainfall thresholds for landslide triggering. Rain gauges are the main source of rainfall information (Segoni et al., 2018), and it is very seldom that landslide initial locations are gauged. Due to the highly heterogeneous spatial distribution of precipitation, especially over complex terrain, there exists great uncertainty when rainfall is not directly measured from landslide initial points. Additionally, Marra et al. (2016) found that the initial points of shallow landslides and debris flows generally correspond to the local peak of rainfall. Rain depth decreases with distance, causing an underestimation when rainfall is measured away from the landslide initial point. Traditionally, the nearest gauge is used to represent the weather condition at the landslide site, which sometimes can be kilometers away. Nikolopoulos et al. (2015) examined other more complicated interpolation methods, such as inverse distance weighting and ordinary kriging, and concluded that these methods did not bring any particular added value to the simplest nearest neighbor method.

Using gridded data can avoid this allocation problem (Leonarduzzi et al., 2017). But uncertainties still exist since gridded data only represent the grid-mean value but not the "true" weather conditions at landslide sites. Nevertheless, it is still essential that the gridded data used in our study can accurately represent the grid-mean value. The WRF model configurations of the HAR v2, such as the forcing strategy and physical parameterization schemes, were carefully chosen to ensure its quality (Wang et al., 2021). Several studies (Pritchard et al., 2019; Li et al., 2020) indicate the high accuracy and quality of the old version of the High Asia Refined Analysis (HAR) (Maussion et al., 2014). Wang et al. (2021) compared the performance of the two versions of the HAR against in situ observations from 57 GSOD stations over the High Mountain Asia in terms of daily precipitation and air temperature at 2 m. It was concluded that compared to the old version, HAR v2 generally produces slightly higher precipitation amounts with a mean bias of $0.36\,\mathrm{mm\,d^{-1}}$. Furthermore, Hamm et al. (2020) compared the HAR v2 with other gridded precipitation data sets at different spatial resolutions, including reanalysis data and satellite-based precipitation retrieval, over a rugged terrain of the central Himalayas and the southwestern Tibetan Plateau. It was concluded that the HAR v2 is the only product that can resolve orographic precipitation, which is a fundamental process over complex terrain. Simulation of air temperature at 2 m in the HAR v2 is better than the old version due to the snow depth correction approach (Wang et al., 2021). Snowmelt in the HAR v2 is resolved by the Noah LSM, which only considers a single layer of snowpack (Koren et al., 1999). Several studies found uncertainty of the Noah LSM in reproducing the snow-related process, e.g., the overestimation of snow albedo (e.g., Chen et al., 2014; Minder et al., 2016; Tomasi et al., 2017). Nevertheless, the snow-related process is the major weakness of LSMs and needs further improvement in the future (Chen et al., 2014).

4.1.3 Impact of spatial resolution of atmospheric data

Previous studies have shown that the spatial resolutions of gridded rainfall data have impacts on identifying landslide-triggering thresholds (Marra et al., 2017; Nikolopoulos et al., 2017). To investigate the influence of spatial resolution of rainfall + snowmelt data on the event properties of landslide-triggering weather events and the triggering thresholds, we resampled the rainfall + snowmelt data from HAR v2 to lower resolutions (20, 30, and 40 km). Then, we repeated the procedure described in Sect. 2.2.2 to determine the event properties of LTE UTL events and their associated thresholds. The results are presented in Fig. 7. There are nine "not detected" events when using the original HAR v2 10 km data (Fig. 3), which means the rainfall + snowmelt amounts at these landslide grid points are near zero ($\leq 0.2\,\mathrm{mm\,d^{-1}}$) on the day and 1 d before landslide occurrence. By lowering the spatial resolution, more events can be detected. This implies the uncertainty in the reported landslide location since resampling of rainfall + snowmelt encompasses rainfall + snowmelt information from nearby grid points. In general, I_{mean} and I_{max} decrease with the increase in grid size, which is in line with the findings of Hamm et al. (2020) that higher-resolved products generally capture more extreme events than coarser products. I_{mean} and I_{max} thresholds defined by coarser products are also generally lower. The impact of grid size on Q is the opposite: larger grid size leads to higher Q and threshold value. This is closely associated with the increase in event duration with the increase in grid spacing, resulting from the fact that the resampling process can blend several localized events temporally together. However, lowering the spatial resolution does not lead to worse predictive performance. This, on the one hand, implies again that lower resolution can partly compensate for the uncertainty in landslide locations. On the other hand, it indicates that although landslide initiation itself is a highly localized phenomenon, the weather processes that ensure sufficient water

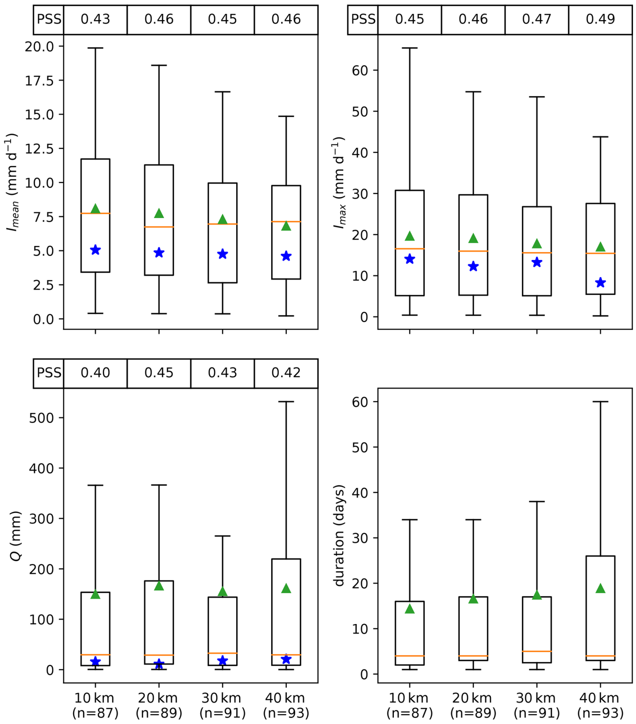

Figure 7. Boxplots demonstrating the impact of spatial resolution of atmospheric data on I_{mean}, I_{max}, Q, and duration of LTE UTL events, as well as the associated landslide-triggering thresholds (blue stars). The yellow line denotes the median, and the green triangle indicates the mean. Outliers are not shown for a better intercomparison, and n denotes the number of landslide events detected by rainfall + snowmelt.

input into the system and that trigger landslides can be clearly identified at the mesoscale (Prenner et al., 2018).

Based on the above analysis, it can be expected that a convection-permitting-scale (< 10 km) downscaling simulation would provide a more realistic representation of weather events that initialized landslides. Compared to such a high-resolution simulation, the HAR v2 10 km data would underestimate the intensity and overestimate the duration of landslide-triggering rainfall + snowmelt events. Moreover, the 10 km resolution of the HAR v2 is not able to explicitly resolve convection processes. Convection-permitting-scale simulations show improvement over simulations applying cumulus parameterization schemes in several aspects, such as more accurate reproduction of the timing of precipitation peaks (Ou et al., 2020; Zhou et al., 2021). However, a finer resolution has a lower tolerance for uncertainty in the landslide location. The potential of a kilometer-scale simulation cannot be realized if the landslide location uncertainty is larger than the grid size. Thus, for our study region, future studies should not focus only on acquiring high-resolution and high-quality atmospheric data but also on developing landslide inventories with higher location accuracy.

4.2 Climatic disposition

In probabilistic risk analysis (e.g., Scherer et al., 2013), the risk that a system experiences an adverse effect caused by a hazardous process is given as the product of hazard and vulnerability. Vulnerability itself depends on exposure and sensitivity. Adverse effects only occur when the elements at risk are exposed to a hazardous event. Thus, risk is a function of hazard, exposure, and sensitivity. Applying this risk concept to our case, the adverse effect is a landslide triggered by rainfall + snowmelt, and the hazardous process is a rainfall + snowmelt event that exceeds the defined thresholds. The risk that a location experiences a landslide triggered by rainfall + snowmelt depends on two factors: (a) how frequent a location is exposed to rainfall + snowmelt events that exceed $I_{\mathrm{mean,th}}$, $I_{\mathrm{max,th}}$, and Q_{th}, and (b) how sensitive slope instability can be triggered at this location. Climatic disposition represented by mean annual exceedance is actually factor (a) and comprises both aspects of hazard and exposure. Sensitivity is non-climatic landslide susceptibility that is only controlled by terrestrial characteristics. Thus, to assess landslide susceptibility, both climatic and non-climatic aspects need to be included.

The majority of landslide susceptibility studies only considered non-climatic factors. We compared our mean annul exceedance maps with a non-climatic landslide susceptibility map developed by Stanley and Kirschbaum (2017) at a resolution of approximately 1 km (Fig 8). This non-climatic susceptibility map was generated using a heuristic fuzzy approach, in which slope, faults, geology, forest loss, and road networks were taken into account. This map is chosen because it covers the whole of Kyrgyzstan and Tajikistan. Even though the non-climatic susceptibility map and our mean annual exceedance maps were generated by totally different methods, they share some similarities. They both show higher values over areas with steep slopes and lower values in intermontane basins and valleys. This is because topographic relief is considered the best first-order rainfall predictor (Bookhagen and Strecker, 2008). The non-climatic susceptibility map includes information on topography, and topography is explicitly resolved during dynamical downscaling. Mean annual exceedance maps not only display these local-scale features caused by topography but also comprise general atmospheric circulation processes. Around 23% of landslide events are located in zones with low and very low susceptibility. Landslide locations with low susceptibility in the eastern and southern rims of the Fergana Basin exhibit high climatic disposition (Fig. 6). This discrepancy between the non-climatic landslide susceptibility and our mean annual exceedance maps suggests that both climatic and non-climatic aspects need to be considered for landslide susceptibility mapping. Some event locations show both low susceptibility and low climatic disposition (e.g., in southwestern Tajikistan), which implies the uncertainty in reported landslide locations.

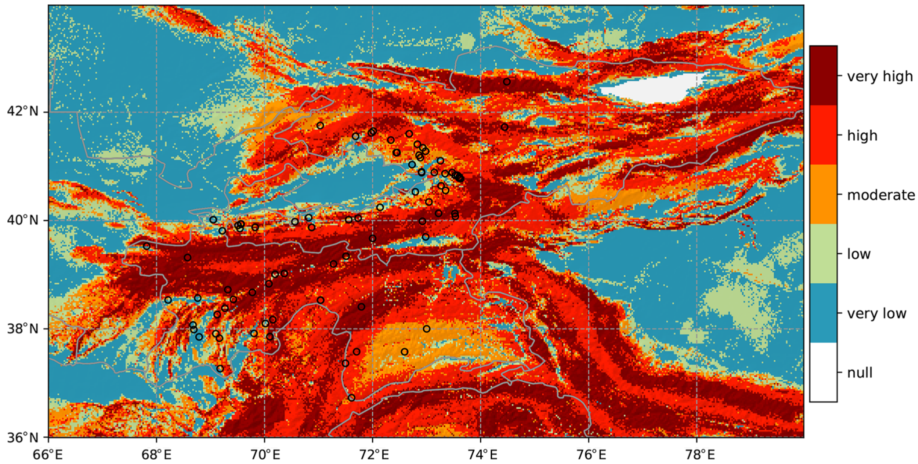

Figure 8. Non-climatic landslide susceptibility map computed using slope, geology, fault zones, road networks, and forest loss developed by Stanley and Kirschbaum (2017). Black circles: landslide events from the GLC and the GFLD. Topographic shading is based on DEM data from SRTM.

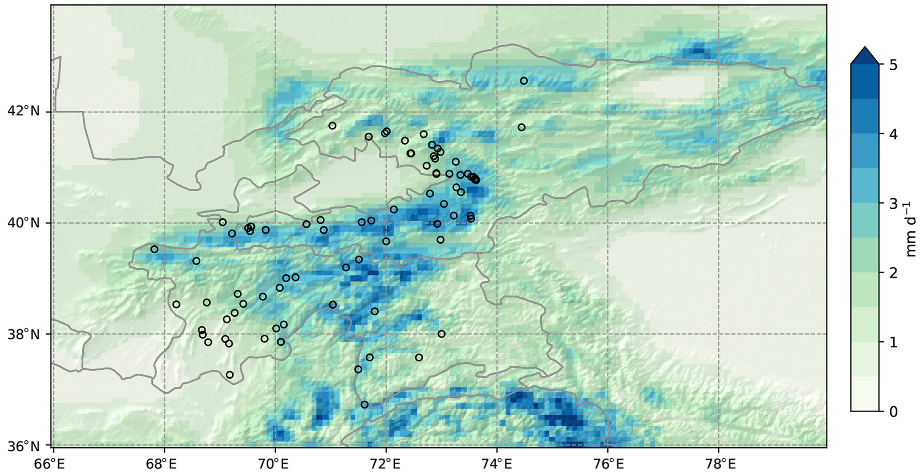

Figure 9. Annual sum of rainfall and snowmelt averaged over 2014–2018 from HAR v2. Black circles: landslide events from the GLC and the GFLD. Topographic shading is based on DEM data from SRTM.

In addition, some landslide susceptibility studies took climate into account, but they often simply applied averaged annual precipitation (e.g., Shahabi et al., 2014; Havenith et al., 2015b; Wang et al., 2015). Averaged annual precipitation only shows the climatological conditions in general. Mean annual exceedance is derived from weather-scale triggering conditions, and therefore, it also contains information on extreme processes. In our case, for instance, the mean annual rainfall + snowmelt map does not correspond well with landslide occurrences, especially in the Tajik Basin and the northeastern rim of the Fergana Basin (Fig. 9). But these landslide events are captured better in both mean annual exceedance maps (Fig. 6). This indicates the added value of climatic disposition derived from triggering conditions.

4.3 Thresholds for different landslide size

The GLC provides six categorized landslide sizes. Landslide events in Kyrgyzstan and Tajikistan fall into the following categories: (1) small – small landslide affecting one hill slope or small area; (2) medium – moderately sized landslide that could be either a single event or multiple landslides within an area and that involves a large volume of material; (3) large – large landslide or series of landslides that occur in one general area but cover a wide area; and (4) unknown (Kirschbaum et al., 2015). The GFLD does not contain information about landslide size. Therefore, for landslide events from the GFLD, we set the landslide size as "unknown". Table 2 presents the calibrated thresholds and corresponding statistical scores for these categories for UTL events. Using entire events leads to similar results (not presented here).

Table 2. Calibrated thresholds of I_{mean} (mm d^{-1}), I_{max} (mm d^{-1}), and Q (mm) for UTL events of the sum of rainfall and snowmelt (rainfall + snowmelt), as well as corresponding performance statistics for different categories of landslide size, and n refers to the number of landslides in each category.

Landslide size	Property	Threshold	HR	FAR	d	PSS	AUC
Small	I_{mean}	9.85	1.00	0.07	0.07	0.93	0.97
($n=5$)	I_{max}	21.55	1.00	0.07	0.07	0.93	0.97
	Q	124.25	1.00	0.04	0.04	0.96	0.98
Medium	I_{mean}	4.80	0.63	0.25	0.44	0.39	0.71
($n=41$)	I_{max}	14.05	0.49	0.12	0.53	0.37	0.73
	Q	9.65	0.73	0.35	0.44	0.38	0.72
Large	I_{mean}	8.10	0.55	0.11	0.47	0.44	0.72
($n=11$)	I_{max}	21.75	0.45	0.05	0.55	0.40	0.73
	Q	2.85	1.00	0.63	0.63	0.37	0.73
Unknown	I_{mean}	5.25	0.77	0.26	0.35	0.51	0.80
($n=30$)	I_{max}	13.25	0.73	0.17	0.32	0.57	0.81
	Q	16.90	0.77	0.25	0.34	0.51	0.79

Interestingly, the thresholds for landslides with small sizes are higher than other categories and have the best predictive performance. All of these five small-sized landslide events are snowmelt-contributed events that occurred from March to May. The worse predictive performance for landslides with larger sizes could indicate that for those events, the triggering mechanism is much more complicated than small-sized events, and other non-atmospheric factors might also play a role. However, the sample size of small-sized landslide events is too small to draw a robust conclusion. The number of small-sized landslides is expected to be underreported since media reports are biased towards events with more severe impacts.

5 Conclusions

In this study, we combined gridded atmospheric data from the HAR v2 with 87 landslide records extracted from the GLC and the GFLD to analyze rainfall and snowmelt conditions that triggered landslides in Kyrgyzstan and Tajikistan. Thresholds for landslide triggering were determined for different event properties for rainfall, snowmelt, and rainfall + snowmelt. Mean annual exceedance maps were generated based on the defined thresholds.

Monthly landslide counts in Kyrgyzstan and Tajikistan correspond well with the monthly distribution of rainfall + snowmelt. An exception is March when soil temperature at the top soil layer (0–0.1 m) and air temperature at 2 m are both below zero. Investigation of the relationship between landslides and soil temperature could be a topic for future studies. Snowmelt plays a crucial role in landslide triggering in Kyrgyzstan and Tajikistan since it contributes to the triggering of 40 % of landslide events.

By including snowmelt as an additional trigger, the skill of landslide prediction was significantly improved. I_{mean}, I_{max}, and Q have similar predictive performance. Thresholds of $I_{\text{mean}} = 5.05$ mm d^{-1}, $I_{\text{max}} = 14.05$ mm d^{-1}, and $Q = 15.65$ mm for UTL events were defined for landslide triggering in Kyrgyzstan and Tajikistan. Using the entire period of weather events leads to similar threshold values but better predictive performance. This could indicate uncertainty in landslide timing. Mean annual exceedance maps derived from these thresholds depict climatic disposition and have added value in landslide susceptibility mapping.

The majority of previous studies applied rainfall estimates from in situ gauges or satellite retrievals. Our study demonstrates the potential of the regional climate model (RCM) in landslide prediction. Dynamical downscaling products generated by RCMs can provide physically consistent, high-resolution data that are extremely valuable for data-scarce areas. Given the global applicability of the dynamical downscaling method, our approach can also be applied in other regions as long as the number and quality of landslide records are sufficient. Even though a higher-resolved downscaling product can reproduce landslide-triggering weather events more realistically, it has a lower tolerance for the uncertainty in landslide locations and does not necessarily lead to better predictive performance. Future studies in Kyrgyzstan and Tajikistan should focus on developing landslide inventories with both high location accuracy and timing accuracy to reduce the uncertainty in triggering thresholds.

