#### **ORIGINAL PAPER**



# Modeling regional precipitation over the Indus River basin of Pakistan using statistical downscaling

Muhammad Saleem Pomee 1,2 . Moetasim Ashfaq 3 · Bashir Ahmad 2 · Elke Hertig 1

Received: 3 October 2019 / Accepted: 27 April 2020 / Published online: 23 June 2020 © The Author(s) 2020

#### **Abstract**

Complex processes govern spatiotemporal distribution of precipitation within the high-mountainous headwater regions (commonly known as the upper Indus basin (UIB)), of the Indus River basin of Pakistan. Reliable precipitation simulations particularly over the UIB present a major scientific challenge due to regional complexity and inadequate observational coverage. Here, we present a statistical downscaling approach to model observed precipitation of the entire Indus basin, with a focus on UIB within available data constraints. Taking advantage of recent high altitude (HA) observatories, we perform precipitation regionalization using K-means cluster analysis to demonstrate effectiveness of low-altitude stations to provide useful precipitation inferences over more uncertain and hydrologically important HA of the UIB. We further employ generalized linear models (GLM) with gamma and Tweedie distributions to identify major dynamic and thermodynamic drivers from a reanalysis dataset within a robust cross-validation framework that explain observed spatiotemporal precipitation patterns across the Indus basin. Final statistical models demonstrate higher predictability to resolve precipitation variability over wetter southern Himalayans and different lower Indus regions, by mainly using different dynamic predictors. The modeling framework also shows an adequate performance over more complex and uncertain trans-Himalayans and the northwestern regions of the UIB, particularly during the seasons dominated by the westerly circulations. However, the cryosphere-dominated trans-Himalayan regions, which largely govern the basin hydrology, require relatively complex models that contain dynamic and thermodynamic circulations. We also analyzed relevant atmospheric circulations during precipitation anomalies over the UIB, to evaluate physical consistency of the statistical models, as an additional measure of reliability. Overall, our results suggest that such circulation-based statistical downscaling has the potential to improve our understanding towards distinct features of the regional-scale precipitation across the upper and lower Indus basin. Such understanding should help to assess the response of this complex, data-scarce, and climatesensitive river basin amid future climatic changes, to serve communal and scientific interests.

#### 1 Introduction

The trans-Himalayan and transboundary Indus River flows approximately 3200 km across its 1.12 million square kilometer basin to ultimately descend into the Arabian Sea (FAO 2011). The upper Indus basin (UIB) marked by the confluence

- Muhammad Saleem Pomee saleemwrri@gmail.com
- Regional Climate Change and Health, Faculty of Medicine, University of Augsburg, Augsburg, Germany
- Pakistan Agricultural Research Council (PARC), Islamabad, Pakistan
- <sup>3</sup> Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

of the Hindukush, Karakoram, and Himalayans mountain ranges (HKH) contains the largest cryosphere volumes outside the Poles (Soncini et al. 2015; Qiu 2008). The HKH topography modulates two contrasting synoptic-scale processes: the South Asian Summer Monsoon and Western Disturbances, to determine highly variable patterns of observed precipitation over the UIB (Filippi et al. 2014; Palazzi et al. 2013). Snow and glacial melt within the UIB mainly govern the hydrological regime of the Indus River (e.g., Tahir et al. 2011; Archer and Fowler 2004), which sustains the livelihood of nearly 215 million downstream inhabitants (Latif et al. 2018; Lutz et al. 2016). In contrast, the lower Indus (LI) primarily has an arid to semi-arid climate and represents large fertile plains, spate irrigation regions, and a diverse coastal ecology.

An accurate knowledge of the spatial and temporal distribution of precipitation is largely unknown in the high altitudes



(HA) of the UIB (Immerzeel et al. 2015; Hewitt 2005; Winiger et al. 2005). This lack of certainty in precipitation characteristics stems from an insufficient number of HA meteorological observatories, inaccuracies in available HA measurements (Rasmussen et al. 2012), concerns about the data quality (Hewitt 2011; Winiger et al. 2005), and serious constraints on transboundary data sharing due to regional politics (Dahri et al. 2016). It should be noted that most of the observatories with long-term records are sparsely located in the low altitudes, while short-term and inconsistent observations of the HA limit our understanding about the regional orography.

Given an elevated susceptibility to climate change and the associated high social-ecological vulnerability (MRI 2015; Nepal and Shrestha 2015), a reliable knowledge towards the spatiotemporal organization of regional precipitation is imperative to implement efficient adaptations at the basin level.

However, the observational constraints hamper a reliable modeling of the processes that govern precipitation distribution within the basin. While different gridded products (e.g., reanalysis, remote sensing or interpolated station observations) can potentially help to overcome these observational deficiencies (Immerzeel et al. 2015), the scope of such improvements is somewhat limited over the complex and largely snow-covered HA of the UIB (Immerzeel et al. 2015; Huffman et al. 2007; Turner and Annamalai 2012).

Given the unique observational challenges, different strategies have been adapted to improve the confidence in precipitation simulations over the UIB. However, studies often find contrasting climate change signals over these regions, which range from rapidly retreating glaciers (Kääb et al. 2012, 2015; Wiltshire 2014; Jacob et al. 2012; Cogley 2011) to rather glacial expansions in the Karakoram, also known as the Karakoram anomaly (e.g., Bashir et al. 2017; Kapnick et al. 2014; Tahir et al. 2014; Bhambri et al. 2013; Minora et al. 2013). The lack of robustness in the projected change signal and associated glacial response over this region implicates the development of effective policy measures to reduce projected vulnerability over the basin. Therefore, it is imperative to understand the influence of existing observational and modeling challenges on projected climatic changes over this region and to devise strategies for improving spatial inferences of regional precipitation. However, such improvements potentially have to be derived through methodological considerations, as the availability of additional high-quality data is still an ongoing issue.

In this context, the reliability of spatially distributed precipitation estimates may suffer from (i) altitudinal biases; (ii) issues with observational data (homogeneity, time series length, consistency), and/or (iii) the downscaling approaches. We refer to these challenges as the type 1, type 2, and type 3 uncertainties, respectively. Many previous studies exhibit the type 1 uncertainty because they rely on the low-altitudinal observations to infer seasonal and

annual precipitation statistics over the UIB (e.g., Iqbal and Athar 2018, Ali et al. 2015; Khan et al. 2015a; Akhtar et al. 2008; Archer and Fowler 2004, Latif et al. 2018 and Khattak et al. 2011). Estimates in most of these earlier studies neither represent the HA nor provide any quantifiable mechanism for drawing logical inferences between the low and high altitudes, and therefore may lack the reliability. Some of the recent studies (e.g., Immerzeel et al. 2015; Dahri et al. 2016, 2018; Hasson 2016) have addressed this deficiency, either by incorporating the HA perspective using short-term data (e.g., glacial mass balance and or fragmented HA monitoring) or by other assumptions. These studies significantly differ in regional precipitation estimates over the UIB, when compared with historically observed estimates recorded by the valley stations. However, although estimates in these studies are derived by maximizing spatial coverage within the UIB, they may suffer from some other shortcomings. For example, the use of a limited number of HA stations, which not only contain short-term observations (e.g., Hasson 2016) but may also exhibit issues related to homogeneity (e.g., Dahri et al. 2016), measurement errors, and assumption of a linear precipitation gradient along the altitudes (e.g. Immerzeel et al. 2013), and therefore may induce type 2 uncertainties in respective simulations.