Appendix A

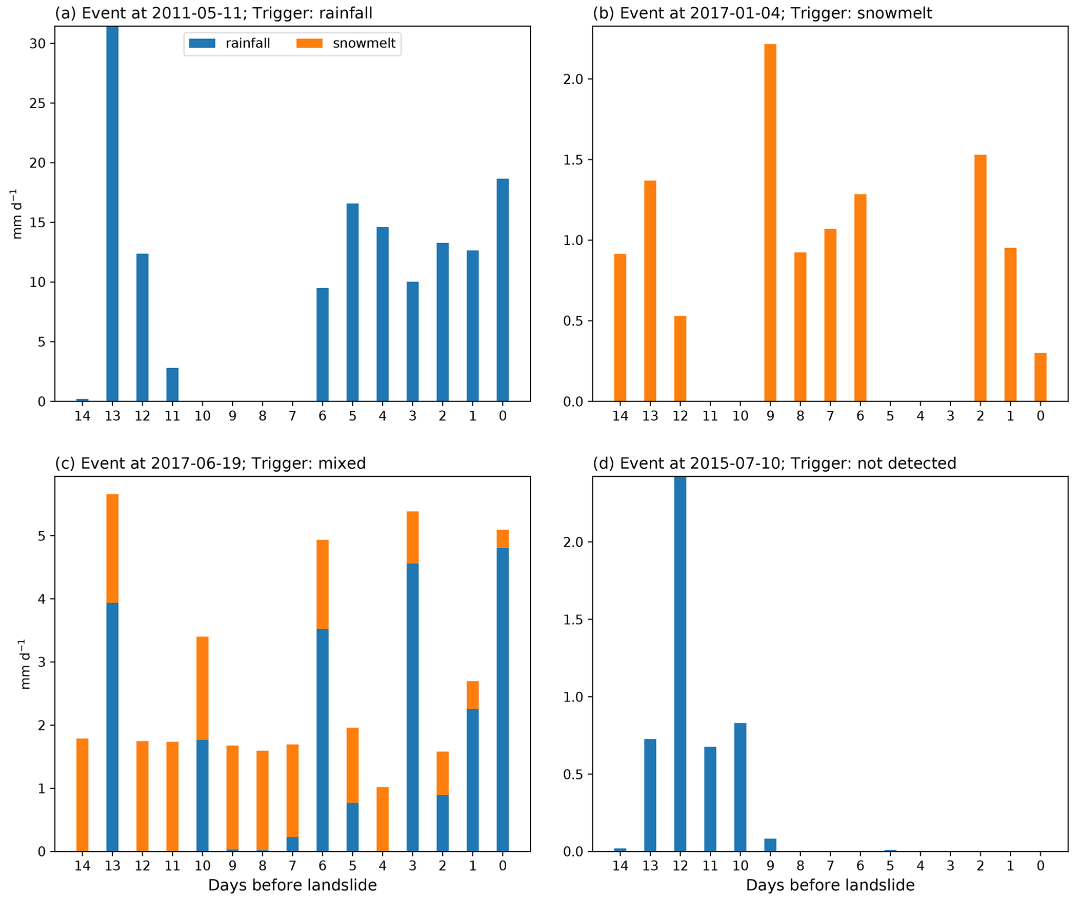

Figure A1. Event-based temporal process of rainfall and snowmelt for selected landslide events with landslide triggers defined as **(a)** rainfall, **(b)** snowmelt, **(c)** mixed, and **(d)** not detected, according to the method described in Sect. 2.2.1.

Table A1. Landslide events in Kyrgyzstan and Tajikistan extracted from the GLC and the GFLD from 2004 to 2018. The column "trigger" indicates the trigger of landslide events detected by the HAR v2.

Event date	Source	Longitude	Latitude	Country	Trigger
17 Apr 2004	GFLD	73.0420	40.3428	Kyrgyzstan	Mixed
22 May 2004	GFLD	69.2172	39.8106	Tajikistan	Rainfall
14 Jun 2004	GFLD	70.8718	39.8734	Kyrgyzstan	Rainfall
17 Nov 2004	GFLD	70.0802	38.8324	Tajikistan	Mixed
13 Mar 2005	GFLD	69.0502	40.0141	Tajikistan	Mixed
9 Apr 2005	GFLD	69.2656	38.3801	Tajikistan	Mixed
25 Mar 2007	GLC	70.1951	39.0071	Tajikistan	Mixed
1 Apr 2007	GLC	72.5920	37.5760	Tajikistan	Mixed
5 Apr 2007	GLC	71.6110	36.7270	Tajikistan	Snowmelt
17 Apr 2007	GLC	71.6849	41.5552	Kyrgyzstan	Rainfall
17 Apr 2007	GLC	68.2140	38.5330	Tajikistan	Rainfall
22 Apr 2007	GLC	73.1416	40.8870	Kyrgyzstan	Rainfall
5 Jun 2007	GFLD	69.1633	37.8276	Tajikistan	Rainfall
21 Jul 2007	GLC	73.0000	38.0000	Tajikistan	Mixed
22 Jul 2007	GLC	70.4400	40.7500	Tajikistan	Not detected
22 Jul 2007	GFLD	71.0363	38.5289	Tajikistan	Rainfall
16 Apr 2009	GFLD	71.9767	41.6184	Kyrgyzstan	Rainfall
21 Apr 2009	GLC	68.7882	37.8515	Tajikistan	Rainfall
5 May 2009	GFLD	70.1529	38.1701	Tajikistan	Rainfall
7 May 2009	GFLD	69.7741	38.6726	Tajikistan	Rainfall
11 May 2009	GFLD	71.0363	38.5289	Tajikistan	Snowmelt
14 May 2009	GLC	68.6900	37.9867	Tajikistan	Rainfall
16 May 2009	GFLD	71.0363	38.5289	Tajikistan	Snowmelt
20 May 2009	GFLD	69.3199	38.7221	Tajikistan	Rainfall
13 Mar 2010	GFLD	69.0502	40.0141	Tajikistan	Snowmelt
7 May 2010	GLC	69.8054	37.9148	Tajikistan	Rainfall
7 May 2010	GFLD	70.0994	37.8560	Tajikistan	Rainfall
3 Jun 2010	GLC	72.9227	39.9854	Kyrgyzstan	Mixed
11 May 2011	GLC	72.8282	41.4088	Kyrgyzstan	Rainfall
12 Jun 2011	GLC	69.1238	38.2644	Tajikistan	Rainfall
12 Jun 2011	GLC	69.5667	39.9342	Kyrgyzstan	Rainfall
12 May 2012	GLC	70.8159	40.0538	Kyrgyzstan	Rainfall
13 May 2012	GFLD	70.8718	39.8734	Kyrgyzstan	Rainfall
28 Jun 2013	GLC	72.0106	41.6518	Kyrgyzstan	Rainfall
12 Apr 2014	GLC	69.0971	37.9107	Tajikistan	Rainfall
12 Apr 2014	GFLD	70.0994	37.8560	Tajikistan	Rainfall
16 Apr 2014	GFLD	68.6749	38.0710	Tajikistan	Rainfall
26 Apr 2014	GFLD	68.7626	38.5685	Tajikistan	Rainfall
3 Apr 2015	GFLD	69.4222	38.5428	Tajikistan	Rainfall
8 May 2015	GLC	70.0162	38.0991	Tajikistan	Rainfall
24 May 2015	GLC	72.9053	40.8986	Kyrgyzstan	Rainfall
24 May 2015	GFLD	73.2559	41.1036	Kyrgyzstan	Rainfall
10 Jul 2015	GLC	70.4275	39.0712	Tajikistan	Not detected
16 Jul 2015	GLC	71.7041	37.5773	Tajikistan	Rainfall
21 Jul 2015	GFLD	71.7929	38.4071	Tajikistan	Rainfall
26 Apr 2016	GLC	72.9071	40.8894	Kyrgyzstan	Not detected
9 May 2016	GLC	68.5748	39.3160	Tajikistan	Mixed
15 May 2016	GLC	72.9293	41.3431	Kyrgyzstan	Rainfall
23 May 2016	GLC	72.7907	40.5304	Kyrgyzstan	Rainfall
27 May 2016	GLC	69.8266	39.8751	Kyrgyzstan	Rainfall
28 May 2016	GLC	71.5577	40.0150	Kyrgyzstan	Mixed
16 Jun 2016	GLC	72.3374	41.4850	Kyrgyzstan	Rainfall
20 Jun 2016	GLC	73.5233	40.1293	Kyrgyzstan	Rainfall
27 Jun 2016	GLC	74.4438	41.7246	Kyrgyzstan	Rainfall
29 Jun 2016	GLC	73.1415	41.7649	Kyrgyzstan	Not detected

Table A1. Continued.

Event date	Source	Longitude	Latitude	Country	Trigger
29 Jul 2016	GLC	69.5597	39.9377	Kyrgyzstan	Rainfall
16 Aug 2016	GLC	78.3019	42.6831	Kyrgyzstan	Not detected
18 Aug 2016	GLC	70.5626	39.9790	Tajikistan	Rainfall
4 Jan 2017	GLC	71.9999	39.6699	Kyrgyzstan	Snowmelt
26 Jan 2017	GLC	72.8834	40.8960	Kyrgyzstan	Not detected
26 Mar 2017	GFLD	73.5725	40.8316	Kyrgyzstan	Mixed
7 Apr 2017	GLC	73.6257	40.7733	Kyrgyzstan	Snowmelt
9 Apr 2017	GLC	73.5335	40.8320	Kyrgyzstan	Snowmelt
10 Apr 2017	GLC	69.5091	39.9095	Kyrgyzstan	Mixed
11 Apr 2017	GLC	72.8601	41.2047	Kyrgyzstan	Mixed
14 Apr 2017	GFLD	73.5725	40.8316	Kyrgyzstan	Mixed
16 Apr 2017	GLC	73.2668	40.6430	Kyrgyzstan	Snowmelt
16 Apr 2017	GLC	73.6000	40.7836	Kyrgyzstan	Snowmelt
17 Apr 2017	GLC	73.6047	40.8044	Kyrgyzstan	Mixed
18 Apr 2017	GLC	71.4973	37.3628	Tajikistan	Mixed
18 Apr 2017	GLC	72.9069	40.8838	Kyrgyzstan	Rainfall
22 Apr 2017	GLC	73.3402	40.8663	Kyrgyzstan	Mixed
23 Apr 2017	GLC	71.5074	39.3410	Tajikistan	Snowmelt
23 Apr 2017	GLC	72.8835	41.1610	Kyrgyzstan	Rainfall
23 Apr 2017	GFLD	72.9801	41.2790	Kyrgyzstan	Mixed
29 Apr 2017	GLC	73.4724	40.8864	Kyrgyzstan	Mixed
29 Apr 2017	GFLD	73.2203	40.1325	Kyrgyzstan	Mixed
30 Apr 2017	GLC	72.4381	41.2550	Kyrgyzstan	Rainfall
30 Apr 2017	GLC	73.5310	40.0774	Kyrgyzstan	Mixed
10 May 2017	GLC	74.4847	42.5635	Kyrgyzstan	Mixed
11 May 2017	GLC	73.3497	40.5560	Kyrgyzstan	Rainfall
16 May 2017	GLC	71.0302	41.7545	Kyrgyzstan	Rainfall
17 May 2017	GLC	72.6771	41.6014	Kyrgyzstan	Rainfall
28 May 2017	GLC	71.2755	39.1978	Tajikistan	Mixed
19 Jun 2017	GLC	72.9814	39.6978	Kyrgyzstan	Mixed
19 Jun 2017	GLC	71.7318	40.0439	Kyrgyzstan	Rainfall
26 Jun 2017	GLC	67.8173	39.5267	Tajikistan	Rainfall
28 Jun 2017	GLC	68.5480	39.3951	Tajikistan	Not detected
29 Jun 2017	GLC	72.7303	41.0321	Kyrgyzstan	Rainfall
29 Jun 2017	GLC	72.4521	41.2557	Kyrgyzstan	Rainfall
3 Jul 2017	GLC	70.3650	39.0219	Tajikistan	Rainfall
3 Jul 2017	GLC	68.4838	39.1172	Tajikistan	Not detected
4 Jul 2017	GLC	69.5279	39.8102	Kyrgyzstan	Not detected
13 May 2018	GLC	69.5445	39.8526	Kyrgyzstan	Rainfall
16 May 2018	GLC	69.1773	37.2642	Tajikistan	Rainfall
21 May 2018	GLC	72.1386	40.2437	Kyrgyzstan	Mixed

Table A2. K-fold validation results. Mean values and standard deviations (in parentheses) for thresholds of I_{mean} (mm d^{-1}), I_{max} (mm d^{-1}), and Q (mm) for entire events of rainfall, snowmelt, and the sum of rainfall and snowmelt (rainfall + snowmelt), as well as corresponding performance statistics.

Predictor	Property	Threshold	HR	FAR	d	PSS	AUC
Rainfall	I_{mean}	3.76 (0.33)	0.56 (0.14)	0.33 (0.03)	0.56 (0.10)	0.23 (0.13)	0.62 (0.01)
	I_{max}	11.06 (0.66)	0.46 (0.16)	0.18 (0.02)	0.57 (0.15)	0.28 (0.15)	0.65 (0.01)
	Q	12.31 (3.88)	0.53 (0.16)	0.25 (0.07)	0.55 (0.10)	0.27 (0.10)	0.67 (0.01)
Snowmelt	I_{mean}	7.06 (0.02)	0.22 (0.14)	0.06 (0.01)	0.78 (0.14)	0.16 (0.14)	0.31 (0.02)
	I_{max}	13.61 (0.44)	0.23 (0.13)	0.04 (0.01)	0.77 (0.13)	0.19 (0.12)	0.32 (0.01)
	Q	122.38 (7.93)	0.23 (0.13)	0.03 (0.01)	0.77 (0.13)	0.20 (0.12)	0.33 (0.01)
Rainfall + snowmelt	I_{mean}	4.96 (0.02)	0.70 (0.13)	0.25 (0.02)	0.40 (0.08)	0.45 (0.14)	0.78 (0.01)
	I_{max}	12.93 (0.37)	0.65 (0.15)	0.15 (0.01)	0.39 (0.13)	0.49 (0.15)	0.81 (0.01)
	Q	17.20 (0.14)	0.71 (0.15)	0.23 (0.02)	0.38 (0.10)	0.48 (0.13)	0.81 (0.01)

Table A3. K-fold validation results. Mean values and standard deviations (in parentheses) for thresholds of I_{mean} (mm d^{-1}), I_{max} (mm d^{-1}), and Q (mm) for UTL events of rainfall, snowmelt, and the sum of rainfall and snowmelt (rainfall + snowmelt), as well as corresponding performance statistics.

Predictor	Property	Threshold	HR	FAR	d	PSS	AUC
Rainfall	I_{mean}	4.04 (1.47)	0.45 (0.13)	0.33 (0.10)	0.66 (0.08)	0.12 (0.08)	0.59 (0.01)
	I_{max}	10.94 (1.47)	0.34 (0.06)	0.18 (0.04)	0.68 (0.05)	0.16 (0.06)	0.58 (0.01)
	Q	10.21 (2.22)	0.46 (0.09)	0.29 (0.04)	0.62 (0.09)	0.17 (0.11)	0.59 (0.01)
Snowmelt	I_{mean}	7.14 (0.26)	0.21 (0.10)	0.06 (0.02)	0.79 (0.10)	0.15 (0.09)	0.31 (0.02)
	I_{max}	12.88 (0.23)	0.23 (0.12)	0.05 (0.01)	0.77 (0.12)	0.18 (0.11)	0.32 (0.02)
	Q	99.95 (4.67)	0.22 (0.13)	0.04 (0.01)	0.78 (0.13)	0.18 (0.13)	0.32 (0.02)
Rainfall + snowmelt	I_{mean}	5.35 (0.85)	0.61 (0.22)	0.23 (0.04)	0.47 (0.17)	0.38 (0.18)	0.76 (0.01)
	I_{max}	13.54 (0.56)	0.56 (0.15)	0.14 (0.01)	0.47 (0.14)	0.42 (0.14)	0.77 (0.01)
	Q	15.83 (0.44)	0.63 (0.13)	0.25 (0.02)	0.45 (0.10)	0.38 (0.12)	0.76 (0.01)

Code and data availability. The landslide data and atmospheric data used in this study are freely available from the following links:

- Global Landslide Catalog: https://maps.nccs.nasa.gov/arcgis/home/item.html?id=eec7aee8d2e040c7b8d3ee5fd0e0d7b9 (NASA, 2021)
- Global Fatal Landslide Database: https://shefuni.maps.arcgis.com/apps/webappviewer/index.html?id=8458951270904fc29527254492517063 (UOS, 2021)
- High Asia Refined Analysis version 2: https://www.klima.tu-berlin.de/HARv2 (TUB, 2021).

The source code used in this study is freely available upon request.

Author contributions. All authors were involved in study conceptualization and writing of the manuscript. XW collected the data, carried out the analyses, and produced the visualizations.

Competing interests. The authors declare that they have no conflict of interest.

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

Acknowledgements. This work was supported by the German Federal Ministry of Education and Research (BMBF) under the framework of the "Climatic and Tectonic Natural Hazards in Central Asia (CaTeNA)" project (grant no. FKZ 03G0878G).

Financial support. This work was conducted under the framework of the "Climatic and Tectonic Natural Hazards in Central Asia (CaTeNA)" project (grant no. FKZ 03G0878G) supported by the German Federal Ministry of Education and Research (BMBF).

This open-access publication was funded by Technische Universität Berlin.

Review statement. This paper was edited by David J. Peres and reviewed by two anonymous referees.

References

Barbosa, N., Andreani, L., Gloaguen, R., and Ratschbacher, L.: Window-Based Morphometric Indices as Predictive Variables for Landslide Susceptibility Models, Remote Sens., 13, 451, https://doi.org/10.3390/rs13030451, 2021.

Behling, R. and Roessner, S.: Multi-temporal landslide inventory for a study area in Southern Kyrgyzstan derived from RapidEye satellite time series data (2009–2013), V. 1.0. GFZ Data Services, Potsdam, Germany, https://doi.org/10.5880/GFZ.1.4.2020.001, 2020.

Berti, M., Martina, M., Franceschini, S., Pignone, S., Simoni, A., and Pizziolo, M.: Probabilistic rainfall thresholds for landslide occurrence using a Bayesian approach, J. Geophys. Res.-Earth, 117, F04006, https://doi.org/10.1029/2012JF002367, 2012.

Bookhagen, B. and Strecker, M. R.: Orographic barriers, high-resolution TRMM rainfall, and relief variations along the eastern Andes, Geophysical Res. Lett., 35, L06403, https://doi.org/10.1029/2007GL032011, 2008.

Braun, A., Fernandez-Steeger, T., Havenith, H.-B., and Torgoev, A.: Landslide Susceptibility Mapping with Data Mining Methods – a Case Study from Maily-Say, Kyrgyzstan, in: Engineering Geology for Society and Territory – Volume 2, Springer, Cham, 995–998, 2015.

Bui, D. T., Pradhan, B., Lofman, O., Revhaug, I., and Dick, Ø. B.: Regional prediction of landslide hazard using probability analysis of intense rainfall in the Hoa Binh province, Vietnam, Nat. Hazards, 66, 707–730, 2013.

Carrara, A., Crosta, G., and Frattini, P.: Geomorphological and historical data in assessing landslide hazard, Earth Surf. Proc. Land., 28, 1125–1142, 2003.

Chen, F., Barlage, M., Tewari, M., Rasmussen, R., Jin, J., Lettenmaier, D., Livneh, B., Lin, C., Miguez-Macho, G., Niu, G. Y., and Wen, L.: Modeling seasonal snowpack evolution in the complex terrain and forested Colorado Headwaters region: A model intercomparison study, J. Geophys. Res.-Atmos., 119, 13–795, 2014.

Dai, F., Lee, C., and Ngai, Y. Y.: Landslide risk assessment and management: an overview, Eng. Geol., 64, 65–87, 2002.

Dietz, A. J., Conrad, C., Kuenzer, C., Gesell, G., and Dech, S.: Identifying changing snow cover characteristics in central Asia between 1986 and 2014 from remote sensing data, Remote Sens., 6, 12752–12775, 2014.

Fawcett, T.: An introduction to ROC analysis, Pattern Recognit. Lett., 27, 861–874, 2006.

Froude, M. J. and Petley, D. N.: Global fatal landslide occurrence from 2004 to 2016, Nat. Hazards Earth Syst. Sci., 18, 2161–2181, https://doi.org/10.5194/nhess-18-2161-2018, 2018.

Gariano, S. L., Brunetti, M. T., Iovine, G., Melillo, M., Peruccacci, S., Terranova, O., Vennari, C., and Guzzetti, F.: Calibration and validation of rainfall thresholds for shallow landslide forecasting in Sicily, southern Italy, Geomorphology, 228, 653–665, 2015.

Giannecchini, R., Galanti, Y., Avanzi, G. D., and Barsanti, M.: Probabilistic rainfall thresholds for triggering debris flows in a human-modified landscape, Geomorphology, 257, 94–107, 2016.

Guzzetti, F., Peruccacci, S., Rossi, M., and Stark, C. P.: Rainfall thresholds for the initiation of landslides in central and southern Europe, Meteorol. Atmos. Phys., 98, 239–267, 2007.

Hamm, A., Arndt, A., Kolbe, C., Wang, X., Thies, B., Boyko, O., Reggiani, P., Scherer, D., Bendix, J., and Schneider, C.: Intercomparison of Gridded Precipitation Datasets over a Sub-Region of the Central Himalaya and the Southwestern Tibetan Plateau, Water, 12, 3271, https://doi.org/10.3390/w12113271, 2020.

Hanssen, A. W. and Kuipers, W. J. A.: On the relationship between the frequency of rain and various meteorological parameters, Meded. Verh., 81, 2–15, 1965.

Havenith, H.-B., Strom, A., Torgoev, I., Torgoev, A., Lamair, L., Ischuk, A., and Abdrakhmatov, K.: Tien Shan geohazards database: Earthquakes and landslides, Geomorphology, 249, 16–31, 2015a.

Havenith, H.-B., Torgoev, A., Schlögel, R., Braun, A., Torgoev, I., and Ischuk, A.: Tien Shan geohazards database: Landslide susceptibility analysis, Geomorphology, 249, 32–43, 2015b.

Hunt, K. M. and Dimri, A.: Synoptic-scale precursors of landslides in the western Himalaya and Karakoram, Sci. Total Environ., 776, 145895, https://doi.org/10.1016/j.scitotenv.2021.145895, 2021.

Ilyasov, S., Zabenko, O., Gaydamak, N., Kirilenko, A., Myrsaliev, N., Shevchenko, V., and Penkina, L.: Climate profile of the Kyrgyz Republic, The State Agency for Environmental Protection and Forestry under the Government of the Kyrgyz Republic and The United Nations Development Programme, Bishkek, Kyrgyzstan, 2013.

Jarraud, M.: Guide to meteorological instruments and methods of observation, WMO-No. 8, World Meteorological Organisation, Geneva, Switzerland, 2008.

Jia, G., Tang, Q., and Xu, X.: Evaluating the performances of satellite-based rainfall data for global rainfall-induced landslide warnings, Landslides, 17, 283–299, 2020.

Khan, Y. A., Lateh, H., Baten, M. A., and Kamil, A. A.: Critical antecedent rainfall conditions for shallow landslides in Chittagong City of Bangladesh, Environ. Earth Sci., 67, 97–106, 2012.

Kirschbaum, D. and Stanley, T.: Satellite-based assessment of rainfall-triggered landslide hazard for situational awareness, Earth's Future, 6, 505–523, 2018.

Kirschbaum, D., Stanley, T., and Zhou, Y.: Spatial and temporal analysis of a global landslide catalog, Geomorphology, 249, 4–15, 2015.

Kirschbaum, D. B., Adler, R., Hong, Y., Hill, S., and Lerner-Lam, A.: A global landslide catalog for hazard applications: method, results, and limitations, Nat. Hazards, 52, 561–575, 2010.

Koren, V., Schaake, J., Mitchell, K., Duan, Q.-Y., Chen, F., and Baker, J.: A parameterization of snowpack and frozen ground intended for NCEP weather and climate models, J. Geophys. Res.-Atmos., 104, 19569–19585, 1999.

Leonarduzzi, E., Molnar, P., and McArdell, B. W.: Predictive performance of rainfall thresholds for shallow landslides in Switzerland from gridded daily data, Water Resour. Res., 53, 6612–6625, 2017.

Li, D., Yang, K., Tang, W., Li, X., Zhou, X., and Guo, D.: Characterizing precipitation in high altitudes of the western Tibetan plateau with a focus on major glacier areas, Int. J. Climatol., 40, 5114–5127, 2020.

Marra, F.: Rainfall thresholds for landslide occurrence: systematic underestimation using coarse temporal resolution data, Nat. Hazards, 95, 883–890, 2019.

Marra, F., Nikolopoulos, E., Creutin, J., and Borga, M.: Space–time organization of debris flows-triggering rainfall and its effect on the identification of the rainfall threshold relationship, J. Hydrol., 541, 246–255, 2016.

Marra, F., Destro, E., Nikolopoulos, E. I., Zoccatelli, D., Creutin, J. D., Guzzetti, F., and Borga, M.: Impact of rainfall spatial aggregation on the identification of debris flow occurrence thresholds, Hydrol. Earth Syst. Sci., 21, 4525–4532, https://doi.org/10.5194/hess-21-4525-2017, 2017.

Maussion, F., Scherer, D., Mölg, T., Collier, E., Curio, J., and Finkelnburg, R.: Precipitation seasonality and variability over the Tibetan Plateau as resolved by the High Asia Reanalysis, J. Climate, 27, 1910–1927, 2014.

Minder, J. R., Letcher, T. W., and Skiles, S. M.: An evaluation of high-resolution regional climate model simulations of snow cover and albedo over the Rocky Mountains, with implications for the simulated snow-albedo feedback, J. Geophys. Res.-Atmos., 121, 9069–9088, 2016.

Mostbauer, K., Kaitna, R., Prenner, D., and Hrachowitz, M.: The temporally varying roles of rainfall, snowmelt and soil moisture for debris flow initiation in a snow-dominated system, Hydrol. Earth Syst. Sci., 22, 3493–3513, https://doi.org/10.5194/hess-22-3493-2018, 2018.

NASA: Glabal Landslide Catalog Points, available at: https://maps.nccs.nasa.gov/arcgis/home/item.html?id=eec7aee8d2e040c7b8d3ee5fd0e0d7b9, last access: 6 July 2021.

Nikolopoulos, E., Borga, M., Creutin, J., and Marra, F.: Estimation of debris flow triggering rainfall: Influence of rain gauge density and interpolation methods, Geomorphology, 243, 40–50, 2015.

Nikolopoulos, E., Destro, E., Maggioni, V., Marra, F., and Borga, M.: Satellite rainfall estimates for debris flow prediction: an evaluation based on rainfall accumulation–duration thresholds, J. Hydrometeorol., 18, 2207–2214, 2017.

Nikolopoulos, E. I., Crema, S., Marchi, L., Marra, F., Guzzetti, F., and Borga, M.: Impact of uncertainty in rainfall estimation on the identification of rainfall thresholds for debris flow occurrence, Geomorphology, 221, 286–297, 2014.

Ou, T., Chen, D., Chen, X., Lin, C., Yang, K., Lai, H.-W., and Zhang, F.: Simulation of summer precipitation diurnal cycles over the Tibetan Plateau at the gray-zone grid spacing for cumulus parameterization, Clim. Dynam., 54, 3525–3539, 2020.

Ozturk, U., Pittore, M., Behling, R., Roessner, S., Andreani, L., and Korup, O.: How robust are landslide susceptibility estimates?, Landslides, 18, 681–695, 2020.

Palmer, J.: A slippery slope: Could climate change lead to more landslides?, Eos, 101, https://doi.org/10.1029/2020EO151418, 2020.

Peres, D. J., Cancelliere, A., Greco, R., and Bogaard, T. A.: Influence of uncertain identification of triggering rainfall on the assessment of landslide early warning thresholds, Nat. Hazards Earth Syst. Sci., 18, 633–646, https://doi.org/10.5194/nhess-18-633-2018, 2018.

Piciullo, L., Gariano, S. L., Melillo, M., Brunetti, M. T., Peruccacci, S., Guzzetti, F., and Calvello, M.: Definition and performance of a threshold-based regional early warning model for rainfall-induced landslides, Landslides, 14, 995–1008, 2017.

Postance, B., Hillier, J., Dijkstra, T., and Dixon, N.: Comparing threshold definition techniques for rainfall-induced landslides: A national assessment using radar rainfall, Earth Surf. Proc. Land., 43, 553–560, 2018.

Prenner, D., Kaitna, R., Mostbauer, K., and Hrachowitz, M.: The value of using multiple hydrometeorological variables to predict temporal debris flow susceptibility in an alpine environment, Water Resour. Res., 54, 6822–6843, 2018.

Pritchard, D. M., Forsythe, N., Fowler, H. J., O'Donnell, G. M., and Li, X.-F.: Evaluation of Upper Indus near-surface climate representation by WRF in the high Asia refined analysis, J. Hydrometeorol., 20, 467–487, 2019.

Rodwell, M. J., Richardson, D. S., Hewson, T. D., and Haiden, T.: A new equitable score suitable for verifying precipitation in numerical weather prediction, Q. J. Roy. Meteorol. Soc., 136, 1344–1363, 2010.

Rossi, M., Luciani, S., Valigi, D., Kirschbaum, D., Brunetti, M., Peruccacci, S., and Guzzetti, F.: Statistical approaches for the definition of landslide rainfall thresholds and their uncertainty using rain gauge and satellite data, Geomorphology, 285, 16–27, 2017.

Saponaro, A., Pilz, M., Wieland, M., Bindi, D., Moldobekov, B., and Parolai, S.: Landslide susceptibility analysis in data-scarce regions: the case of Kyrgyzstan, Bull. Eng. Geol.e Environ., 74, 1117–1136, 2015.

Scherer, D., Fehrenbach, U., Lakes, T., Lauf, S., Meier, F., and Schuster, C.: Quantification of heat-stress related mortality hazard, vulnerability and risk in Berlin, Germany, Erde, 144, 238–259, 2013.

Segoni, S., Piciullo, L., and Gariano, S. L.: A review of the recent literature on rainfall thresholds for landslide occurrence, Landslides, 15, 1483–1501, 2018.

Shahabi, H., Khezri, S., Ahmad, B. B., and Hashim, M.: Landslide susceptibility mapping at central Zab basin, Iran: A comparison between analytical hierarchy process, frequency ratio and logistic regression models, Catena, 115, 55–70, 2014.

Stanley, T. and Kirschbaum, D. B.: A heuristic approach to global landslide susceptibility mapping, Nat. Hazards, 87, 145–164, 2017.

Stanley, T., Kirschbaum, D. B., Pascale, S., and Kapnick, S.: Extreme Precipitation in the Himalayan Landslide Hotspot, in: Satellite Precipitation Measurement, Springer, Cham, 1087–1111, 2020.

Tewari, M., Chen, F., Wang, W., Dudhia, J., LeMone, M., Mitchell, K., Ek, M., Gayno, G., Wegiel, J., and Cuenca, R.: Implementation and verification of the unified NOAH land surface model in the WRF model, in: vol. 1115, 20th conference on weather analysis and forecasting/16th conference on numerical weather prediction, American Meteorological Society, Seattle, WA, 2165–2170, 2004.

Tomasi, E., Giovannini, L., Zardi, D., and de Franceschi, M.: Optimization of Noah and Noah_MP WRF land surface schemes in snow-melting conditions over complex terrain, Mon. Weather Rev., 145, 4727–4745, 2017.

Torgoev, I., Alioshin, Y. G., and Torgoev, A.: Monitoring landslides in Kyrgyzstan, FOG – Freiberg Online Geoscience, Freiberg, 130–139, 2012.

TUB: The High Asia Refined analysis version 2 (HAR v2), available at: https://www.klima.tu-berlin.de/HARv2, last access: 6 July 2021.

UOS: Global Fatal Landslide Database (Version 2; 2004 to 2017), available at: https://shefuni.maps.arcgis.com/apps/webappviewer/index.html?id=8458951270904fc29527254492517063, last access: 6 July 2021.

Wang, Q., Wang, D., Huang, Y., Wang, Z., Zhang, L., Guo, Q., Chen, W., Chen, W., and Sang, M.: Landslide susceptibility mapping based on selected optimal combination of landslide predisposing factors in a large catchment, Sustainability, 7, 16653–16669, 2015.

Wang, X., Tolksdorf, V., Otto, M., and Scherer, D.: WRF-based Dynamical Downscaling of ERA5 Reanalysis Data for High Mountain Asia: Towards a New Version of the High Asia Refined Analysis, Int. J. Climatol., 41, 743–762, 2021.

Wieczorek, G. F.: Landslides: investigation and mitigation, in: chap. 4 – Landslide triggering mechanisms, Transportation Research Board Special Report, Transportation Research Board, Washington, DC, 1996.

Woodcock, F.: The evaluation of yes/no forecasts for scientific and administrative purposes, Mon. Weather Rev., 104, 1209–1214, 1976.

Zhou, X., Yang, K., Ouyang, L., Wang, Y., Jiang, Y., Li, X., Chen, D., and Prein, A.: Added value of kilometer-scale modeling over the third pole region: a CORDEX-CPTP pilot study, Clim. Dynam., 1–15, 2021.

Zhuo, L., Dai, Q., Han, D., Chen, N., and Zhao, B.: Assessment of simulated soil moisture from WRF Noah, Noah-MP, and CLM land surface schemes for landslide hazard application, Hydrol. Earth Syst. Sci., 23, 4199–4218, https://doi.org/10.5194/hess-23-4199-2019, 2019.

Paper IV

Sensitivity of water balance in the Qaidam Basin to the mid-Pliocene climate

Wang, X.[1], Schmidt, B.[1], Otto, M., Ehlers, A.T., Mutz, S., Botsyun, S. and Scherer, D., 2021: Sensitivity of Water Balance in the Qaidam Basin to the Mid-Pliocene Climate. *J. Geophys. Res. Atmos.*, 126(16), https://doi.org/10.1029/2020JD033965.

Status: Published.

Own contribution:

- design of the study (with co-authors)
- numerical simulations (with co-authors)
- data preparation (with co-authors)
- data analysis and visualization (with co-authors)
- writing and revision of manuscript (with co-authors)

[1]Schmidt, B. and Wang, X. contributed equally to this work

JGR Atmospheres

RESEARCH ARTICLE
10.1029/2020JD033965

X. Wang and B. Schmidt should be considered joint first author.

Key Points:
- ECHAM5 global climate simulations for the present day and the mid-Pliocene were dynamically downscaled over the Qaidam Basin (QB)
- Results show a positive imbalance in water balance when the mid-Pliocene climate is imposed on the QB with its modern surface settings
- This imbalance in water balance is closely associated with changes in the midlatitude westerlies and in the East Asian Summer Monsoon

Supporting Information:
Supporting Information may be found in the online version of this article.

Correspondence to:
X. Wang and B. Schmidt,
xun.wang@tu-berlin.de;
benjamin.schmidt@tu-berlin.de

Citation:
Wang, X., Schmidt, B., Otto, M., Ehlers, T. A., Mutz, S. G., Botsyun, S., & Scherer, D. (2021). Sensitivity of water balance in the Qaidam Basin to the mid-Pliocene climate. *Journal of Geophysical Research: Atmospheres*, *126*, e2020JD033965. https://doi.org/10.1029/2020JD033965

Received 25 SEP 2020
Accepted 14 JUN 2021

Sensitivity of Water Balance in the Qaidam Basin to the Mid-Pliocene Climate

Xun Wang[1], Benjamin Schmidt[1], Marco Otto[1], Todd A. Ehlers[2], Sebastian G. Mutz[2], Svetlana Botsyun[2], and Dieter Scherer[1]

[1]Chair of Climatology, Technische Universität Berlin, Berlin, Germany, [2]Department of Geosciences, University Tübingen, Tübingen, Germany

Abstract The Qaidam Basin (QB) in the northeastern Tibetan Plateau held a megalake system during the Pliocene. Today, the lower elevations in the basin are hyperarid. To understand to what extent the climate plays a role in the maintenance of the megalake system during the Pliocene, we applied the Weather Research and Forecasting model for dynamical downscaling of ECHAM5 global climate simulations for the present day and the mid-Pliocene. When imposing the mid-Pliocene climate on the QB with its modern land surface settings, the annual water balance (ΔS), that is, the change in terrestrial water storage within the QB, increases. This positive imbalance of ΔS induced solely by the changes in the large-scale climate state would lead to a readjustment of lake extent, until a new equilibrium state is reached, where loss due to evaporation over lake areas compensates for the input by runoff and precipitation. Atmospheric water transport (AWT) analysis at each border of the QB reveals that this imbalance of ΔS is caused by stronger moisture influx across the western border in winter, spring, and autumn and weaker moisture out-flux across the eastern border in summer. These changes in AWT are associated with the strengthening of the midlatitude westerlies in all seasons, except for summer, and the intensification of the East Asian Summer Monsoon. Given that the mid-Pliocene climate is an analog to the projected warm climate of the near future, our study contributes to a better understanding of climate change impacts in central Asia.

1. Introduction

The Qaidam Basin (QB) is an intermontane endorheic drainage basin located in the northeastern Tibetan Plateau (TP) (Figure 1). The central and lower elevation part of the QB is hyperarid today, but paleogeographic studies revealed that the QB contained a freshwater megalake system during the Pliocene (Chen & Bowler, 1986; Mischke et al., 2010; J. Wang et al., 2012) even though the basin and surrounding mountain areas experienced general aridification throughout the Pliocene (Y. F. Miao et al., 2013; Rieser et al., 2009). With the beginning of the Pleistocene, the megalake system began to shrink. This process continued throughout the Pleistocene until today when only a few playas and saline lakes remain (J. Wang et al., 2012). In the mid-Pliocene (~3 Ma), the global climate was warmer and wetter (Ravelo et al., 2004), while paleogeographic features were similar to those of today (Dowsett et al., 2010). The principal objective of this study is to investigate to what extent climate plays a role in the maintenance of the Qaidam megalake system during the mid-Pliocene.