The type 3 uncertainty originates from the choice of the adapted statistical downscaling methodology. For instance, bias correction methods, which have been widely used in these previous studies, highly depend on the observations for an appropriate characterization of the region of interest. However, the observation sparsity over the UIB makes the use of bias correction relatively ineffective. Lastly, the current generation of General Circulation Models (GCMs) has persistently failed in the representation of dynamic and thermodynamic processes that dictate precipitation variability over this region at varying time scales (e.g., Ashfaq et al. 2017; Palazzi et al. 2015; Rastogi et al. 2018). Alternatively, regional climate models (RCMs) generally offer improved regional simulations due to better topographic representation and the flexibility of region-specific tuning of the model parameters. However, an evaluation of the seven fine-scale (0.44°) CORDEX-SA experiments has also shown a limited success over this region since they still derive boundary forcing from the biased GCMs (e.g., Hasson et al. 2019; Mishra 2015; Syed et al. 2014). The High Asia Refined analysis (HAR), which is based on dynamical modeling, has shown promising results in reproducing overall regional precipitation climatology and in explaining sub-regional precipitation variability over time (e.g., Maussion et al. 2014; Curio and Scherer 2016). More recently, Pritchard et al. (2019) have evaluated HAR performance specifically over the UIB in terms of precipitation and other important near-surface variables. Although the annual precipitation cycle is reasonably simulated, there exists some



biases at seasonal scales, which lead to wetter and colder conditions across the HA areas.

It is clear that further efforts are needed to reduce uncertainties in precipitation estimates over the UIB. For example, incorporating a time series analysis along with careful data quality checks, an avoidance of the direct use of available HA observations, and the adaption of other, potentially better suited statistical downscaling approaches, may help to resolve some of these uncertainties. To this end, perfect prognosis statistical downscaling (Wilks 2006) is a promising statistical alternative, which offers distinct advantages. However, so far only a few studies are reported in the region that make use of this technique (e.g., Kazmi et al. 2016; Mahmood and Babel 2012). This approach employs large-scale atmospheric circulations within a statistical downscaling framework to resolve the observed precipitation distributions. The use of atmospheric circulations is advantageous as these variables are relatively better modeled in the GCMs (Kaspar-Ott et al. 2019), and can explain the physical mechanisms that govern regional precipitation. Given the observational uncertainty in the UIB, such physical explanations of the downscaled precipitation may further add confidence in the statistical modeling results. Moreover, knowledge towards the large-scale dynamics that influence precipitation distribution over the UIB is relatively robust, as many of the previous studies provide a good overview in this regard that can serve as the basis for comparison. For instance, Syed et al. (2006, 2010), Curio and Scherer (2016), Cannon et al. (2015), Ahmad et al. (2015), and Kazmi et al. (2016) explain different aspects of the precipitation governing circulations over the region.

Considering such advantages and to present a different perspective, we have adapted perfect prognosis downscaling to investigate the observed precipitation dynamics within the Indus basin with a focus on the UIB, by accounting for spatiotemporal precipitation variability. Due to the dominant role of synoptic-scale circulations in regional precipitation over the UIB, we assume that despite the differences in magnitudes, distinct linear relationships of temporal precipitation variability between the low and HA should exist, at least on subregional scales. Such intra-regional signals, if properly identified and captured, can significantly improve the understanding of precipitation variability over the relatively uncertain HA. The availability of the recent HA observations can help to validate this assumption. Especially, this study aims to:

- Evaluate the effectiveness of low altitude stations in representing HA precipitation dynamics across the UIB without using actual precipitation measurements
- ii) Identify precipitation governing circulations within a robust statistical downscaling framework to explain observed spatiotemporal precipitation patterns
- iii) Analyze the physical consistency of final statistical models through composites of governing circulations

Moreover, extending the analysis to the LI further helps to understand the water demand perspective, which is necessary for effective water resources planning at the basin scale.

In Section 2, we provide a brief description of the study area. Data and details of the adapted downscaling framework and the approach for seasonal composites are described in Section 3. This is followed by the description of results and discussion in Section 4 and the details of composite analysis and sources of uncertainty in Section 5. Finally, the conclusions are drawn in Section 6.

#### 2 Study area

The Indus River originates from the southwestern Tibetan Plateau and ultimately descends into the Arabian Sea after traversing through the entirety of Pakistan. The River basin area stretches across four geopolitically complex South Asian countries (i.e., China, India, Afghanistan, and Pakistan) and is ranked 12th in the world in terms of its size (Fig. 1a). Annual precipitation substantially varies across the basin with a magnitude as low as 150 mm in the south to more than 2000 mm in the northern highlands and follows a nonlinear variation along the altitudes (Archer and Fowler 2004; Dahri et al. 2016). It should be noted that this study only focuses on the basin area controlled by Pakistan (the largest basin area among all four countries and contains the highest altitudes), due to severe constraints on availability of data over rest of the region (Fig. 1b). Within this study, the UIB outlines the drainage area above the Mangla dam over River Jehlum in Pakistan, as shown in blue color in Fig. 1a.

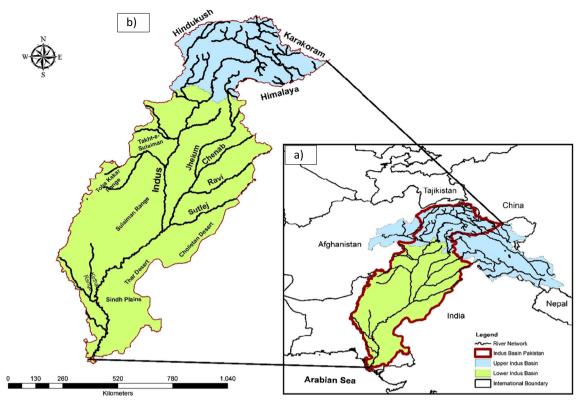
#### 3 Data and methodology

### 3.1 Precipitation: sources, homogeneity, and distributional considerations

The Pakistan Meteorological Department (PMD) operates the largest monitoring network across Pakistan that also includes the Indus River basin. However, these observatories mostly represent low-altitude regimes (only two stations are located above 2000 m altitude with a maximum at 2394 m). Since the mid-1990s, the Water and Power Development Authority (WAPDA) of Pakistan has initiated a systematic monitoring program to represent major sub-catchments within the UIB and that has significantly improved the spatial and altitudinal coverage. Additionally, the University of Bonn under Cultural Areas Karakoram (CAK) Program has some operational observatories in the HA of Hunza and Gilgit sub-catchments of the UIB.