In an endorheic basin, the steady-state water balance (ΔS) is zero (Broecker, 2010; Ibarra et al., 2018). ΔS is defined as the total change in terrestrial water storage (TWS) within all the reservoirs inside the basin. However, changes in climate state can alter the TWS and lead to an imbalanced ΔS (e.g., Jiao et al., 2015; Y. Li et al., 2019; J. Wang et al., 2018). A recent study by Scherer (2020) revealed that under the present climate conditions, the ΔS in the QB is close to zero, and specific humidity is the main climate driver of the annual ΔS. He found that the annual ΔS positively correlates with the annual specific humidity. Based on these findings, we hypothesize that a wetter climate state like that of the mid-Pliocene would cause a positive imbalance of the ΔS in the QB, which would lead to recharging of groundwater reservoirs and eventually to a rising lake level and an extension of the lake area in the QB. This readjustment of lake extent would continue until a new equilibrium state is reached, where loss due to evaporation over lake areas compensates for the input by runoff and precipitation.

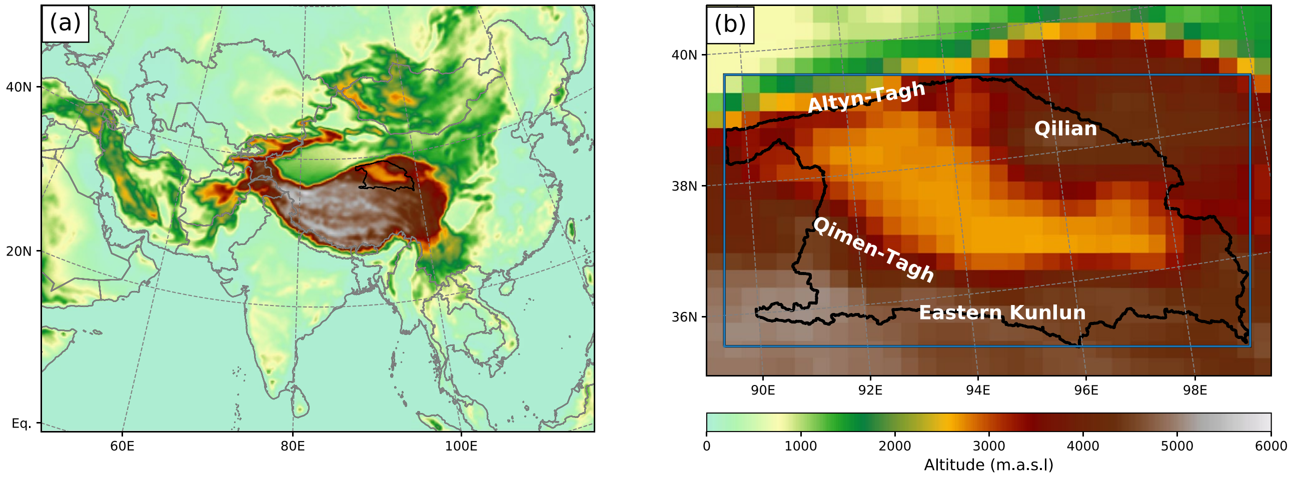

Figure 1. (a) Map of Weather Research and Forecasting (WRF) model domain and (b) overview of the Qaidam Basin. Black line: boundary of the Qaidam Basin (Lehner & Grill, 2013). Blue rectangle: Qaidam box for atmospheric moisture budget analysis.

By imposing the mid-Pliocene climate state on the QB with its modern land surface settings (without a megalake system), an imbalance of ΔS is induced. From that, we can estimate the lake extent needed to remove this imbalance (hereinafter referred to as the equilibrium lake extent). This can give us an insight into the influence of climate on the existence of the megalake system during the mid-Pliocene. To achieve this, we conducted two regional climate simulations to examine the sensitivity of ΔS to changes in large-scale climate state. These two simulations were driven by present-day and mid-Pliocene global climate simulations, respectively. The intention of this study is not to reconstruct the regional climate in the QB during the mid-Pliocene. Instead, we simulated the regional climate of the QB and its surrounding areas with modern surface conditions. In this way, we can analyze the large-scale controls of ΔS in the QB independently from other land surface controls.

Scherer (2020) revealed the regional-scale climate driver of the ΔS. However, the large-scale controlling mechanisms of ΔS in the QB are still not fully understood. On the climatological scale, ΔS and net moisture transport into and out of an endorheic basin through its lateral boundary are in balance (Brubaker et al., 1993). Consequently, the changes in ΔS in the QB are linked to the changes in atmospheric water transport (AWT). The mid-Pliocene climate is considered as an analog of near-future climates (Burke et al., 2018). Thus, investigating the response of ΔS to the mid-Pliocene climate can help us understand the lake development in the future and contribute to water resource management in central Asia.

The goal of this study is to answer the following research questions:

1. How large is the imbalance in ΔS in the QB if the mid-Pliocene climate is imposed on the QB with its modern land surface settings? What equilibrium lake extent would be needed to remove this imbalance?
2. How does the AWT differ from the mid-Pliocene climate to the present-day climate?
3. Which large-scale systems regulate the changes in AWT?

The paper is organized as follows: we describe the methods used in this study in the following section. Section 3 presents the dynamical downscaling results for differences in ΔS, AWT, and large-scale circulation patterns between the mid-Pliocene and the present climate. Results are discussed and compared with other studies in Section 4. Conclusions are drawn in Section 5.

2. Data and Methods

2.1. Global Climate Simulations

For the global climate simulations, we used isotope tracking ECHAM5-wiso atmospheric General Circulation Model (GCM), which is developed at Alfred Wegener Institute and based on the ECHAM5 model of the Max Planck Institute for Meteorology, Hamburg (Roeckner et al., 2003). The model is well established and included in the Coupled Model Intercomparison Projects (Meehl et al., 2007; Taylor et al., 2012). The ability of ECHAM5-wiso to reproduce modern and paleoclimates on both global and regional scales has been shown in multiple studies (Botsyun et al., 2020; Mutz et al., 2016, 2018).

We performed ECHAM5-wiso simulations at a T159 spectral resolution (equivalent to a grid spacing of ~0.75°) with a vertical resolution of L31 (31 levels up to 10 hPa). The control simulation (PD_GCM) used present-day boundary conditions including the AMIP2 sea-surface temperature and sea ice data from 1957 to 2014 and observed greenhouse gas concentrations for the same period (Nakicenovic et al., 1990). The simulation was conducted for more than 40 model years. A climatological reference period of 15 years was established for the analysis presented here using the simulation years 2000–2014 to represent the most recent climate conditions. We also performed a paleoclimate experiment, representing the climate conditions of the mid-Pliocene (PLIO_GCM, ~3 Ma). The setups and boundary conditions of PLIO_GCM for ECHAM5 are identical to those of Mutz et al. (2018) and Botsyun et al. (2020). In PLIO_GCM, we accounted for changing pCO_2, land surface conditions including vegetation change and land ice, albedo, orbital variation, and sea-surface temperatures, which potentially cause changes in the hydrological cycle. The PLIO_GCM experiment was conducted for 18 years, including 3 years necessary for model spin-up. Both PD_GCM and PLIO_GCM experiments were validated against observed and modeled climate patterns (Botsyun et al., 2020; Mutz et al., 2018).

2.2. Dynamical Downscaling

We employed Weather Research and Forecasting (WRF) model version 4.1.2 (Skamarock et al., 2019) as the Regional Climate Model for the dynamical downscaling of GCM data. WRF is a fully compressible nonhydrostatic model. Two sensitivity experiments were conducted using 15-year time slices from PD_GCM and PLIO_GCM simulated by ECHAM5 (Section 2.1) as initial and boundary conditions. These two WRF experiments are referred to as PD and PLIO in the following text. Except for the atmospheric forcing data, other parameters were kept the same in both experiments (Table 1).

We set the model domain's (Figure 1) grid spacing to 30 km. In the vertical direction, 28 terrain-following eta-levels were used. The model time steps are 120 s with a 6 hourly data output. The boundary conditions were updated every 6 h. We employed the daily reinitialization strategy from Maussion et al. (2011, 2014), where we initialize a model run for every 24 h period. Each simulation starts at 12:00 UTC and contains 36 h, with the first 12 h as the spin-up time. This strategy kept the large-scale circulation patterns simulated by WRF closely constrained by the forcing data, while concurrently allowing WRF to develop the mesoscale atmospheric features. Physical parameterization schemes were consistent with the ones used for high-resolution dynamical downscaling in High Mountain Asia in X. Wang et al. (2021).

To examine the ability of the model to reproduce the present-day climate patterns, PD predicted climate is compared with ERA5 reanalysis data of the European Centre for Medium-Range Weather Forecasts (Copernicus Climate Change Service (C3S), 2017) with regard to precipitation (P) and AWT (Figure S1). PD generally reproduces the spatial patterns of P and AWT. But there exist some discrepancies in the amount of P and AWT.

2.3. Data Analysis

2.3.1. Water Balance

In this study, ΔS is defined as the total change in TWS within the basin's reservoirs. For endorheic drainage basins like the QB, surface runoff across the basin's border is zero by definition and groundwater runoff can be neglected. The QB has been a closed system at least since the Oligocene (Herb et al., 2015; J. Wang et al., 2012). Thus, the ΔS of the QB can be expressed as the spatial average of net precipitation ($P - ET$), that is, the difference between P and evapotranspiration (ET), over the total area of the QB:

$$\Delta S = \left\langle P - ET \right\rangle \quad (1)$$

Angle brackets indicate a spatial average over the whole area of the QB (black line in Figure 1b).

Table 1
Basic WRF Model Configurations

	Dynamics
Dynamical solver	Advanced Research WRF (ARW), nonhydrostatic
	Maps and grids
Map projection	Lambert conformal conic
Horizontal grid spacing	30 km (281 × 217 grid points)
Vertical levels	28 Eta-level
	Forcing strategy
Forcing data	ECHAM5 time slices PD_GCM and PLIO_GCM
Lake surface temperature	Substituted by daily mean surface air temperature
Initialization	Daily
Runs starting time	Daily at 12:00 UTC
Runs duration	36 h
Spin-up time	12 h
	Physical parameterization schemes
Longwave radiation	RRTM scheme (Mlawer et al., 1997)
Shortwave radiation	Dudhia scheme (Dudhia, 1989)
Cumulus	Kain-Fritsch cumulus potential scheme (Berg et al., 2013)
Microphysics	Morrison 2-moment scheme (Morrison et al., 2009)
Planetary boundary layer	Yonsei University scheme (Hong et al., 2006)
Land surface model	Unified Noah land surface model (Tewari et al., 2004)
Surface layer	Revised MM5 surface layer scheme (Jiménez et al., 2012)

Note. Additional model information can be found in namelists attached in supporting information.

2.3.2. Equilibrium Lake Extent Analysis

Several previous studies applied a lake mass balance equation to estimate the equilibrium lake extent (Broecker, 2010; Ibarra et al., 2018). The basic concept is that (a) in the equilibrium state of an endorheic basin, water that feeds the basin's lake (precipitation and runoff from land areas) is in balance with lake evaporation and (b) runoff is generated over land areas, where $P - ET$ is positive and accumulates in the low-altitude parts of the basin to form or feed lakes:

$$(P_{land} - ET_{land})(A_{QB} - A_{lake}) + P_{lake}A_{lake} = ET_{lake}A_{lake} \tag{2}$$

where A refers to area and subscripts *land*, *lake*, and *QB* denote corresponding quantities over land, lake, and the whole basin. Rearranging Equation 2 yields:

$$A_{lake} = \frac{P_{land} - ET_{land}}{P_{land} - ET_{land} + ET_{lake} - P_{lake}} A_{QB} \tag{3}$$

Since our PD and PLIO simulations do not contain a lake in the QB, the difference of ΔS between PLIO and PD can be considered as the change in $P_{land} - ET_{land}$. Thus, by adding the change signal of ΔS to the present-day $P_{land} - ET_{land}$ and assuming constant values of E_{lake} and P_{lake}, the equilibrium lake extent that is needed to remove the imbalance in ΔS can be estimated.

The present-day $P_{land} - ET_{land} = 2.4$ mm a^{-1} and $P_{lake} = 40$ mm a^{-1} are derived from the High Asia Refined (HAR) analysis 10 km products (Maussion et al., 2011, 2014). The HAR is an atmospheric data set generated by dynamical downscaling using WRF. It was comprehensively validated and analyzed (D. Li et al., 2020; Maussion et al., 2014; Pritchard et al., 2019). We calculated three projections with different ET_{lake}: 600, 800, and 1,000 mm a^{-1}. These three values represent lower, medium, and upper estimations of lake evaporation rates over the TP based on previous studies (Haginoya et al., 2009; Lazhu et al., 2016; X. Li et al., 2016; J. Xu et al., 2009; S. Yu et al., 2011).

2.3.3. Atmospheric Water Transport

Following Curio et al. (2015), AWT is calculated as the vertical integration of water flux over the whole atmospheric column along the model eta-levels from surface (z_{sfc}) to top (z_{top}):

$$Q = \int_{z=z_{sfc}}^{z_{top}} v_h \rho q \Delta z \tag{4}$$

where v_h is the horizontal wind vector, ρ is the dry air density, q is the specific humidity for all water species, which is converted from the mixing ratio of water vapor, liquid water, and solid water, and Δz is the thickness of each eta-level, which changes over time and increases with height.

The shape of the QB is very irregular (black line in Figure 1b). A simple rectangle covering the QB (hereafter referred to as Qaidam box) was defined to perform budget analysis on the AWT across the four borders (blue rectangle in Figure 1). This method is widely used to estimate the moisture input and output within a certain area (e.g., Feng & Zhou, 2012; Koffi et al., 2013; Z. Wang et al., 2017). The AWT at each border was calculated and converted to the theoretical precipitation amount. The atmospheric moisture budget of the Qaidam box was then calculated as the sum of the AWT at all borders.

2.3.4. Statistical Method

We applied two-sided Welch's t-tests to assess the uncertainty in the change signal of 15-year means between PD and PLIO. We present p-values in the form of maps and tables to give a transparent and open assessment of the uncertainty in the change signal without dichotomous treatment of it, following Wasserstein and Lazar (2016). The interpretation of a large number of local hypothesis tests requires a conservative and critical interpretation of the uncertainty of each local test. Thus, the p-values presented in these maps are adjusted using the false discovery rate adjustment approach by Yekutieli and Benjamini (1999). This method allows a more realistic representation of uncertainty in a field of correlated tests.

3. Results

3.1. Comparison of ΔS and its Components

In this section, we focus on the changes in ΔS and its components in the High Mountain Asia region and the QB. Figure 2 shows the downscaling results for P, ET, and $P - ET$ for PLIO and PD and the difference between the two. Maps of adjusted p-values from two-sided Welch's t-tests are presented in Figure S2 to show the uncertainty in the change signal of 15-year means between PD and PLIO. Table 2 presents seasonal and annual values of P, ET, ΔS, air temperature at 2 m ($T2$), and specific humidity at 2 m ($Q2$) averaged over the QB (black line in Figure 1). The associated p-values are presented in Table S1.

Over the High Mountain Asia region, simulated P is enhanced in PLIO over the Himalayas, the northern TP, the Tarim Basin, and the Tien Shan. Lower P in PLIO can be found in the central TP and Pamir–Karakoram (Figure 2c). In the QB, P strongly correlates with the altitude for both simulations, due to orographically induced precipitation (Figures 2a and 2b). PLIO generally has higher values of P in the QB, and the largest difference can be found in the Qilian Mountains, the Altyn-Tagh Mountains, and the Qimen-Tagh ranges. However, the change signal shows large uncertainty on the effect in the Eastern Kunlun Mountains (Figure S2a), which are located in the transition area of positive and negative values of P difference between PLIO and PD (Figure 2c). Averaged over the whole QB, we see increased P in all seasons in PLIO with an annual difference between PLIO and PD of 63 mm a^{-1} (Table 2).

The spatial patterns of the differences in ET between PLIO and PD (Figure 2f) generally follow those of the differences in P (Figure 2c) in large parts of the domain. In these regions, it is not the availability of energy for latent heat, but water availability that limits ET. We find a different situation in the Pamir–Karakoram region, the northern slopes of the central Himalayas, and the south-eastern part of the QB. This opposite change of P and ET in the above-mentioned regions indicates that energy availability must be the limiting factor for ET. In the QB, ET mainly takes place in the mountain areas, where most of the precipitation

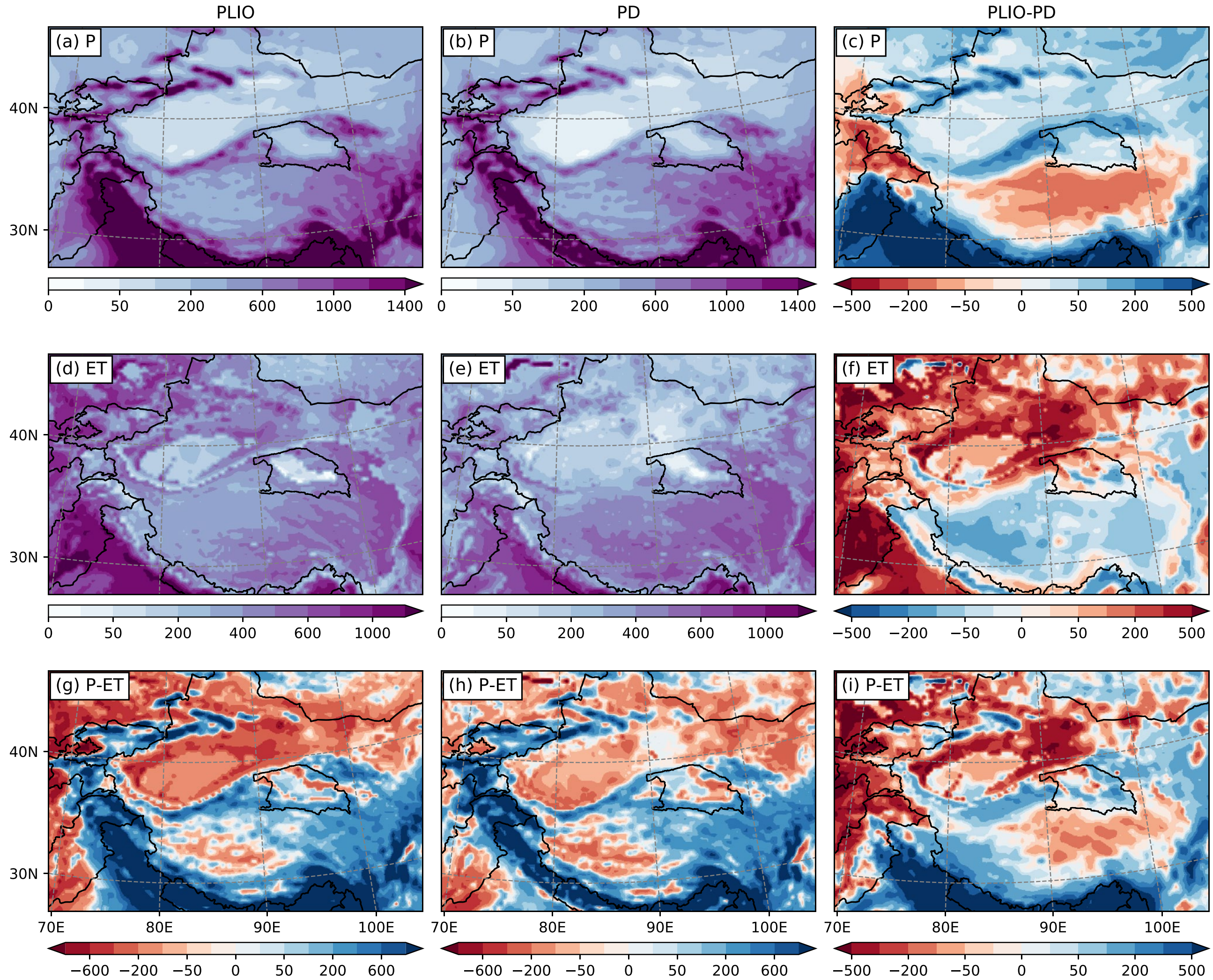

Figure 2. Fifteen-year average of annual precipitation P (a–c), evapotranspiration ET (d–f), and net precipitation $P - ET$ (g–i) for PLIO, PD, and difference between PLIO and PD. The unit for all subplots is mm a^{-1}. p-values from two-sided Welch's t-test for difference of means between PLIO and PD are presented in Figure S2.

occurs (Figures 2d and 2e). Averaged over the whole QB, PLIO shows an increase in ET of 36 mm a^{-1} compared to PD.