We obtain precipitation time series of 58 stations (shown in Fig. 2) from these three sources for further investigations. Out of these 58 stations, 42 stations are located within the UIB to





**Fig. 1** (a) The transboundary Indus River basin, shown as colored regions, along with the major tributary river network. The color scheme differentiates between the upper and lower Indus basin. (b) The study area

(Indus basin of Pakistan) with the same color scheme to represent the upper Indus basin (UIB) and lower Indus (LI) basin of Pakistan. The Mountains of the HKH region are also shown around the UIB

account for its topographic complexity and significance for the Indus waters. The mean data length of the 35 historic and the 23 recent HA stations is 37 (1979–2015) and 17 (1994–2015) years, respectively. The inclusion of the HA stations has extended the observational coverage of the UIB up to 4730 m (20, 11, and 8 stations are located above 2000, 3000, and 3500 meters, respectively), to understand orographic dynamics, which otherwise is least understood, and govern the basin hydrology. Information about the selected stations is given in Table 1.

Based upon station precipitation analysis, we identify three major seasons that cover the winter period spanning from December through March (WS, DJFM), the pre-monsoon spanning from April to June (PMS, AMJ), and the summer monsoon season that spans from July to September (MS, JAS). These seasons contribute about 34%, 22%, and 36%, respectively, of the annual total of the basin precipitation and provide the basis for further analysis in our study.

We group monthly time series in each of the three seasons and check the data for its completeness (Moberg et al. 2006), homogeneity (Wijngaard et al. 2003; Alexandersson 1986), and underlying data structure. We make use of four different statistical procedures for homogeneity testing to group the stations as "useful" (where at least three tests indicate homogeneity), "doubtful" (where at least two tests indicate inhomogeneity), and "suspect" (where none or only one test indicates

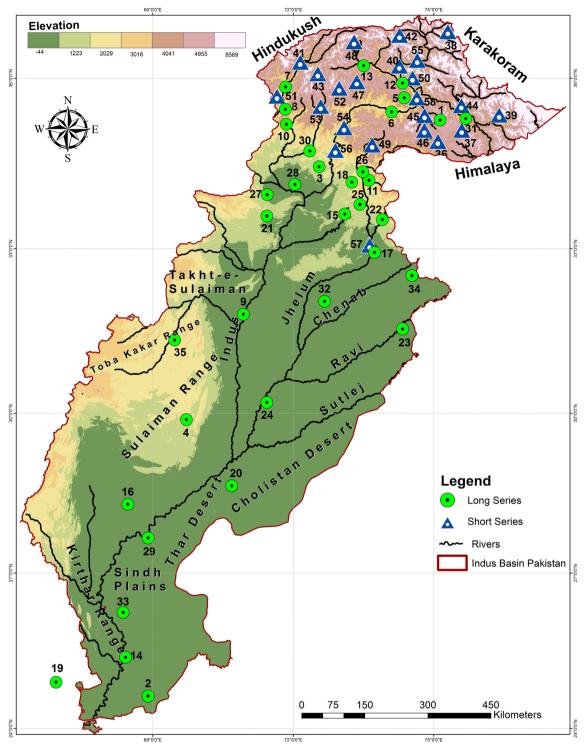
homogeneity) after Wijngaard et al. (2003). Although most of the stations appeared as "useful," some seasonal inhomogeneity has also been identified. For example, we find seven stations as "suspect" and seven stations as "doubtful" in the WS, four stations as "suspect" and eight stations as "doubtful" in the PMS, and four stations as "suspect" and nine stations as "doubtful" in the MS. Most of the stations with homogeneity issues are located within the UIB. Therefore, the use of these stations without homogeneity considerations, as it has been done in previous studies, can potentially lead to errors in the seasonal precipitation estimates.

We further use several goodness-of-fit measures like the KS test (Smirnov 1939), Anderson-Darling Statistics (Anderson and Darling 1952), and Akaike and Bayesian Information Criterions (Akaike 1973; Stone 1979), to identify the underlying distribution for the observed precipitation. Our analysis suggests that gamma distribution is the best statistical representative for nearly all of the study stations. This identification helps to facilitate the choice of an appropriate statistical model, which is described in detail in Section 3.5.

### 3.2 Precipitation regionalization and selection of representative stations

We assume that large-scale atmospheric circulations drive precipitation variability within the UIB. Therefore, it is





**Fig. 2** Location of the (numbered) study stations over the Indus basin of Pakistan. The green circles represent the locations of historic stations (1979–2015), while triangles show the recent HA installations (1994–

2015) within the UIB. The color scheme here represents the altitudinal perspective across the River basin. The major river network is also shown and named. For information about the numbers, please see Table 1

expected that despite differences in the magnitudes, commonalities in temporal precipitation variability should exist among the low and HA across the UIB. To test this assumption, we employ a K-means cluster analysis on all 58 stations to

identify the precipitation regions with similar covariance, using Spearman correlation as a distance measure (Wilks 2006). Such a precipitation analysis, which avoids actual precipitation amounts, is advantageous as it can potentially



**Table 1** Overview of the meteorological stations used in this study. The Long and Lat are longitude and latitude measurements, expressed in decimal degrees (dd). Altitudes represent average station elevation above mean sea level in meters. Source: *PMD* Pakistan Meteorological

Department, *WAPDA* Water and Power Development Authority of Pakistan, *CAK* University of Bonn under Cultural Areas Karakoram Program (synthesized from Weniger et al. 2005 and in personal communication with the lead author of Dahri et al. 2016)