For $P - ET$, PLIO and PD yield similar spatial patterns over High Mountain Asia, with negative $P - ET$ over the western TP, the Tarim Basin, and some parts of the QB (Figures 2g and 2h). In PLIO, $P - ET$ is decreased over the central TP, the Tarim Basin, and Pamir–Karakoram. The annual ΔS in the QB as a whole is 26 mm a^{-1} higher in PLIO (Table 2) with a p-value of 0.058 (Table S1), which is higher than the p-values of the other quantities. This suggests that the observed effect is real but more research on the topic should be conducted to lower the level of uncertainty. Spatially, a higher ΔS in PLIO can be found in the eastern and central parts of the QB (Figure 2i). Seasonally, PLIO has a higher ΔS in the QB in winter, spring, and autumn. In summer, the difference in ΔS between PLIO and PD is negative (Table 2).

The global climate in the mid-Pliocene is believed to be warmer and wetter than the modern climate, which is also shown in our ECHAM5 global simulations. However, this is not true for the regional climate signal in the QB. While the local $Q2$ signal averaged over the QB is consistent with the global signal, the local $T2$ in the QB is 2 K lower in PLIO (Table 2).

Table 2
Fifteen-Year Average of Seasonal and Annual Precipitation P (mm Season^{-1} or mm a^{-1}), Evapotranspiration ET (mm Season^{-1} or mm a^{-1}), Water Balance ΔS (mm Season^{-1} or mm a^{-1}), air Temperature at 2 m T2 (°C), and Specific Humidity at 2 m Q2 (g kg^{-1}) Averaged Over the Qaidam Basin for PLIO, PD, and the Difference Between PLIO and PD (PLIO − PD)

	PLIO					PD					PLIO − PD				
	P	*ET*	Δ*S*	*T*2	*Q*2	*P*	*ET*	Δ*S*	*T*2	*Q*2	*P*	*ET*	Δ*S*	*T*2	*Q*2
DJF	51	22	29	−13.9	1.3	43	25	19	−12.2	1.3	8	−3	11	−1.7	0.0
MAM	131	87	43	−2.7	3.4	111	83	27	0.0	3.1	20	4	17	−2.7	0.3
JJA	163	152	12	8.7	7.4	151	122	29	9.9	6.0	13	30	−17	−1.2	1.4
SON	74	62	12	−3.3	3.3	52	56	−4	−1.0	3.0	22	6	16	−2.3	0.3
Annual	419	322	96	−2.8	3.9	356	286	70	−0.8	3.4	63	36	26	−2.0	0.5

Note. *p*-values from two-sided Welch's *t*-tests for difference of means between PLIO and PD can be found in Table S1.

3.2. Lake Extent Analysis

In Figure 3, we summarize lake extents (A_{lake}), lake levels (z_{lake}), and the rise in lake levels (Δz) of three equilibrium lake states under the increase in mean annual ΔS of 26 mm a^{-1} and three different estimations of ET_{lake}. Figure 3 also illustrates these three projected states, where lake extents are estimated using WRF model topography for accumulations of $P - ET$ and subsequent runoff originating from land in grid points at or below equilibrium lake levels. An increase in mean annual ΔS of 26 mm a^{-1} would be sufficient to sustain a lake in the QB with an extent ranging from 7,298 to 12,260 km^2.

3.3. Comparison of AWT

Since ΔS is linked to the large-scale AWT, in this section, we compare AWT between PLIO and PD. Table 3 presents the seasonal and annual AWT through each border of the Qaidam box (blue rectangle in Figure 1b), as well as the sum of AWT from all borders, that is, the atmospheric moisture budget. In both PLIO and PD, the western and eastern borders of the Qaidam box serve as the dominant moisture input and output channel, respectively. The higher annual moisture budget in PLIO derives from the increased moisture influx across the western border and the decreased moisture export at the eastern border (Table 3). In PLIO, the increased moisture influx at the western border occurs in winter, spring, and autumn, while a strong reduction of moisture export at the eastern border occurs in summer. This indicates that moisture budget and ΔS in the QB is related to the large-scale systems that influence the AWT at the western and eastern borders.

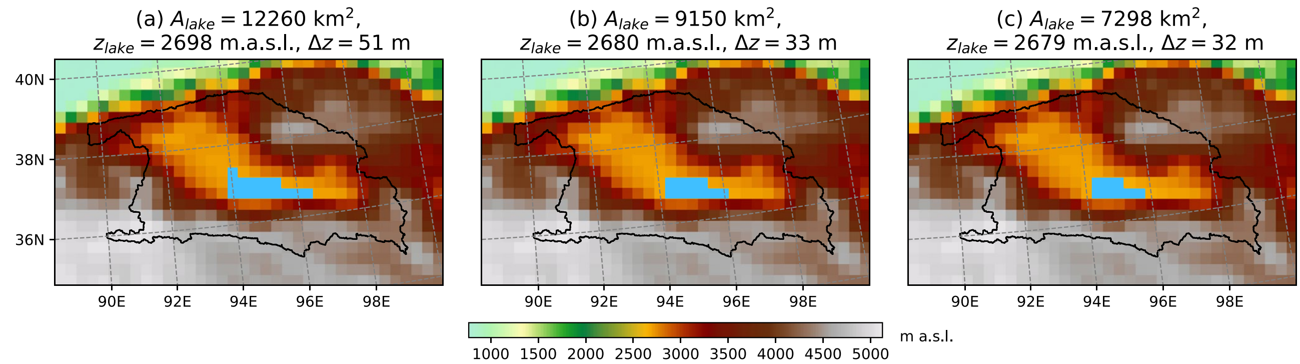

Figure 3. Illustrations of simulated lake extents of equilibrium lake states using the present-day model topography from WRF for accumulation of net precipitation in the Qaidam Basin (QB) and subsequent runoff originating from land in areas at or below equilibrium lake levels (marked in blue). The equilibrium lake extent (A_{lake}), lake level (z_{lake}), and the rise in lake level (Δz) were estimated applying the method described in Section 2.3 using 26 mm a^{-1} as input change in mean annual ΔS and (a) 600 mm a^{-1}, (b) 800 mm a^{-1}, and (c) 1,000 mm a^{-1} as input lake evaporation (E_{lake}).

Table 3
Fifteen-Year Average of Seasonal and Annual Atmospheric Water Flux Converted to Theoretical Precipitation Amount (mm Season^{-1} or mm a^{-1}) Through Each Border of the Qaidam Box (Blue Rectangle in Figure 1b)

	PLIO					PD					PLIO − PD				
	West	East	South	North	Sum	West	East	South	North	Sum	West	East	South	North	Sum
DJF	327	−298	−38	28	19	277	−248	−7	−2	19	50	−50	−31	30	0
MAM	433	−399	50	−49	36	390	−368	19	−14	28	43	−31	31	−35	8
JJA	344	−283	120	−174	8	407	−459	159	−95	12	−63	176	−39	−79	−4
SON	415	−414	71	−79	−7	368	−394	95	−80	−11	47	−20	−24	1	4
Annual	1,519	−1,393	204	−274	56	1,441	−1,469	267	−191	48	78	76	−63	−83	8

Note. Positive values indicate moisture input into the Qaidam Basin, while negative values represent moisture output. p-values from two-sided Welch's t-tests for difference of means between PLIO and PD can be found in Table S2.

Figure 4 illustrates the route and the magnitude of seasonal moisture propagation across the whole domain. In both PD and PLIO, the midlatitude westerlies control the AWT in the QB throughout the year. In winter, spring, and autumn, there exists an eastward moisture transport in the difference plots (Figures 4c, 4f, and 4l), indicating a stronger moisture transport by the midlatitude westerlies in PLIO in these seasons. This is supported by higher moisture input at the western border and higher output at the eastern border in these seasons. In summer months, the influence of the midlatitude westerlies is weaker in PLIO, as shown by the anomalous westward moisture transport over the QB in Figure 4i. Accordingly, the moisture input at the western border and the output at the eastern border are lower in PLIO.

In the summer months, moisture transport is predominantly from the southwest due to the Indian Summer Monsoon (ISM). This northeastward transport over the TP into the QB through the southern border of the Qaidam box is stronger in PD than that in PLIO (Figures 4g–4i and Table 3). Additionally, the moisture transport over eastern Asia by the East Asian Summer Monsoon (EASM) is stronger in PLIO. However, from Figure 4, it is unclear whether the EASM also contributes to the water transport into the QB. Analysis of the daily zonal water transport on the eastern border of the Qaidam box shows moisture input from the eastern border into the QB in PLIO from the middle of July to the beginning of August (Figure 5), which indicates the influence of the EASM in PLIO. This pattern cannot be observed in PD. The additional moisture input by the EASM in PLIO also contributes to the lower values for total moisture output at the eastern border in summer.

3.4. Comparison of Large-Scale Circulation Patterns

The results in Section 3.3 show the higher ΔS and moisture budget in PLIO are caused by differences in the AWT at the western and eastern borders of the Qaidam box. In this section, we examine the large-scale systems that control the AWT across these two borders.

The AWT at both the western and eastern borders is under the influence of the midlatitude westerlies throughout the year. The seasonal zonal wind speed along a latitude–pressure transect across the longitudinal range of the QB (89°E–100°E) for PLIO and PD is presented in Figure 6. In both PLIO and PD, the midlatitude westerlies show seasonal migrations: a southward extension from summer to winter and a northward contraction from winter to summer. The maximum zonal wind speed occurs at around 200 hPa in all seasons. PLIO shows a poleward shifted and contracted westerly zone. Following the methods used by J. Sun et al. (2020), we defined a strength index as the average maximum zonal wind speed at 200 hPa at each longitude over the Qaidam box to quantify the strength of the westerlies over the QB. Under the mid-Pliocene conditions, the westerlies over the QB are stronger in all seasons except for summer (Figure 7), which explains the larger AWT through the western and eastern borders (Table 3).

Figures 8a–8c show the 500 hPa geopotential height and wind field in summer for PLIO and PD. The Northwest Pacific subtropical high (NPSH) intensifies and extends westwards in PLIO, indicated by higher geopotential and stronger anticyclonic circulation over eastern China (Figure 8a). The NPSH is a major

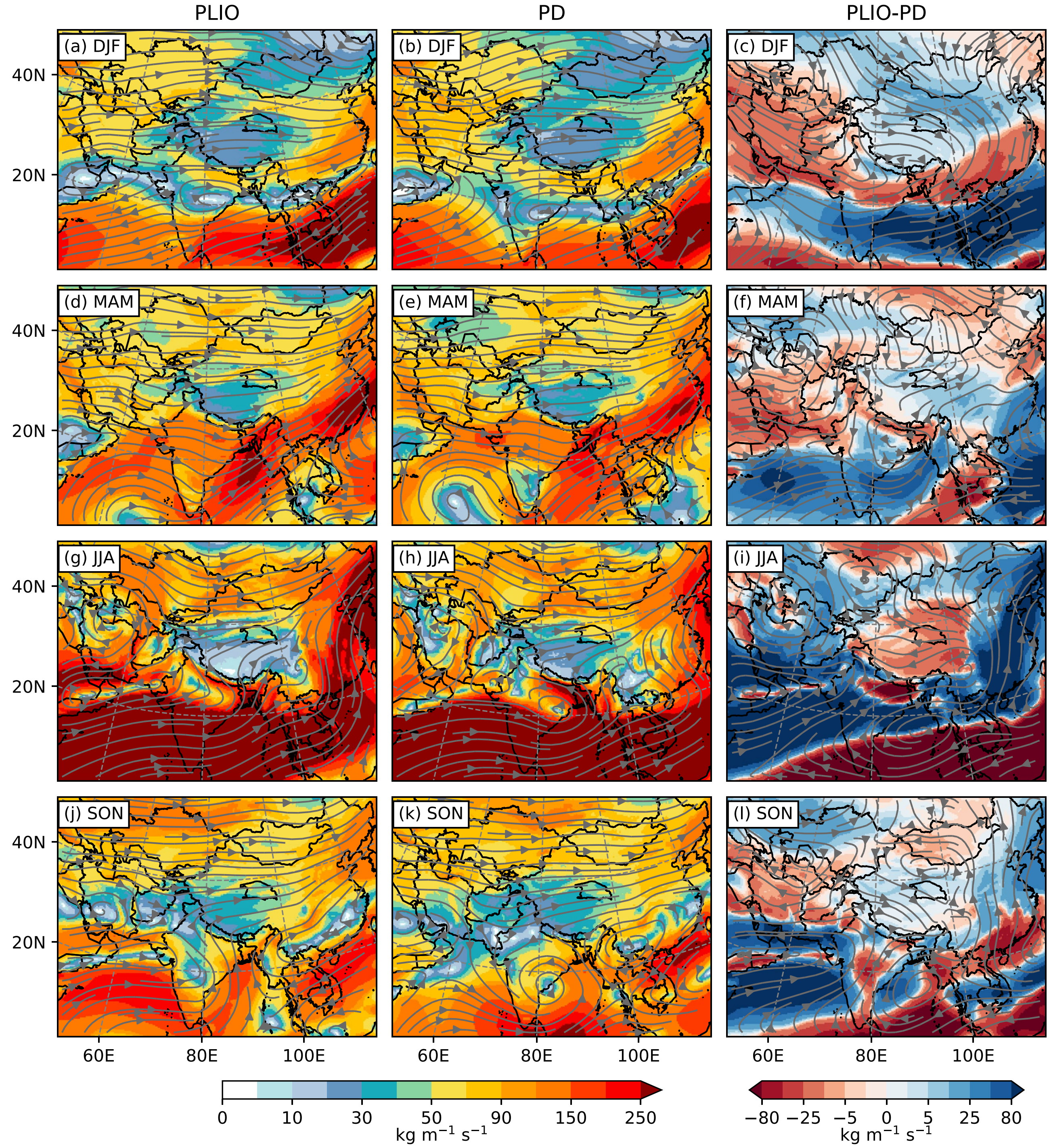

Figure 4. Fifteen-year mean seasonal atmospheric water transport ($kg\,m^{-1}\,s^{-1}$) in DJF (a–c), MAM (d–f), JJA (g–i), and SON (j–l) for PLIO (a, d, g, and j), PD (b, e, h, and k), and the difference between PLIO and PD (c, f, i, and l). Colors represent the strength of water vapor flux; arrows indicate transport direction.

component of the subtropical EASM and regulates the northern edge of the EASM (Huang et al., 2019). The strengthening of the EASM in PLIO is coupled with a tilted jet stream axis over northeastern China. The influence of the EASM on the QB is not visible in the climatology of the summer wind field at the 500 hPa level. Therefore, we selected the period from July 17 to July 27, when the daily AWT averaged over all the grid points at the eastern border of the Qaidam box is negative (westward), to define the period, when the

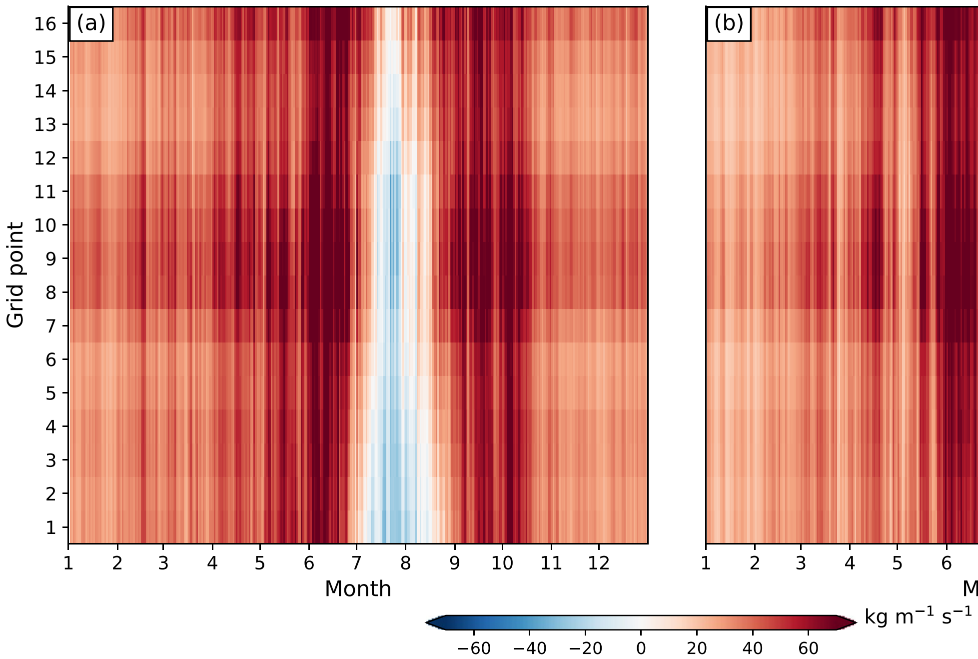

Figure 5. Fifteen-year average of daily zonal atmospheric water transport (kg m^{-1} s^{-1}) of 16 grid poinrts at the eastern border of the Qaidam box (blue rectangle in Figure 1b) for (a) PLIO and (b) PD. Grid point 1 and grid point 16 represent the south-most and north-most grid point. Reddish colors indicate water transport away from the Qaidam Basin, while blueish colors indicate water transport toward the Qaidam Basin.

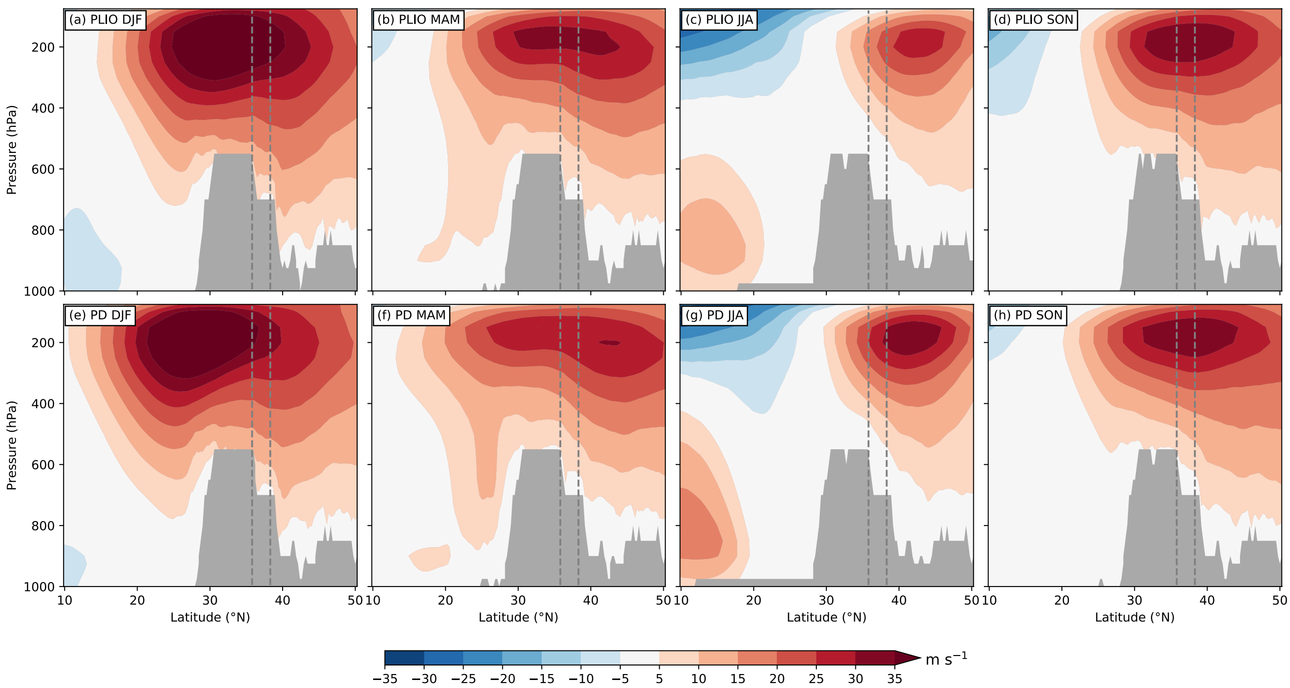

Figure 6. Fifteen-year average of seasonal zonal wind speed (m s^{-1}) along a latitude–pressure transect averaged from 89°E–100°E for PLIO (a–d) and PD (e–h) in DJF (a and e), MAM (b and f), JJA (c and g), and SON (d and h). Dashed lines show the latitudinal range of the Qaidam Basin.

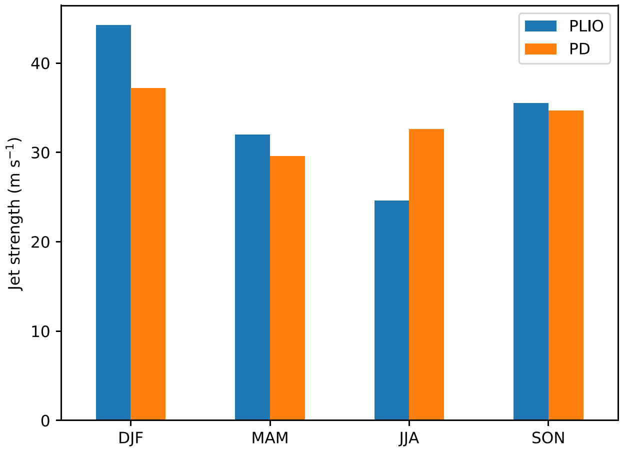

Figure 7. Fifteen-year average of seasonal strength index of jet stream (m s^{-1}) over the Qaidam Basin for PLIO and PD, calculated as defined in J. Sun et al. (2020).

QB is under the direct influence of the EASM. The geopotential height and wind field at 500 hPa in this period are presented in Figures 8d–8f. During this period, the NPSH intensifies and extends further northwestward in both PLIO and PD, as compared to the climatology in the summer. In PLIO, the southern part of the QB is clearly under the control of easterly winds in this period (Figure 8d).

4. Discussion

4.1. Implications of the Higher ΔS in the QB in PLIO

The PLIO simulation presented here cannot be considered as a reconstruction of the regional climate in the QB for the mid-Pliocene. A reconstruction is neither the intention of this study nor is possible at this time at a grid spacing of 30 km. This would require Pliocene surface conditions, such as topography and land cover, at an equivalent or even higher resolution, which are not available to date. The most commonly used surface boundary conditions for mid-Pliocene simulations are from the US Geological Survey's Pliocene Research Interpretation and Synoptic Mapping (PRISM) project (Dowsett et al., 1994). The most recently released PRISM4 data set has a $1^\circ \times 1^\circ$ grid spacing (Dowsett et al., 2016), which is too coarse for regional climate simulations at a grid spacing of 30 km. Moreover, the exact extent and location of the Qaidam megalake system in the mid-Pliocene is unknown. We took only required meteorological fields from the ECHAM5 model to drive the WRF model and applied the same geographical static data from WRF Preprocessing System for both PLIO and PD simulations. This approach has the benefit of isolating the influence of atmospheric variables on the hydroclimate of the QB from other factors, such as land cover and vegetation changes.

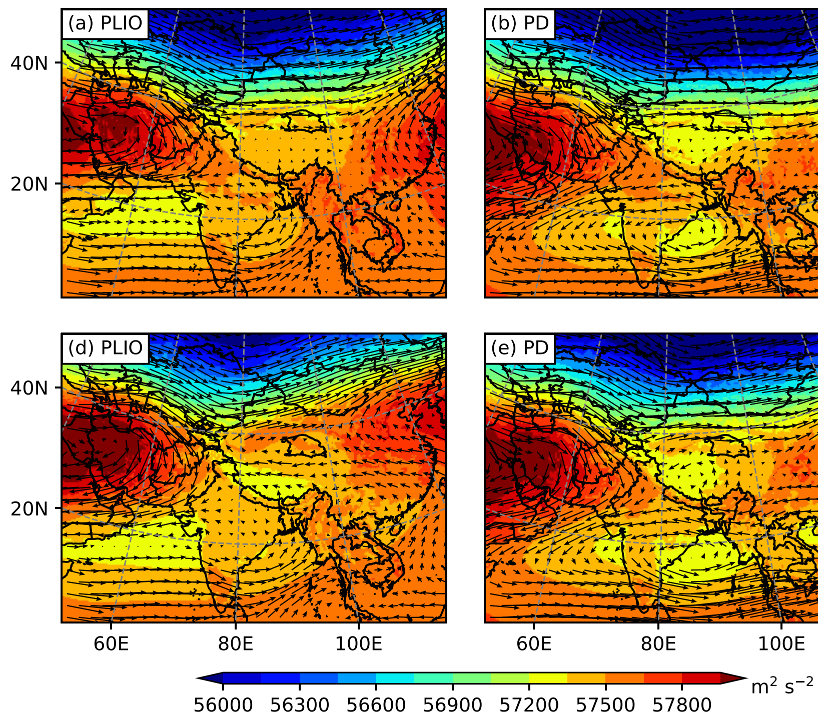

Figure 8. Fifteen-year climatology of geopotential (shading, m^{2}s^{-2}) and wind field (arrows, m s^{-1}) at 500 hPa averaged in JJA (a–c) and from July 17 to July 27 (d–f) for PLIO, PD, and PLIO − PD.

Therefore, the PLIO simulation cannot be interpreted on its own but in conjunction with the PD simulation. Interpretation of the simulation results should focus on the differences between PLIO and PD simulations. The QB has a higher ΔS in PLIO than in PD. This imbalance in ΔS induced solely by the changes in large-scale atmospheric circulations could result in a lake much larger than Lake Qinghai, the largest salt lake in China (4,317 km^2), as shown in our lake extent analysis (Section 3.2). The maximum lake extent of 59,000 km^2 approximated from proxy data (Chen & Bowler, 1986) is still much larger than the estimations we present here. This indicates that factors besides large-scale atmospheric circulation also contributed to sustaining the megalake system during the Pliocene.

4.2. Comparison With Pliocene Model Intercomparison Project Simulations

The changes in large-scale circulation patterns in PLIO are in good agreement with recent modeling studies (Huang et al., 2019; X. Li et al., 2015; Zhang et al., 2013). The results of these studies are based on simulations in the Pliocene Model Intercomparison Project (PlioMIP) using PRISM3D as boundary conditions, which is the same as our forcing ECHAM5 simulation for the mid-Pliocene. X. Li et al. (2015) found a global poleward shift of the midlatitude westerlies and an increase of zonal wind on the poleward flank of the westerly jet in the mid-Pliocene. The poleward shift of the midlatitude westerlies is accompanied by a poleward shift of Hadley and Ferrel cells. The intensification and northwestward extension of the EASM in the mid-Pliocene are also simulated by models in PlioMIP (Huang et al., 2019; Zhang et al., 2013). Huang et al. (2019) compared modeled results to paleontological data from 43 sites throughout China. The northern margin of the EASM indicated by the wet–dry boundary located further northwest in the mid-Pliocene (Figure 4 in Huang et al. [2019]), which is roughly consistent with their modeled results.

In addition, we compared PD and PLIO simulations, and PD_GCM and PLIO_GCM simulations generated by ECHAM5 with COSMOS (Stepanek & Lohmann, 2012) and HadCM3 (Bragg et al., 2012) simulations, in terms of P. These two are the end-members of PlioMIP in simulated, regionally averaged annual precipitation deviations from the preindustrial, as identified by Zhang et al. (2013). As shown in Figure 9, the coarse end-members COSMOS and HadCM3 are not able to produce orographic precipitation patterns correctly, such as those along the Himalayan orogen. Our high-resolution ECHAM5 simulations already show improvements. Nevertheless, the skill of GCMs in predicting orographic precipitation remains limited (e.g., Meehl et al., 2007). This is particularly true at the scale of interest in this study. At coarse grid resolutions, orography is systematically lower than at fine grid resolutions, such that blocking of air masses and orographically induced precipitation are generally underestimated. While the latter effect generally leads to an underestimation of total precipitation, less blocking could act in both directions. Less blocking could result in excessive AWT, which does not necessarily lead to more precipitation since for this atmospheric water to become precipitable, a trigger is needed (Lin et al., 2018). Consequently, the direction in which less blocking changes precipitation remains complicated. In addition, the WRF runs use a nonhydrostatic pressure solver while GCMs assume hydrostatic conditions, which further improves predictions of relatively small-scale phenomena that departures from the hydrostatic balance (Yang et al., 2017). In total, the WRF dynamical downscaling is able to physically resolve mesoscale atmospheric processes that are not included in GCM simulations. Given the aforementioned differences between WRF and GCMs, it is therefore not surprising that mean annual mid-Pliocene precipitation deviation simulated with WRF shows some discrepancies to GCM (Figure 10). Although the big picture of precipitation deviations from WRF is in line with ECHAM5, disagreement can be found over the complex terrains on the TP. WRF exhibits possible improvements, for example, in the eastern Himalayas and over the TP, where the spatial patterns of precipitation deviations are more coherent than in ECHAM5.

4.3. Large-Scale Systems Controlling the ΔS in the QB

There exists a debate on which large-scale system dominates the hydroclimate in the QB in previous studies based on proxy reconstructions. Caves et al. (2015) summarized published data of the spatial distribution of oxygen isotopes in precipitation over High Mountain Asia since the early Eocene (~56 Ma BP). They found that sites at the QB show consistently higher δO_{18} values in the paleo-precipitation than sites in the southern TP, which indicates that westerlies have dominated the moisture supply over the QB since the early Eocene. The authors concluded that the reduction in moisture supply by the westerlies since the early Eocene,

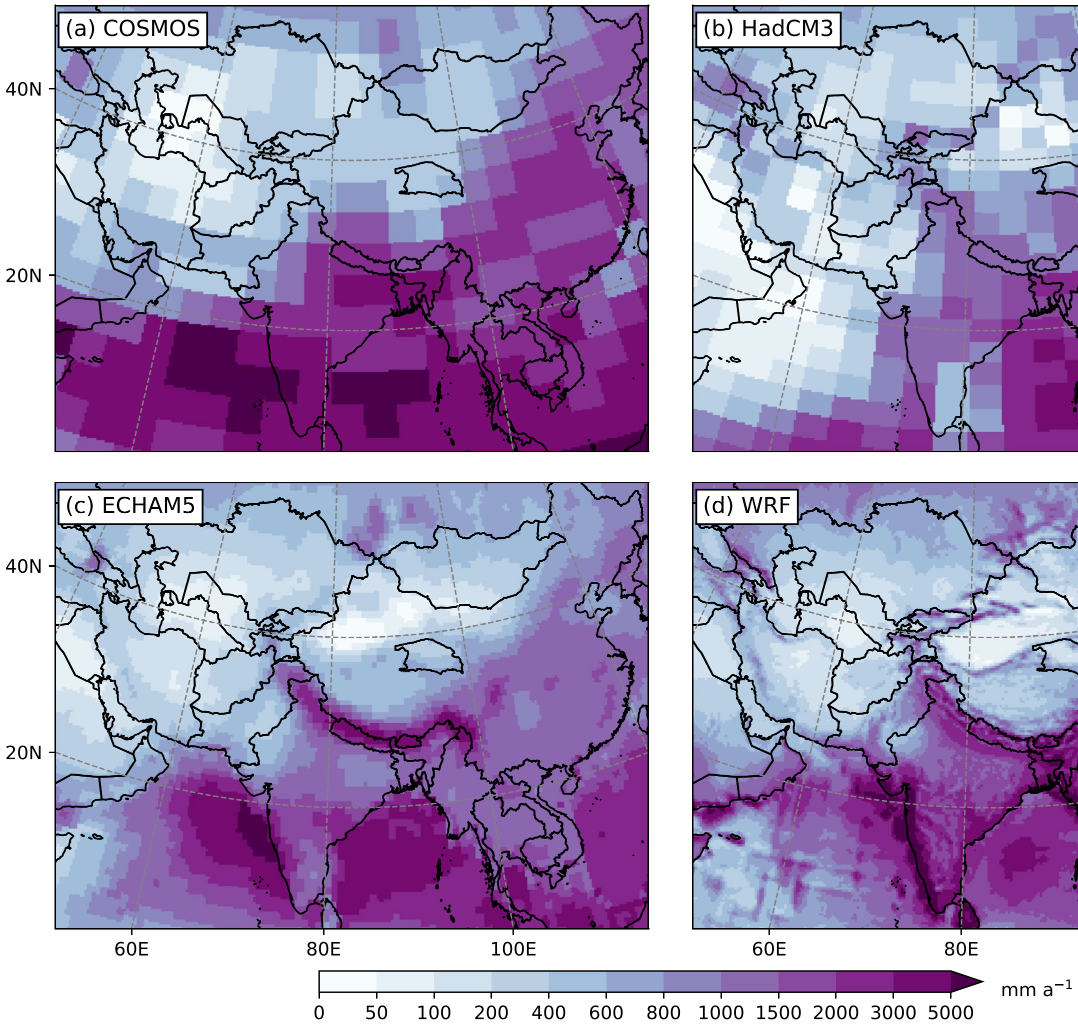

Figure 9. Mean annual mid-Pliocene precipitation climatologies for Pliocene Model Intercomparison Project (PlioMIP) simulations conducted with (a) COSMOS and (b) HadCM3, and for this study's simulations conducted with (c) ECHAM5 and (d) WRF.

rather than uplift of the TP or changes in ISM strength, is the driver of the step-wise drying in central Asia. The modern EASM is believed to have a weak or no influence on the QB (e.g., Huang et al., 2019; L. Yu & Lai, 2012). This is also shown in our PD simulation. There is no westward moisture transport across the eastern border of the Qaidam box in the course of the year (Figure 5b). G. Xu et al. (2011) found correlation between the tree ring δO_{18} series collected from the eastern margin of the QB and the strength of EASM from 1873 to 1975. Y. Miao et al. (2011) examined a pollen record from the KC-1 core, which was extracted in the western part of the QB and covers a period of 18–5 Ma. A transition from a warm–wet climate to a cold–dry climate was found in the QB during this period, and the authors concluded that this transition was driven by the evolution of the East Asian Monsoon system. Our moisture budget analysis shows that the midlatitude westerlies dominate the AWT over the QB in both PLIO and PD (Table 3). However, the difference in the moisture budget in the QB between PLIO and PD is a combined effect of changes in the midlatitude westerlies and the EASM.