Sr. No	Station name	Long	Lat	Altitude	Time series	Source	Sr. No	Station	Long	Lat	Altitude	Time series	Source
		(dd)	(dd)	(m)	(length)		INU	name	(dd)	(dd)	(m)	(length)	
1	Astore	74.90	35.33	2394	1979–2015	PMD	41	Zaini	72.15	36.28	3000	1994–2015	WAPDA
2	Badin	68.90	24.63	9	1979–2015	PMD	42	Ziarat	74.28	36.83	3669	1994–2015	WAPDA
3	Balakot	72.55	34.55	995	1979–2015	PMD	43	Schndor	72.53	36.09	3719	1994–2015	WAPDA
4	Barkhan	69.72	29.88	1097	1979–2015	PMD	44	Shigar	75.59	35.53	2470	1996-2015	WAPDA
5	Bunji	74.63	35.67	1372	1979–2015	PMD	45	Rama	74.81	35.36	3140	1999–2015	WAPDA
6	Chilas	74.10	35.42	1251	1979–2015	PMD	46	Rattu	74.81	35.15	2920	1999–2015	WAPDA
7	Chitral	71.83	35.85	1498	1979–2015	PMD	47	Ushkore	73.36	36.02	3353	1999–2015	WAPDA
8	Darosh	71.78	35.57	1464	1979–2015	PMD	48	Yasin	73.30	36.63	3353	1999–2015	WAPDA
9	DI Khan	70.93	31.82	172	1979–2015	PMD	49	SaifulMaluk	73.69	34.84	3200	2000-2015	PMD
10	Dir	71.85	35.20	1425	1979–2015	PMD	50	Bagrot	74.55	36.01	2310	1994–2010	CAK
11	Ghari Duptta	73.62	34.22	814	1979–2015	PMD	51	Kelash	71.65	35.70	2810	2003-2015	PMD
12	Gilgit	74.33	35.92	1460	1979–2015	PMD	52	Kalam	72.98	35.83	2744	1995–2015	PMD
13	Gupis	73.40	36.17	2156	1979–2015	PMD	53	Khot	72.59	36.52	3505	1994–2015	WAPDA
14	Hyderabad		25.38		1979–2015	PMD	54	Pattan		35.06		2004–2015	PMD
15	Islamabad		33.62		1979–2015		55	Hunza-PMD	74.65			2007–2015	
16	Jacobabad		28.30		1979–2015		56	Malam Jaba		34.75		2003–2015	
17	Jehlum	73.73	32.93	287	1979–2015		57	Mangla		33.10		1995–2015	
18	Kakul		34.18		1979–2015		58	MirKhani		35.50		2008–2015	
19	Karachi-AP	66.93	24.90	22	1979–2015								
20	Khanpur-PBO	70.68	28.65	93	1979–2015								
21	Kohat	71.43	33.58	327	1979–2015	PMD							
22	Kotli	73.90	33.52	610	1979–2015	PMD							
23	Lahore-PBO	74.33	31.55	214	1979–2015	PMD							
24	Multan-PBO	71.43	30.20	122	1979–2015	PMD							
25	Muree	73.40	33.90	2168	1979–2015	PMD							
26	Muzafarabad	73.48	34.37	702	1979–2015	PMD							
27	Peshawer AP	71.51	33.99	353	1979–2015	PMD							
28	Risalpur	71.98	34.08	308	1979–2015	PMD							
29	Rohri	68.90	27.67	66	1979–2015	PMD							
30	Saidu Sharif		34.73		1979–2015	PMD							
31	Sakardu-PBO	75.68	35.30	2210	1979–2015	PMD							
32	Sargodha	72.67	32.05	187	1979–2015	PMD							
33	Shaheed		26.25		1979–2015								
2.4	Banazirabad	74.52	22.52	255	1070 2015	DMD							
34	Sialkot Cant		32.52		1979–2015								
35 26	Zhob-PBO		31.35		1979–2015								
36	Burzil		34.91		1999–2015								
37	Deosai		35.10		1995–2015								
38	Kunjrab		36.85		1994–2015								
39	Hushey		35.37		1994–2015								
40	Naltar	74.27	36.22	2810	1995–2015	WAPDA							



account for the erroneous HA precipitation measurements (e.g., Curio and Scherer 2016; Tahir et al. 2011). During clustering, the objective function is set to maximize (minimize) correlation within (across) the regions to define sharp regional boundaries. Due to greater uncertainty in precipitation observations within the HA of the UIB, we also perform another regionalization experiment that only covers the HA part of the basin. Subsequently, the cluster centroids are computed and correlated with each of the respective regional members. The regional representative (RR) stations for each region are selected by using multiple considerations, which include a correlation between centroid and station, number of missing values, the level of homogeneity, and the length of observational record. The time series of these RR stations serve as predictands during the subsequent modeling process to draw sub-regional precipitation inferences.

#### 3.3 Predictor data

We use variables of ERA-Interim reanalysis (Dee et al. 2011) at a  $2 \times 2$  grid resolution as predictors. Using the understanding established in earlier studies (e.g., Ahmad et al. 2015; Kazmi et al. 2016; Mahmood and Babel 2012; Syed et al. 2010), a number of dynamic and thermodynamic variables at different vertical levels have been selected. These include geopotential heights (zg) at three atmospheric levels (200, 500, 700 hPa), meridional and zonal winds (ua and va) at four atmospheric levels (200, 500, 700, 850 hPa) and mean sea level pressure (psl) as dynamic variables, and relative and specific humidity (hur and hus) at two atmospheric levels (1000 and 700 hPa) as thermodynamic predictors. A larger domain was considered for the dynamic variables (10° E to 100° E, 10° N to 60° N) compared with the domain used for thermodynamic variables (64° E to 80° E, 22° N to 40° N) to account for both, the large-scale dynamical and more regional-scale thermodynamic influences.

#### 3.4 Principle component analysis

We perform s-mode Varimax-rotated PCA, separately for each predictor to identify important centers of variation and to reduce dimensionality of the predictors (see Preisendorfer 1988). The number of PCs is extracted using a modified dominance criterion (Philipp 2003) with some additional constraints that each retained PC should dominate all other PCs by more than one standard deviation over at least seven grid boxes, and explains at least 3 % of the total variance. The PC scores serve as the predictor time series and the PC loadings define the location of centers of variation.

#### 3.5 Generalized linear models

We adapt GLM framework (Mc Cullagh and Nelder 1989) to model the relationships between the atmospheric variables (PC scores) and observed precipitation of the RR stations. We selected the GLM framework due to gamma-distributed monthly precipitation. Within a GLM framework, the maximum likelihood estimation is used to estimate the model parameters for the expected value E of a random variable ( $Y_t$ ) at any given time t:

$$E(Y_t) = \mu_t = g^{-1}(\eta_t) = g^{-1} \sum_{i=1}^{n} X_{ti} \beta_i$$
 (1)

where  $\mu_t$  is the mean for the probability density function (PDF) at time t,  $\eta_t$  is a combination of linear predictors, g is the canonical link function,  $X_{tj}$  is the value of the jth covariate for observation t, n is the total number of covariates, and  $\beta_j$  are parameters whose values have to be estimated from the data. We use log-canonical link function in this study.

For those cases having exact zeros in observed time series of the RR, we use Tweedie exponential dispersion models within a GLM framework, due to their ability for simultaneous modeling of discrete and continuous precipitation features (e.g., Dunn 2004; Hertig and Jacobeit 2015; Hertig et al. 2017). The Tweedie family of distributions has three parameters (i.e.,  $\mu$ , mean, dispersion parameter  $\phi > 0$ , and the Tweedie index parameter p). The index p defines the particular distribution and a typical case (1 representsPoisson-gamma models, which are suitable for modeling positive continuous data with exact zeros (Hasan and Dunn 2011). The variance of the distribution is var  $[Y] = \phi \mu^p$ . The estimation of the index parameter requires sophisticated numerical computations that have been resolved using the profile maximum likelihood estimate provided in the Tweedie R package (Dunn 2010). In summary, we use Tweedie model only when exact zeros appear in the predictand time series and otherwise gamma models for regression analysis.