4.4. Implications for the Future

The mid-Pliocene warm period is a potential analog of future anthropogenic warming since both periods feature a higher CO_2 concentration than the modern level (Burke et al., 2018). Previous studies reveal that both the mid-Pliocene and projected future climates show similar large-scale atmospheric circulations to

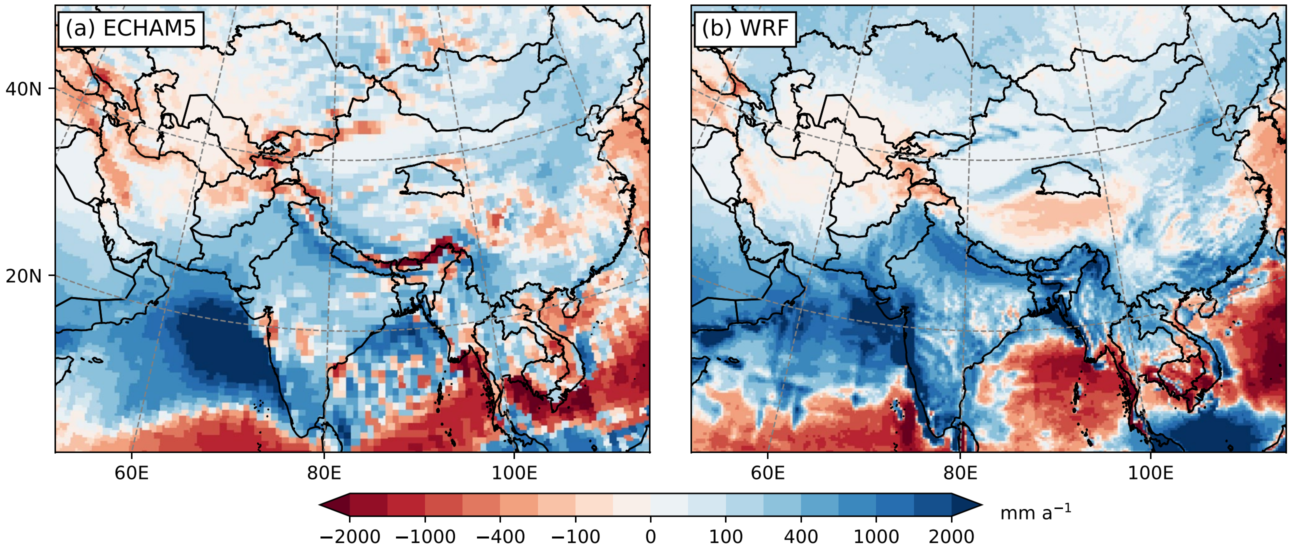

Figure 10. Mean annual mid-Pliocene precipitation deviations for (a) ECHAM5 and (b) WRF from present-day precipitation.

some extent. For example, Y. Sun et al. (2013) applied three Atmosphere–Ocean GCM simulations for Representative Concentration Pathway (RCP) 4.5 scenario, mid-Pliocene, and present day. Their results showed a poleward expansion and an intensification of the Hadley cell over the subtropics under both mid-Pliocene and future climates, but the response of the Walker cell depicts some discrepancies. A conclusion was drawn that the Hadley cell in the mid-Pliocene can be a good analog of the Hadley cell in the future due to the linear relationship between the south-north thermal contrast and the CO_2 concentration. But this analog hypothesis has limitations since the response of the east-west thermal contrast to CO_2 concentration is more complicated and not similar. The monsoon dynamics between the mid-Pliocene climate and the projected future climate under the Extended Concentration Pathway version 4.5, which is an extension of RCP4.5 beyond 2100, were also investigated in a prior study (Y. Sun et al., 2018), who found that both climates show large-scale similarities and an enhanced EASM, which is due to the increase in thermally controlled large-scale moisture transport.

Combined with our results and discussions, the high similarity of large-scale circulations between the mid-Pliocene climate and the projected future climate implies a possible transient state in the QB, during which ΔS will increase, if the EASM extends into the QB and AWT by the westerlies increases. The increased ΔS would lead to the recharge of groundwater reservoirs and subsequent rise of lake levels in the QB until ΔS is balanced again.

5. Conclusions

In this study, we utilized the WRF model for dynamical downscaling of ECHAM5 global simulations for the present day and the mid-Pliocene. The downscaling results show that when imposing the mid-Pliocene climate to the QB with its modern land surface settings, the annual ΔS would increase by 26 mm a^{-1}. This positive imbalance of ΔS induced only by the changes in the large-scale climate state would be sufficient to sustain a lake in the QB with an extent ranging from 7,298 to 12,260 km^2, depending on the value of ET_{lake}.

The annual moisture budget in the QB is higher in PLIO, corresponding with the higher ΔS. Higher moisture input from the western border and lower moisture output at the eastern border are the main reasons for the higher annual moisture budget and higher ΔS in PLIO. These two borders are both under the influence of midlatitude westerlies. The eastern border is additionally regulated by the EASM in PLIO. In PLIO, the midlatitude westerlies contract poleward and intensify over the QB in winter, spring, and autumn, transporting more moisture into the QB through its western border. In summer, the strengthening of the EASM accompanied by weakened westerlies in PLIO leads to the decrease of moisture output at the eastern border.

We conclude that the strengthening of the midlatitude westerlies in all seasons, except for summer, and the intensification of the EASM lead to higher moisture budget and higher ΔS in the QB in PLIO.

The mid-Pliocene climate shares similarities in large-scale circulations with projected future climate scenarios. Thus, our results can contribute to a better understanding of the impacts of climate change and future lake development in central Asia. Our results also highlight the added values of applying high-resolution GCM and dynamical downscaling by WRF. High resolution can resolve fundamental processes over complex terrains, such as orographic precipitation and AWT, more realistically.

Data Availability Statement

The WRF V4.1.2 model used in this study is freely available under the official website of WRF: http://dx.doi.org/10.5065/D6MK6B4K. The setup files for the WRF model are included in the supporting information. The WRF model output data used in this study are available at https://doi.org/10.26050/WDCC/RRA_WRF_simulations.

Acknowledgments
This work was supported by the German Federal Ministry of Education and Research (BMBF) program "Central Asia – Monsoon Dynamics and Geo-Ecosystems II" (CAME II) within Q-TiP project "Quaternary Tipping Points of Lake Systems in the Arid Zone of Central Asia" (code 03G08063C and 03G08063A to D. Scherer and to T. A. Ehlers, respectively). The authors thank Tom Grassmann for his administrative work on the high-performance computing cluster. This open-access publication was funded by Technische Universität Berlin.

References

Berg, L. K., Gustafson, W. I., Kassianov, E. I., & Deng, L. (2013). Evaluation of a modified scheme for shallow convection: Implementation of cup and case studies. *Monthly Weather Review*, *141*(1), 134–147. https://doi.org/10.1175/mwr-d-12-00136.1

Botsyun, S., Ehlers, T., Mutz, S., Methner, K., Krsnik, E., & Mulch, A. (2020). Opportunities and challenges for paleoaltimetry in "small" orogens: Insights from the European Alps. *Geophysical Research Letters*, *47*, e2019GL086046. https://doi.org/10.1029/2019GL086046

Bragg, F., Lunt, D. J., & Haywood, A. (2012). Mid-Pliocene climate modelled using the UK Hadley Centre Model: PlioMIP Experiments 1 and 2. *Geoscientific Model Development*, *5*(5), 1109–1125. https://doi.org/10.5194/gmd-5-1109-2012

Broecker, W. (2010). Long-term water prospects in the western United States. *Journal of Climate*, *23*(24), 6669–6683. https://doi.org/10.1175/2010jcli3780.1

Brubaker, K. L., Entekhabi, D., & Eagleson, P. (1993). Estimation of continental precipitation recycling. *Journal of Climate*, *6*(6), 1077–1089. https://doi.org/10.1175/1520-0442(1993)006<1077:eocpr>2.0.co;2

Burke, K. D., Williams, J. W., Chandler, M. A., Haywood, A. M., Lunt, D. J., & Otto-Bliesner, B. L. (2018). Pliocene and Eocene provide best analogs for near-future climates. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(52), 13288–13293. https://doi.org/10.1073/pnas.1809600115

Caves, J. K., Winnick, M. J., Graham, S. A., Sjostrom, D. J., Mulch, A., & Chamberlain, C. P. (2015). Role of the westerlies in central Asia climate over the Cenozoic. *Earth and Planetary Science Letters*, *428*, 33–43. https://doi.org/10.1016/j.epsl.2015.07.023

Chen, K., & Bowler, J. M. (1986). Late Pleistocene evolution of salt lakes in the Qaidam Basin, Qinghai Province, China. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *54*(1–4), 87–104.

Copernicus Climate Change Service (C3S). (2017). *ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate*. Copernicus Climate Change Service Climate Data Store (CDS). Retrieved from https://cds.climate.copernicus.eu/cdsapp#!/home

Curio, J., Maussion, F., & Scherer, D. (2015). A 12-year high-resolution climatology of atmospheric water transport over the Tibetan Plateau. *Earth System Dynamics*, *6*(1), 109–124. https://doi.org/10.5194/esd-6-109-2015

Dowsett, H., Dolan, A., Rowley, D., Moucha, R., Forte, A. M., Mitrovica, J. X., et al. (2016). The PRISM4 (mid-Piacenzian) paleoenvironmental reconstruction. *Climate of the Past*, *12*(7), 1519–1538. https://doi.org/10.5194/cp-12-1519-2016

Dowsett, H., Robinson, M., Haywood, A., Salzmann, U., Hill, D., Sohl, L., et al. (2010). The PRISM3D paleoenvironmental reconstruction. *Stratigraphy*, *7*(2–3), 123–139.

Dowsett, H., Thompson, R., Barron, J., Cronin, T., Fleming, F., Ishman, S., et al. (1994). Joint investigations of the middle Pliocene climate I: Prism paleoenvironmental reconstructions. *Global and Planetary Change*, *9*(3–4), 169–195. https://doi.org/10.1016/0921-8181(94)90015-9

Dudhia, J. (1989). Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *Journal of the Atmospheric Sciences*, *46*(20), 3077–3107. https://doi.org/10.1175/1520-0469(1989)046<3077:nsocod>2.0.co;2

Feng, L., & Zhou, T. (2012). Water vapor transport for summer precipitation over the Tibetan Plateau: Multidata set analysis. *Journal of Geophysical Research*, *117*, D20114. https://doi.org/10.1029/2011JD017012

Haginoya, S., Fujii, H., Kuwagata, T., Xu, J., Ishigooka, Y., Kang, S., & Zhang, Y. (2009). Air–lake interaction features found in heat and water exchanges over Nam Co on the Tibetan Plateau. *Sola*, *5*, 172–175. https://doi.org/10.2151/sola.2009-044

Herb, C., Koutsodendris, A., Zhang, W., Appel, E., Fang, X., Voigt, S., & Pross, J. (2015). Late Plio-Pleistocene humidity fluctuations in the western Qaidam Basin (NE Tibetan Plateau) revealed by an integrated magnetic-palynological record from lacustrine sediments. *Quaternary Research (United States)*, *84*(3), 457–466. https://doi.org/10.1016/j.yqres.2015.09.009

Hong, S.-Y., Noh, Y., & Dudhia, J. (2006). A new vertical diffusion package with an explicit treatment of entrainment processes. *Monthly Weather Review*, *134*(9), 2318–2341. https://doi.org/10.1175/mwr3199.1

Huang, X., Jiang, D., Dong, X., Yang, S., Su, B., Li, X., et al. (2019). Northwestward migration of the northern edge of the East Asian summer monsoon during the mid-Pliocene warm period: Simulations and reconstructions. *Journal of Geophysical Research: Atmospheres*, *124*, 1392–1404. https://doi.org/10.1029/2018JD028995

Ibarra, D. E., Oster, J. L., Winnick, M. J., Caves Rugenstein, J. K., Byrne, M. P., & Chamberlain, C. P. (2018). Warm and cold wet states in the western United States during the Pliocene–Pleistocene. *Geology*, *46*(4), 355–358. https://doi.org/10.1130/g39962.1

Jiao, J. J., Zhang, X., Liu, Y., & Kuang, X. (2015). Increased water storage in the Qaidam Basin, the North Tibet Plateau from grace gravity data. *PLoS One*, *10*(10), e0141442. https://doi.org/10.1371/journal.pone.0141442

Jiménez, P. A., Dudhia, J., González-Rouco, J. F., Navarro, J., Montávez, J. P., & García-Bustamante, E. (2012). A revised scheme for the WRF surface layer formulation. *Monthly Weather Review*, *140*(3), 898–918. https://doi.org/10.1175/mwr-d-11-00056.1

Koffi, E., Graham, E., & Mätzler, C. (2013). The water vapour flux above Switzerland and its role in the August 2005 extreme precipitation and flooding. *Meteorologische Zeitschrift*, *22*(3), 328–341. https://doi.org/10.1127/0941-2948/2013/0392

Lazhu, Yang, K., Wang, J., Lei, J., Chen, Y., Zhu, Y., et al. (2016). Quantifying evaporation and its decadal change for Lake Nam Co, central Tibetan Plateau. *Journal of Geophysical Research: Atmospheres*, *121*, 7578–7591. https://doi.org/10.1002/2015JD024523

Lehner, B., & Grill, G. (2013). Global river hydrography and network routing: Baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, *27*(15), 2171–2186. https://doi.org/10.1002/hyp.9740

Li, D., Yang, K., Tang, W., Li, X., Zhou, X., & Guo, D. (2020). Characterizing precipitation in high altitudes of the western Tibetan Plateau with a focus on major glacier areas. *International Journal of Climatology*, *40*(12), 5114–5127. https://doi.org/10.1002/joc.6509

Li, X., Jiang, D., Zhang, Z., Zhang, R., Tian, Z., & Yan, Q. (2015). Mid-Pliocene westerlies from PlioMIP simulations. *Advances in Atmospheric Sciences*, *32*(7), 909–923. https://doi.org/10.1007/s00376-014-4171-7

Li, X., Ma, Y., Huang, Y., Hu, X., Wu, X., Wang, P., et al. (2016). Evaporation and surface energy budget over the largest high-altitude saline lake on the Qinghai–Tibet Plateau. *Journal of Geophysical Research: Atmospheres*, *121*, 10–470. https://doi.org/10.1002/2016JD025027

Li, Y., Su, F., Chen, D., & Tang, Q. (2019). Atmospheric water transport to the endorheic Tibetan Plateau and its effect on the hydrological status in the region. *Journal of Geophysical Research: Atmospheres*, *124*, 12864–12881. https://doi.org/10.1029/2019JD031297

Lin, C., Chen, D., Yang, K., & Ou, T. (2018). Impact of model resolution on simulating the water vapor transport through the central Himalayas: Implication for models' wet bias over the Tibetan Plateau. *Climate Dynamics*, *51*(9–10), 3195–3207. https://doi.org/10.1007/s00382-018-4074-x

Maussion, F., Scherer, D., Finkelnburg, R., Richters, J., Yang, W., & Yao, T. (2011). WRF simulation of a precipitation event over the Tibetan Plateau, China—An assessment using remote sensing and ground observations. *Hydrology and Earth System Sciences*, *15*(6), 1795–1817. https://doi.org/10.5194/hess-15-1795-2011

Maussion, F., Scherer, D., Mölg, T., Collier, E., Curio, J., & Finkelnburg, R. (2014). Precipitation seasonality and variability over the Tibetan Plateau as resolved by the High Asia reanalysis. *Journal of Climate*, *27*(5), 1910–1927. https://doi.org/10.1175/jcli-d-13-00282.1

Meehl, G. A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J. F. B., et al. (2007). The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bulletin of the American Meteorological Society*, *88*(9), 1383–1394. https://doi.org/10.1175/bams-88-9-1383

Miao, Y., Fang, X., Herrmann, M., Wu, F., Zhang, Y., & Liu, D. (2011). Miocene pollen record of KC-1 core in the Qaidam Basin, NE Tibetan Plateau and implications for evolution of the East Asian monsoon. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *299*(1–2), 30–38. https://doi.org/10.1016/j.palaeo.2010.10.026

Miao, Y. F., Fang, X. M., Wu, F. L., Cai, M. T., Song, C. H., Meng, Q. Q., & Xu, L. (2013). Late Cenozoic continuous aridification in the western Qaidam Basin: Evidence from sporopollen records. *Climate of the Past*, *9*(4), 1863–1877. https://doi.org/10.5194/cp-9-1863-2013

Mischke, S., Sun, Z., Herzschuh, U., Qiao, Z., & Sun, N. (2010). An ostracod-inferred large Middle Pleistocene freshwater lake in the presently hyper-arid Qaidam Basin (NW China). *Quaternary International*, *218*(1–2), 74–85. https://doi.org/10.1016/j.quaint.2009.03.002

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., & Clough, S. A. (1997). Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research*, *102*(D14), 16663–16682. https://doi.org/10.1029/97jd00237

Morrison, H., Thompson, G., & Tatarskii, V. (2009). Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes. *Monthly Weather Review*, *137*(3), 991–1007. https://doi.org/10.1175/2008mwr2556.1

Mutz, S. G., Ehlers, T. A., Li, J., Steger, C., Paeth, H., Werner, M., & Poulsen, C. J. (2016). Precipitation $\delta^{18}O$ over the Himalaya-Tibet orogen from ECHAM5-wiso simulations: Statistical analysis of temperature, topography and precipitation. *Journal of Geophysical Research: Atmospheres*, *121*, 9278–9300. https://doi.org/10.1002/2016JD024856

Mutz, S. G., Ehlers, T. A., Werner, M., Lohmann, G., Stepanek, C., & Li, J. (2018). Estimates of late Cenozoic climate change relevant to Earth surface processes in tectonically active orogens. *Earth Surface Dynamics*, *6*(2), 271–301. https://doi.org/10.5194/esurf-6-271-2018

Nakicenovic, N., Alcamo, J., Davis, G., De Vries, B., Fenhann, J., Gaffin, S., et al. (1990). *Special report on emissions scenarios: A special report of Working Group III of the Intergovernmental Panel on Climate Change* (Technical Report).

Pritchard, D. M., Forsythe, N., Fowler, H. J., O'Donnell, G. M., & Li, X.-F. (2019). Evaluation of Upper Indus near-surface climate representation by WRF in the High Asia Refined analysis. *Journal of Hydrometeorology*, *20*(3), 467–487.

Ravelo, A. C., Andreasen, D. H., Lyle, M., Lyle, A. O., & Wara, M. W. (2004). Regional climate shifts caused by gradual global cooling in the Pliocene epoch. *Nature*, *429*(6989), 263–267.

Rieser, A. B., Bojar, A.-V., Neubauer, F., Genser, J., Liu, Y., Ge, X.-H., & Friedl, G. (2009). Monitoring Cenozoic climate evolution of northeastern Tibet: Stable isotope constraints from the western Qaidam Basin, China. *International Journal of Earth Sciences*, *98*(5), 1063–1075.

Roeckner, E., Baeuml, G., Bonaventura, L., Brokopf, R., Esch, M., Giorgetta, M., et al. (2003). *The General Circulation Model ECHAM5. Part I: Model description* (Technical Report).

Scherer, D. (2020). Survival of the Qaidam mega-lake system under mid-Pliocene climates and its restoration under future climates. *Hydrology and Earth System Sciences*, *24*(7), 3835–3850. https://doi.org/10.5194/hess-24-3835-2020

Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Zhiquan, L., Berner, J., et al. (2019). *A description of the advanced research WRF model version 4 NCAR technical note* (Technical Report). NCAR-UCAR. Retrieved from http://library.ucar.edu/research/publish-technote

Stepanek, C., & Lohmann, G. (2012). Modeling mid-Pliocene climate with cosmos. *Geoscientific Model Development*, *5*, 1221–1243. https://doi.org/10.5194/gmd-5-1221-2012

Sun, J., Yang, K., Guo, W., Wang, Y., He, J., & Lu, H. (2020). Why has the inner Tibetan Plateau become wetter since the mid-1990s? *Journal of Climate*, *33*, 8507–8522.

Sun, Y., Ramstein, G., Contoux, C., & Zhou, T. (2013). A comparative study of large-scale atmospheric circulation in the context of a future scenario (RCP4.5) and past warmth (mid-Pliocene). *Climate of the Past*, *9*(4), 1613–1627. https://doi.org/10.5194/cp-9-1613-2013

Sun, Y., Ramstein, G., Li, L. Z., Contoux, C., Tan, N., & Zhou, T. (2018). Quantifying East Asian summer monsoon dynamics in the ECP4.5 scenario with reference to the mid-Piacenzian warm period. *Geophysical Research Letters*, *45*, 12523–12533. https://doi.org/10.1029/2018GL080061

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. https://doi.org/10.1175/bams-d-11-00094.1

Tewari, M., Chen, F., Wang, W., Dudhia, J., LeMone, M., Mitchell, K., et al. (2004). Implementation and verification of the unified Noah land surface model in the WRF model. In *Paper presented at 20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction* (Vol. 1115, pp. 2165–2170).

Wang, J., Fang, X., Appel, E., & Song, C. (2012). Pliocene–Pleistocene climate change at the NE Tibetan Plateau deduced from lithofacies variation in the drill core SG-1, western Qaidam Basin, China. *Journal of Sedimentary Research*, *82*(12), 933–952.

Wang, J., Song, C., Reager, J. T., Yao, F., Famiglietti, J. S., Sheng, Y., et al. (2018). Recent global decline in endorheic basin water storages. *Nature Geoscience*, *11*(12), 926–932. https://doi.org/10.1038/s41561-018-0265-7

Wang, X., Tolksdorf, V., Otto, M., & Scherer, D. (2021). WRF-based dynamical downscaling of ERA5 reanalysis data for High Mountain Asia: Toward a new version of the High Asia Refined analysis. *International Journal of Climatology*, *41*(1), 743–762. https://doi.org/10.1002/joc.6686

Wang, Z., Duan, A., Yang, S., & Ullah, K. (2017). Atmospheric moisture budget and its regulation on the variability of summer precipitation over the Tibetan Plateau. *Journal of Geophysical Research: Atmospheres*, *122*, 614–630. https://doi.org/10.1002/2016JD025515

Wasserstein, R. L., & Lazar, N. A. (2016). *The ASA statement on p-values: Context, process, and purpose*. Taylor & Francis.

Xu, G., Chen, T., Liu, X., An, W., Wang, W., & Yun, H. (2011). Potential linkages between the moisture variability in the northeastern Qaidam Basin, China, since 1800 and the East Asian summer monsoon as reflected by tree ring $\delta^{18}O$. *Journal of Geophysical Research*, *116*, D09111. https://doi.org/10.1029/2010JD015053

Xu, J., Yu, S., Liu, J., Haginoya, S., Ishigooka, Y., Kuwagata, T., et al. (2009). The implication of heat and water balance changes in a lake basin on the Tibetan Plateau. *Hydrological Research Letters*, *3*, 1–5. https://doi.org/10.3178/hrl.3.1

Yang, Q., Leung, L. R., Lu, J., Lin, Y.-L., Hagos, S., Sakaguchi, K., & Gao, Y. (2017). Exploring the effects of a nonhydrostatic dynamical core in high-resolution aquaplanet simulations. *Journal of Geophysical Research: Atmospheres*, *122*, 3245–3265. https://doi.org/10.1002/2016JD025287

Yekutieli, D., & Benjamini, Y. (1999). Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *Journal of Statistical Planning and Inference*, *82*(1), 171–196. https://doi.org/10.1016/s0378-3758(99)00041-5

Yu, L., & Lai, Z. (2012). OSL chronology and palaeoclimatic implications of aeolian sediments in the eastern Qaidam Basin of the northeastern Qinghai–Tibetan Plateau. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *337*, 120–129. https://doi.org/10.1016/j.palaeo.2012.04.004

Yu, S., Liu, J., Xu, J., & Wang, H. (2011). Evaporation and energy balance estimates over a large inland lake in the Tibet–Himalaya. *Environmental Earth Sciences*, *64*(4), 1169–1176. https://doi.org/10.1007/s12665-011-0933-z

Zhang, R., Yan, Q., Zhang, Z. S., Jiang, D., Otto-Bliesner, B. L., Haywood, A. M., et al. (2013). Mid-Pliocene East Asian monsoon climate simulated in the PlioMIP. *Climate of the Past*, *9*, 2085–2099. https://doi.org/10.5194/cp-9-2085-2013

Supporting Information for "Sensitivity of Water Balance in the Qaidam Basin to the Mid-Pliocene Climate"

Xun Wang[1], Benjamin Schmidt[1], Marco Otto[1], Todd A. Ehlers[2], Sebastian G. Mutz[2], Svetlana Botsyun[2], Dieter Scherer[1]

[1]Chair of Climatology, Technische Universität Berlin, Berlin, Germany

[2]Department of Geosciences, University Tübingen, Tübingen, Germany

Contents of this file

1. Figure S1
2. Figure S2
3. Table S1
4. Table S2
5. Setup files for the Weather Research and Forecasting model
   - namelist.wps
   - namelist.input

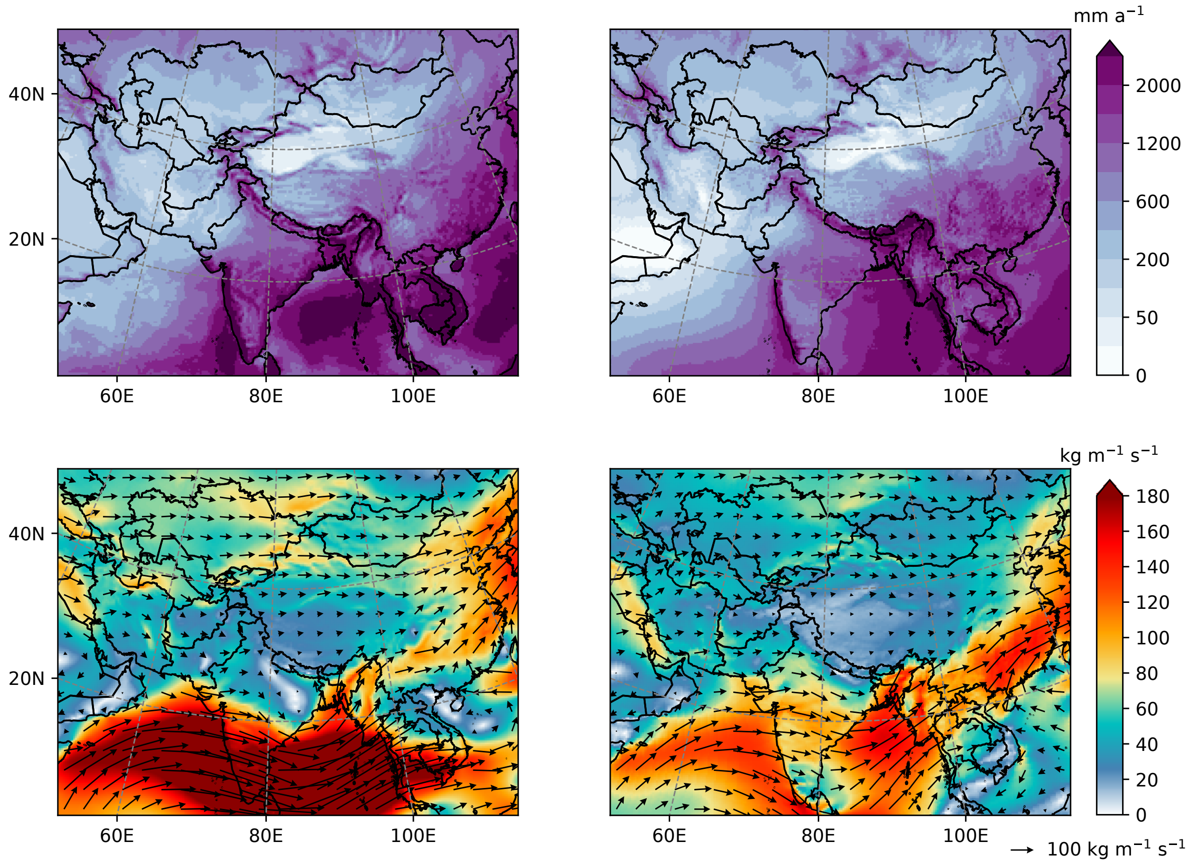

Figure S1. Comparison of annual precipitation (upper) and atmospheric water transport (lower) from PD (left) and ERA5 between 2000 and 2014 (right).

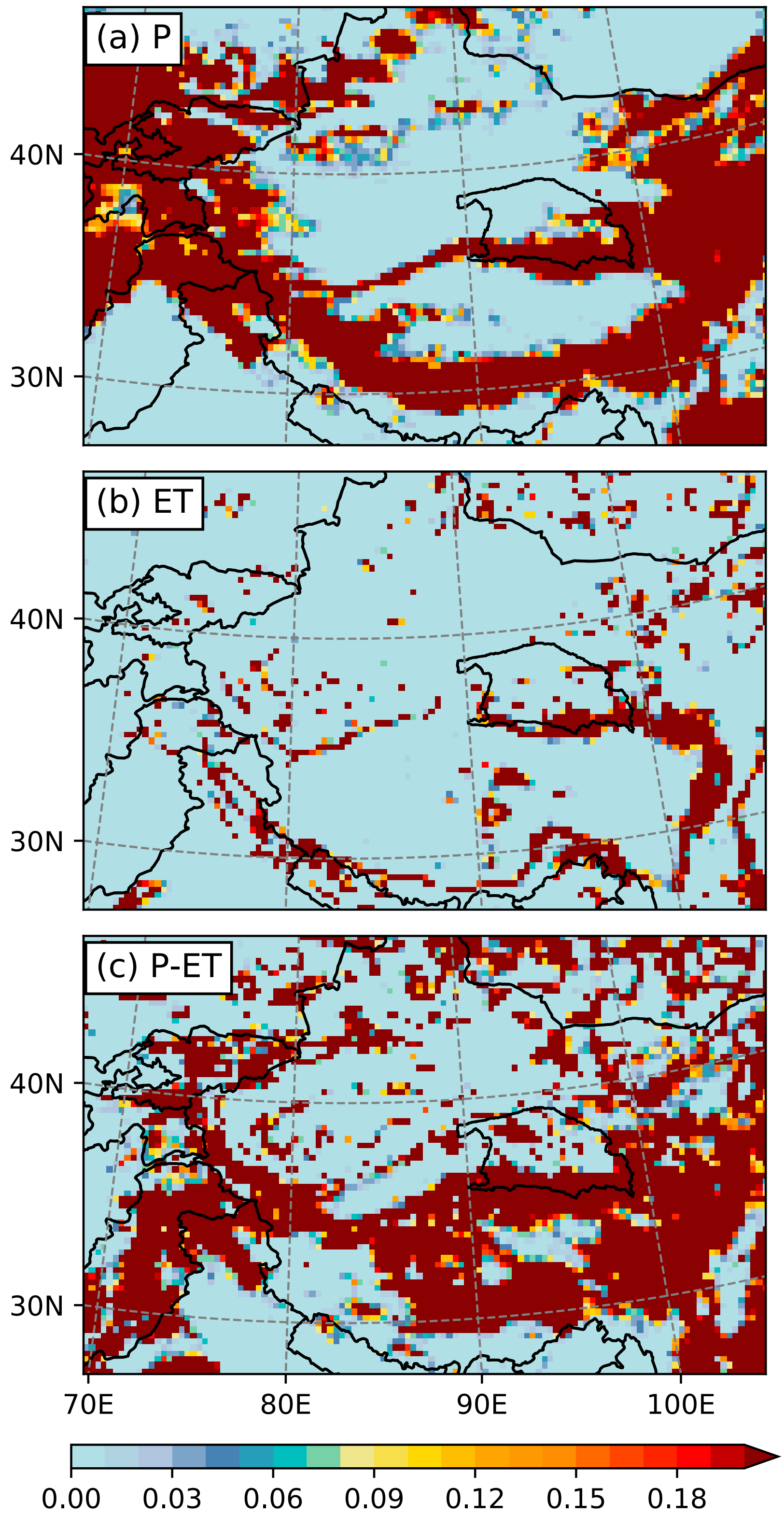

Figure S2. Maps of adjusted p-values from two-sided Welch's t-test for difference between PLIO and PD of 15-year average of annual (a) precipitation (P), (b) evapotranspiration (ET), and (c) net precipitation ($P - ET$).

Table S1. P-values from two-sided Welch's t-test for difference between PLIO and PD of 15-year average of seasonal and annual precipitation (P), evapotranspiration (ET), water balance (ΔS), air temperature at 2 m ($T2$) and specific humidity at 2 m ($Q2$) averaged over the Qaidam Basin (QB).

	P	ET	ΔS	$T2$	$Q2$
DJF	0.080	0.000	0.018	0.000	0.691
MAM	0.002	0.016	0.016	0.000	0.001
JJA	0.051	0.000	0.043	0.000	0.000
SON	0.000	0.000	0.010	0.000	0.001
Annual	0.000	0.000	0.058	0.000	0.000

Table S2. P-values from two-sided Welch's t-test for difference between PLIO and PD of 15-year average of seasonal and annual atmospheric water flux through each border of the Qaidam box (blue rectangle in Figure 1b).

	West	East	South	North	Sum
DJF	0.005	0.001	0.010	0.008	0.349
MAM	0.166	0.266	0.131	0.182	0.010
JJA	0.017	0.000	0.414	0.137	0.050
SON	0.062	0.351	0.297	0.696	0.063
Annual	0.377	0.185	0.246	0.495	0.001

Setup files for the the Weather Research and Forecasting model

namelist.wps

```
&share
 wrf_core = 'ARW',
 max_dom = 1,
 start_date = '2011-07-09_12:00:00',
 end_date = '2011-07-11_00:00:00',
 interval_seconds = 21600,
 io_form_geogrid = 2,
 debug_level = 0,
/

&geogrid
 parent_id         =   1,
 parent_grid_ratio =   1,
 i_parent_start    =   1,
 j_parent_start    =   1,
 e_we              =  281,
 e_sn              =  217,
 geog_data_res = 'usgs_lakes',
 dx = 30000,
 dy = 30000,
 map_proj = 'lambert',
 ref_lat   =  32,
 ref_lon   =  83,
 truelat1  =  32,
 truelat2  =  38,
 stand_lon =  83,
 geog_data_path = '/sim/wrf/static/WRFV4.0/',
 !opt_geogrid_tbl_path = 'geogrid/'

/

&ungrib
 out_format = 'WPS',
 prefix = 'ERA5_pl',
/

&metgrid
 fg_name = '/sim/forcing_data/GCM/PD/Post/IFF/GCM_PD',
 constants_name = './TAVGSFC',
 io_form_metgrid = 2,
/
```

namelist.input

```
&time_control
 run_days = 1,
 run_hours = 12,
 run_minutes = 0,
 run_seconds = 0,
 start_year = 2011,2011,2011,
 start_month = 07, 07, 07,
 start_day = 09, 09, 09,
 start_hour = 12, 12, 12,
 start_minute = 00, 00, 00,
 start_second = 00, 00, 00,
 end_year = 2011,2011,2011,
 end_month = 07, 07, 07,
 end_day = 11, 11, 11,
 end_hour = 00, 00, 00,
 end_minute = 00, 00, 00,
 end_second = 00, 00, 00,
 interval_seconds = 21600,
 input_from_file = .true., .true., .true.,
 history_interval = 180, 60, 60,
 frames_per_outfile = 1000, 1000, 1000,
 restart = .false.,
 restart_interval = 5000,
 io_form_history = 2,
 io_form_restart = 2,
 io_form_input = 2,
 io_form_boundary = 2,
 debug_level = 0,
 auxinput4_inname = "wrflowinp_d<domain>",
 auxinput4_interval = 360,
 io_form_auxinput4 = 2,
 iofields_filename = "./add_out_d01.txt",
 ignore_iofields_warning = .false.,
/

&domains
 max_dom = 1,
 time_step = 120,
 time_step_fract_num = 0,
 time_step_fract_den = 1,
 s_we = 1, 1, 1,
 e_we = 281, 382, 711,
 s_sn = 1, 1, 1,
 e_sn = 217, 253, 296,
 s_vert = 1, 1, 1,
 e_vert = 28, 28, 28,
 eta_levels = 1.000000,0.993000,0.983000,0.970000,0.954000,0.934000,0.909000,0.880000,
              0.829576,0.779151,0.728727,0.678303,0.591744,0.513694,0.443454,0.380375,
              0.323853,0.273326,0.228273,0.188210,0.152689,0.121294,0.093643,0.069378,
              0.048173,0.029725,0.013753,0.000000,
 num_metgrid_levels = 31,
 num_metgrid_soil_levels = 5,
 dx = 30000, 10000, 2000,
 dy = 30000, 10000, 2000,
 grid_id = 1, 2, 3,
 parent_id = 1, 1, 2,
 i_parent_start = 1, 85, 202,
 j_parent_start = 1, 84, 115,
 parent_grid_ratio = 1, 3, 5,
 parent_time_step_ratio = 1, 3, 5,
 feedback = 1,
 smooth_option = 0,
 hypsometric_opt =1,
 interp_theta = .true.,
 lagrange_order = 1,
 /

&physics
 mp_physics = 10, 10, 10,
 ra_lw_physics = 1, 1, 1,
 ra_sw_physics = 1, 1, 1,
 radt = 30, 10, 2,
 sf_sfclay_physics = 1, 1, 1,
 sf_surface_physics = 2, 2, 2,
 bl_pbl_physics = 1, 1, 1,
 bldt = 0, 0, 0,
 cu_physics = 10, 10, 0,
 cudt = 5, 1, 1,
```

```
 isfflx = 1,
 icloud = 1,
 surface_input_source = 1,
 num_soil_layers = 4,
 mp_zero_out = 0,
 num_land_cat = 28,
 sst_update = 1,
 usemonalb = .false.,
 rdmaxalb = .true.,
 seaice_threshold = 271,
/

&fdda
/

&dynamics
 hybrid_opt= 0,
 rk_ord = 3,
 w_damping = 1 ,
 diff_opt = 1, 1, 1,
 km_opt = 4, 4, 4,
 diff_6th_opt = 2, 2, 2,
 diff_6th_factor = 0.12, 0.12, 0.12,
 base_temp = 290.,
 damp_opt = 3,
 zdamp = 5000., 5000., 5000.,
 dampcoef = 0.2, 0.2, 0.2,
 khdif = 0, 0, 0,
 kvdif = 0, 0, 0,
 non_hydrostatic = .true., .true., .true.,
 moist_adv_opt = 1, 1, 1,
 scalar_adv_opt = 1, 1, 1,
 use_theta_m = 0,
 epssm = 0.1, 0.1, 0.5,
/

&bdy_control
 spec_bdy_width = 5,
 spec_zone = 1,
 relax_zone = 4,
 specified = .true., .false., .false.,
 nested = .false., .true., .true.,
/

&namelist_quilt
 nio_tasks_per_group = 0,
 nio_groups = 1,
/

&grib2
/
```