#### 3.5.1 Measures of model performance

Mean squared error skill score (MSESS) is a commonly employed metric to judge the accuracy of continuous non-probabilistic forecasts like precipitation, over the long-term climatology (Wilks 2006). MSESS is defined as:

$$MSESS = 1 - \frac{MSE_{modeled}}{MSE_{reference}} \times 100$$
 (2)

where

MSE modeled = 
$$\frac{1}{n} \sum_{n=1}^{n} (y^* - y)^2$$
 (3)



MSE reference = 
$$\frac{1}{n} \sum_{n=1}^{n} \left( \overline{y} - y \right)^2$$
 (4)

and

MSE mean squared errors  $y^*$  model prediction

y observation

 $\overline{y}$  mean over the observations number of observations

MSESS ranges from 0% (no skill improvement over the reference model, in this case a long-term climatology) to 100% (perfect model).

### 3.6 Downscaling framework: model development and selection

We develop GLM-based regression models to identify the atmospheric variables that exert the strongest influence on precipitation. We use a cross-validation framework by randomly selecting calibration (two-thirds of available time series) and validation periods (remaining one-third) and adapt the following stepwise reduction process, which is randomly repeated 1000 times to identify effective and robust statistical models:

I. The PC time series of each predictor is individually regressed against each of the RR stations. Initially, the regression models are developed using all PCs of an individual predictor as input. Subsequently, the revised models are developed based on an adjusted PC input (i.e., dropping PCs, one by one) and the corresponding MSE is calculated. We use MSE to rank the significance of the PCs and to identify the most influential PCs of each predictor (i.e., PCs whose absence maximized the errors were considered the most important). Next, the models containing only two most influential PCs are developed and the corresponding errors and MSESS are computed. These 2 PC models are further complemented one by one with the remaining ranked PCs to identify the best PC combination per predictor (i.e., which maximizes MSESS during the calibration and validation phases). We repeat this process for all 15 predictors and the model that demonstrates maximum predictability over the calibration and validation periods (sum of the MSESS) is termed as the best signal predictor model (SPM).

II. We further test different predictor combinations to evaluate any performance improvement over the best SPM from step I. For such combinations, the best SPM is complemented one by one with the remaining predictors. If two correlated PCs of different variables (absolute correlation threshold of 0.40) are found, only the more influential PC (showing greater error effect, when removed) was retained for further consideration. Moreover, an addition of a second predictor is only

considered, if it improves the prediction skills by at least 7% over the best SPM. This arbitrary number effectively demonstrates reasonable validation improvements without overfitting the models. The same procedure is repeated during the combinations of three or more predictor variables to identify better performing combinations, wherever applicable.

#### 3.7 Composite analysis

Composites can help to identify atmospheric circulations that are associated with different precipitation regimes to evaluate the physical consistency (reliability) of the statistical models. We construct the sample composites where monthly scores of the dominating PC (as identified in the final regression models) exceed or fall below  $\pm$  1.5 PC scores to subsample the cases of higher positive and higher negative PC scores. These subsets are evaluated to identify wetter (dryer) precipitation regimes of the predictand and their associated summary statistics such as the number of events, mean precipitation rate per event, and the dominant month. The same thresholds are further used to subsample relevant large-scale seasonal circulations to define climatology over respective subsamples and the standardized anomalies of these predictors are plotted to investigate precipitation supporting (suppressing) circulation patterns. Similarly, the difference between these contrasting regimes is also plotted to identify the circulation anomalies that support above-normal precipitation.

#### 4 Results and discussion

#### 4.1 Precipitation regionalization

In our analyses, the K-means clustering identifies six WS, seven PMS, and seven MS precipitation regions across the Indus basin to explain adequately the observed precipitation variability under the basin-wide experiment. The regionalization scheme successfully clusters 57%, 78%, and 48% of the total recent HA stations (23) around the historical observatories within the identified WS (three), PMS (four), and MS (five) precipitation regions in the UIB. The second regionalization experiment identifies four subregions during each of these seasons and further improves the spatial coverage within the UIB by clustering 70%, 83 %, and 83% of the HA stations around different historic stations. Table 2 presents the statistical performance during both of these regionalization experiments. Both experiments show a good inter- and intra-regional performance in terms of selected statistical attributes with an exception of the MS in the second experiment. Thus, both regionalization experiments validate our assumption of similarities



**Table 2** Summary of precipitation regionalization experiments on seasonal scales. Column 1 represents the precipitation seasons, and column 2 defines the number of precipitation regions. Columns 3 and 4 show sum of squared errors (SSE) within (WI) and between (BW)

regions. Total explained variance is shown in column 5. Columns 6 and 7 define the correlation within (WI) and between (BW) the regions, respectively

Season	Final regions	SSE (WI)	SSE (BW)	EVar (%)	Corr (WI)	Corr (BW)
Basin-wide regionalizat	tion		,	,		,
WS ( Dec-Mar)	6	23.44	130.11	85	0.78	0.46
PMS ( Apr-Jun)	7	48.2	346.54	88	0.82	0.18
MS (Jul-Sep)	7	38.05	139.19	78.6	0.77	0.28
HA-UIB regionalization	n					
WS (Dec-Mar)	4	12.84	55.86	81	0.84	0.48
PMS (Apr–Jun)	4	22.25	121.97	84	0.82	0.36
MS (Jul–Sep)	4	26.93	54.17	70	0.76	0.27

in the precipitation variability between the recent HA and historic stations. Therefore, the historic stations can advantageously be used to infer orographic precipitation over more uncertain HA through these regionalization schemes.

Figures 3, 4, 5, and 6 describe the outcome of the precipitation regionalization process and identify the RR and those HA stations that could not be clustered around any of the surrounding historic stations during both regionalization experiments. Averagely, four different regions have been identified through both of these experiments to represent precipitation dynamics of the UIB. These regions represent precipitation variability of the southern Himalayans, northern Himalayans, and the northwestern parts of the UIB and provide an opportunity for a fine-scale analysis. The identification of such distinct precipitation regions further justifies the need for a sub-regional approach, as such heterogeneity may not be properly reflected by considering UIB as a single unit. Similarly, some of the important LI features (e.g., spate irrigation, irrigated plains, and the coastal belt) are also identified during the basin-wide regionalization experiment. These are important reflectors of the water demand and should simultaneously be considered to demonstrate integrated water resources planning at the basin scale.

Overall, the regionalization schemes successfully capture the orographic heterogeneity within the UIB, show a good regional coherence, and significantly cover the altitudinal perspective within the observational profile during all three seasons. Moreover, our regionalization provides a more realistic basis for understanding subregional precipitation characteristics, when compared with the previously used hydrological units (Dahri et al. 2016) or the administrative units (Iqbal and Athar 2018).

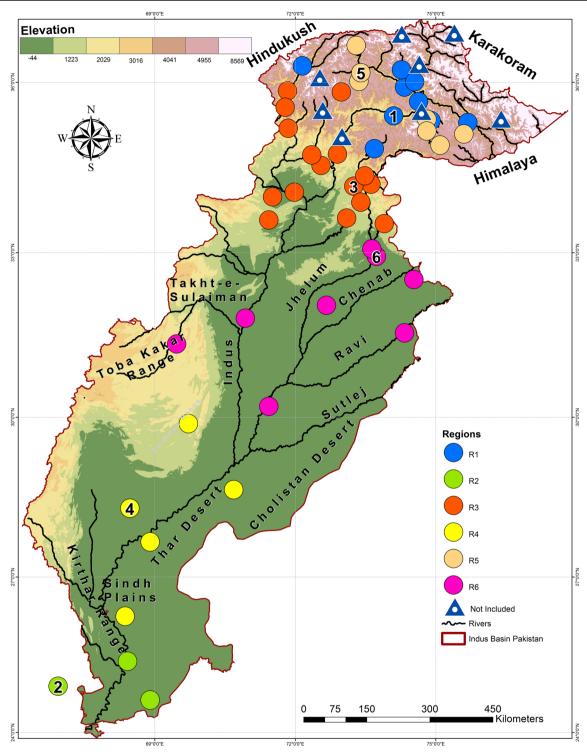
#### 4.2 Seasonal precipitation models

Using the outlined downscaling procedure (Section 3.6), we successfully model observed seasonal precipitation patterns over 29 sub-regions of the study basin under both regionalization experiments. However, we were unable to determine valid models for three smaller regions, representing the coastal belt within the basin-wide experiment (R2 in PMS and R2 in WS) and one UIB region (R3 in MS) during the revised regionalization experiment. This issue partly stems from the relatively large number of dry months (84, 80, and 21, respectively) in the observed time series of RR stations for these regions.

#### 4.2.1 Predictor significance

The final composition of valid seasonal precipitation models (Table 3) indicates a dominant influence of wind predictors in basin-scale precipitation distribution (nearly 80% in total). Particularly, the lower tropospheric winds most effectively resolve the basin-wide regional models (50.9%). The influence of upper tropospheric winds is also significant at the basin scale (25.5 %). Among these wind predictors, the meridional wind velocities in the lower (40%) and upper troposphere (25%) are more prominent in these sub-regional models. However, their effectiveness has a stronger seasonality, where a maximum contribution is realized during the MS (va850 = 52.5%, va200 = 32.5%). After the wind components, the contribution of thermodynamic predictors is at its maximum at the basin scale (about 15%), which also exhibits strong seasonality with the maximum during the PMS (nearly 26%) and WS (about 15 %) and a minimum during the MS





**Fig. 3** The WS (i.e., DJFM) precipitation regionalization under the basin-wide experiment. The colored circles show different precipitation regions with similar co-variability. The numbered circles are the locations of the

RR stations in respective precipitation regions. The triangles define those HA stations that could not be grouped with any of the available historic stations within the UIB

(5%). Among thermodynamic predictors, relative humidity (especially at 1000 hPa level) is the most effective predictor during all precipitation seasons. The lower level dynamic variables (psl, zg700) form the third important predictor set

among regional models (7.6%) and mainly dominate during the WS.

These final predictors capture the seasonally varying regional climatology to explain observed precipitation



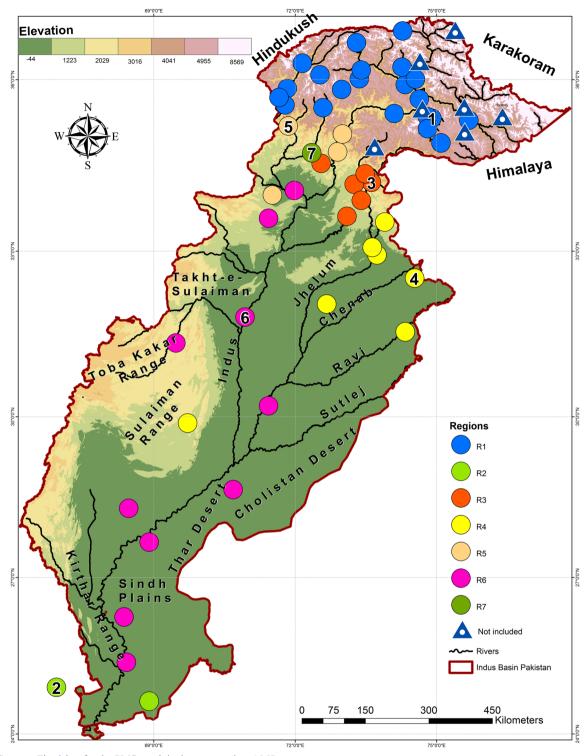


Fig. 4 Same as Fig. 3 but for the PMS precipitation season (i.e., AMJ)

patterns. For instance, the thermally induced local Hadley circulations during the MS are adequately represented by the meridional component of tropospheric winds. Similarly, the atmospheric moisture sources from local evapotranspiration within the basin and advection

from different remote oceanic and terrestrial sources (e.g., the Arabian, Caspian, and Mediterranean Seas) during the PMS and WS play a crucial role in regional precipitation (e.g., Cannon et al. 2015; Curio and Scherer 2016; Mei et al. 2015; Syed et al. 2010). Our



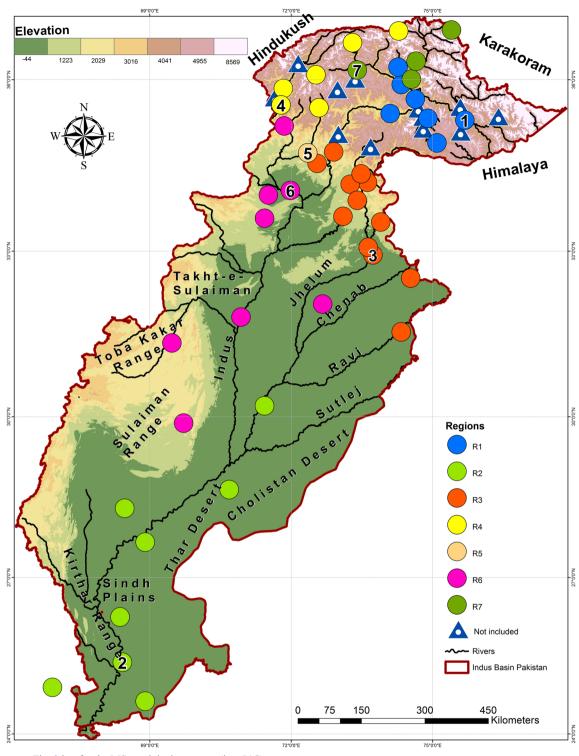


Fig. 5 Same as Fig. 3 but for the MS precipitation season (i.e., JAS)

statistical models represent such moisture influences through the humidity-related predictors particularly during the PMS, when the strength of dynamic forcing reduces over this region. Thus, the identified predictors are representative of the major physical processes that govern the regional precipitation and therefore enhance

the confidence in statistical simulations. Similarly, explanations are also valid for the predictors of the HA-UIB regionalization experiment due to their close similarity with the basin-wide experiment. More details about the physical explanations of the statistical models will follow in Section 5.



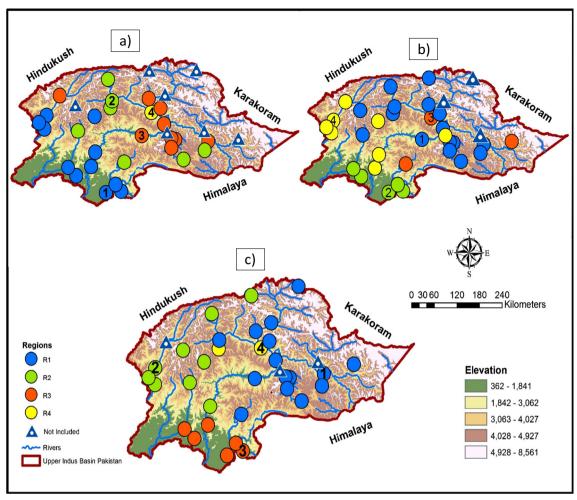


Fig. 6 Same as Fig. 3 but for the HA-UIB precipitation regionalization. (a) WS, (b) PMS, and (c) MS precipitation regions, respectively

#### 4.2.2 Winter season models

The summary of the seasonal modeling process (Table 4) shows that GLM-Tweedie models adequately represent the WS precipitation dynamics across all five precipitation regions of the Indus basin. For information about these regions and the RR stations, please see Fig. 3. Within its three UIB regions (R1, R2, and R5), the mean calibration (validation) root mean squared error (RMSE) amount to 27.14 mm (27.93 mm) with a corresponding MSESS of 41.60% (34.69%). The R3 region (stretching along the foothills of the southern Himalayans and covering the lower Hindukush areas further northwest) contains 16 study stations including some of the HA, and receives the highest precipitation amounts (93 mm per month) among all other UIB regions during the WS. Therefore, precipitation (snow and liquid) within this larger region dominates the river flows during the winter and early spring seasons and its accurate modeling is highly desirable. The RR station strongly represents precipitation characteristics of this region (correlation = 0.93 with regional centroid) and therefore allows to make reliable inferences over a broad altitudinal range (308 to 2744 m) in this region. A single predictor (va850) effectively models the observed precipitation of this RR with a high validation skill (49.57%). Another large UIB region (R1), which mainly covers the Karakorum and trans-Himalayan valleys of the Hindukush, also exhibits an acceptable validation skill (26.1%) despite a large number of dry months. This is mainly a cryosphere-dominated region, which is also strongly represented by its RR station (correlation = 0.83 with regional centroid) and provides inference over the relatively higher altitudes (1251–3200 m) of more uncertain trans-Himalayas. Zonal winds at 700 hPa level (ua700) and specific humidity at 1000 hPa level (hus1000) effectively model the observed precipitation patterns in this region. Similarly, another northern Himalayan region (R5) shows a higher predictability skill (28.39%) despite having the largest numbers of dry months and provides important information for the highest altitudinal bands (2156-4030 m) in our analysis. This region has the largest cryosphere extent, receives mainly solid precipitation and, together with R1, helps to regulate the hydrology of the Indus River. Such a skillful statistical representation of the observed WS precipitation that covers



**Table 3** Frequency of the predictors (in percentage) in final seasonal precipitation models for different sub-regions of the Indus basin of Pakistan. Column 1 shows the large-scale predictors used in this study. The number after each predictor reflects the atmospheric level (pressure level in hPa). Columns 2–4 represent the predictor frequencies in the different seasons. The last column shows the average of the seasonal predictor frequencies

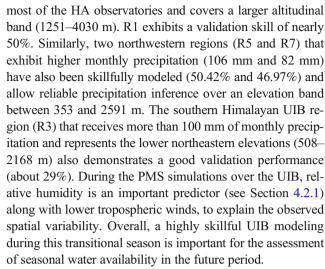
Predictors	WS (%)	PMS (%)	MS (%)	Basin-wide (%)
va200	20.0	19.4	32.5	24
ua200	0.0	0.0	2.5	0.8
zg200	0.0	0.0	0.0	0.0
zg500	0.0	0.0	0.0	0.0
zg700	20.0	0.0	0.0	6.7
hus700	0.0	0.0	2.5	0.8
hur 700	0.0	9.7	0.0	3.2
hur1000	8.6	16.1	2.5	9.1
hus1000	5.7	0.0	0.0	1.9
va500	0.0	0.0	2.5	0.8
ua500	0.0	0.0	2.5	0.8
ua700	14.3	0.0	0.0	4.8
va700	0.0	22.6	0.0	7.5
va850	31.4	12.9	52.5	32.3
ua850	0.0	19.4	0.0	6.5
mslp	0.0	0.0	2.5	0.8
Total	100	100	100	100

the larger spatio-altitudinal scales within the UIB can greatly help in reducing the uncertainty in future projections.

Two LI regions (R4 and R6) also exhibit significant validation skills (25.39 % and 32.89%) for making reliable inferences over diverse conditions of the northeastern rain-fed areas, irrigated plains, and southwestern spate regions. However, precipitation in these regions requires both dynamic and thermodynamic predictors for effective modeling, as shown by the final models (Table 4). The relatively higher error rates in some of the WS cases are discussed in Section 5.3.

#### 4.2.3 Pre-monsoon season models

During the PMS, the GLM-gamma models provide the best modeling performance with an overall skill score of approximately 50% during calibration and 42% in validation (Table 4). Moreover, the average validation performance (43.90%) and the average ability of the RR stations in describing the spatial perspective (correlation coefficient of 0.92) over four different UIB regions (R1, R3, R5, and R7, see also Fig. 4) are at its maximum, compared with other precipitation seasons. Within the UIB, the largest northern Himalayan region (R1) contains 21 study stations including



Additionally, the PMS modeling performance over the two LI regions (R4 and R6) exhibits higher validation skills (40.11% and 37.64%, respectively) to infer precipitation dynamics over the irrigated plains and lower southwestern spate regions. These conditions are simulated by a set of dynamic and thermodynamic predictors.

#### 4.2.4 Monsoon season

Precipitation dynamics are spatially more complex over the UIB during the MS, as it required five different precipitation regions (R1, R3, R4, R5, and R7, see also Fig. 5). Moreover, both the GLM-gamma and Tweedie models are required for statistical simulations of observed precipitation variability over the UIB. Nonetheless, a reasonable average validation performance across all of these UIB regions (32.73%) contributes to confidence in resulting spatial inferences. Within the UIB, maximum validation skill (52.41%) is demonstrated over the wettest southern Himalayan region R3, which receives monthly precipitation sums of 181 mm over a larger vertical horizon (122 to 2591 m). The RR station for this region exhibits a correlation of 0.84 with the regional centroid. The R3 region mainly receives liquid precipitation, which helps to regulate the downstream reservoir operations (e.g., hydropower and irrigation). The modeling process of the two other UIB regions (R1 and R7) also exhibits reasonable validation skills (28.57% and 23.65%, respectively). These regions reflect the highest observed altitudinal range (1251 to 4030 m and 2156 to 4730 m) over the Karakoram and Hindukush areas. Similarly, two northwestern regions (R5 and R4) also demonstrate reasonable validation skills (20.71% and 39%) and represent an altitudinal range from 962 to 3719 m. A strong dynamic forcing during the MS helped to largely resolve the complex MS processes within the UIB, as shown by various predictors of the statistical models (Table 4).



**Table 4** Statistical performance of the seasonal downscaling models over the Indus Basin of Pakistan. Column 1 outlines the identified precipitation region, column 2 indicates the name of the regional representative (RR) station, and its correlation with respective regional centroid is given in the parenthesis. Column 3 (Reg. Alt) indicates the mean regional altitude in meters (m) above mean sea level and in parenthesis the range of regional altitude. Columns 4, 5, 6, 7 and 8 show the adapted GLM, number of dry months in time series of RR

station, number of final predictors, predictor names, and number of PCs in final regression models, respectively. The mean monthly-observed (modeled) precipitation rates of RR station are shown in column 9 (10) respectively. The calibration (validation) root mean squared errors (RMSE) and mean squared error skill scores (MSESS) are reflected in columns 11 (12) and 13 (14) respectively. The blue (green) color indicates UIB (LI) seasonal regions and red color shows the average outcome of seasonal models over the entire Indus basin

WS Mo Region	RR	Reg. Alt	GLM	Dry		Predictor	PCs	Mean	Preci	RN	ISE	MS	ESS
Region	(Corr)	reg. /iii	GLM	Mon		1 redictor	1 03	(mm/			m)		<b>L</b> 55 <b>%</b> )
		(m)		(Nos)	(Nos)	(Name)	(Nos)	Obs	Mod	Cal	Val	Cal	Val
R1	Chilas	2223	Tweedie	39	2	va850+zg700	8	18	17	17.60	18.10	33	26.1
KI	(0.83)	(1251-3200)	1 weedle	37	-	V4030 12g700		10	17	17.00	10.10	33	20.1
R3	Kakul (0.93)	1173.5 (308-2744)	Tweedie	7	1	va850	7	93	93	49.88	50.54	51.92	49.57
R5	Gupis (0.70)	3266 (2156-4030)	Tweedie	56	2	ua700+hus1000	6	11	12	13.95	15.15	39.87	28.39
R4	Jacobabad (0.80)	266 (35-1097)	Tweedie	69	2	va200+hus1000	8	7	7	10.69	11.64	47.86	28.75
R6	Jehlum (0.89)	365 (122-1405)	Tweedie	19	2	hur1000+zg700	6	46	46	37.53	37.98	41.98	36.62
Avg.Basin	( , , ,	(		38	1.8		7.0	35	35	25.93	26.68	42.93	33.89
Avg.UIB				34	1.7		7.0	41	41	27.14	27.93	41.60	34.69
Avg.LI				44	2.0		7.0	27	27	24.11	24.81	44.92	32.69
PMS M	odels												
R1	Astore (0.86)	2627 (1251-4030)	Gamma	0	2	hur700+va850	7	57	57	32.6	34.02	55.44	49.37
R3	Ghari Duputta (0.91)	1327 (508-2168)	Gamma	0	2	va700 + ua850	8	101	101	44.74	48.29	41.14	28.83
R5	Dir (0.94)	1281 (353-2591)	Gamma	0	1	hur1000	2	106	104	45.62	47.63	56.68	50.42
R7	Saidu.Shairf (1)	961 (961)	Gamma	0	2	hur1000+ua850	5	82	80	44.58	45.45	54.53	46.97
R4	Sialkot (0.84)	419 (187-1097)	Gamma	0	1	va200	6	43	42	30.92	31.71	45.98	40.11
R6	DI Khan (0.74)	259 (28-1405)	Tweedie	20	2	hur1000+ua850	3	21	20	17.71	18.23	44.65	37.64
Avg.Basin				3	1.7		5.2	68	67	36.03	37.56	49.74	42.22
Avg.UIB				0	1.8		5.5	87	86	41.89	43.85	51.95	43.90
Avg.LI				10	1.5		4.5	32	31	24.32	24.97	45.32	38.88
MS Mo	dels												
R1	Sakardu (0.74)	2218 (1251-4030)	Tweedie	16	2	va850+hus700	7	13	13	11.32	12.77	44.09	28.57
R3	Jehlum (0.84)	746.25 (122-2591)	Gamma	0	1	va200	7	181	180	76.73	78.63	56.92	52.41
R4	Darosh (0.76)	2868 (1464-3719)	Tweedie	11	1	va200	6	21	21	16.26	16.82	28.82	20.71
R5	Saidu.Shari f (1)	961(961)	Gamma	0	2	va850+hur1000	11	115	116	54.7	60.03	52.71	38.32
R7	Gupis (0.69)	2892 (2156-4730)	Tweedie	20	3	va850+psl+ua200	8	19	20	18.61	20.65	41.69	23.65
R6	Risalpur (0.75)	659 (172-1425)	Gamma	0	2	va700+ua500	7.0	114	115	74.26	76.89	41.13	32.06
R2	Hyderabad (0.88)	52 (9-122)	Tweedie	45	2	ua700+va500	4.0	40	40	45.16	46.58	46.58	40.32
Avg.Basin				8	1.8		7.7	72	72	41.96	44.27	44.23	32.62
Avg.UIB				9	1.8		7.8	70	70	35.52	37.78	44.85	32.73
Avg.LI				23	2.0		5.5	77	78	59.7	61.74	43.86	36.19



In contrast, the homogenous topography and the strong MS influence across the LI helps to relatively better model precipitation over spate and irrigated regions (R4 and R2), with validation skills of 32.06% and 42.32 %, respectively. Spate regions not only provide an abundance of flows during the MS to support diverse livelihood but also cause regular havoc through damage to downstream infrastructure and human lives during extreme events. Therefore, its consideration is important to minimize the socio-economic suffering as well for enhancing national water availability through effective planning.

#### 4.3 HA-UIB modeling

The HA-UIB modeling summary (Table 5) shows some performance compromises. For instance, the mean validation skills over four WS regions show a decrease of 3.5% compared with the corresponding skills in the basin-wide modeling (34.69%). However, this performance loss is primarily due to a poor simulation of a single-station region (R4), which exhibits a lower validation skill (18.42%). The performance for this particular region suffers from a relatively lower precipitation rate (7 mm/month), a higher number of dry months (33), and the presence of some outliers. All other WS regions in the HA-UIB modeling exhibit comparable skills by mainly using different dynamical predictors. However, the trans-Himalayan regions (R2, R4) additionally require specific humidity as a predictor.

Four PMS regions show a very similar average skill (42.62%) compared with the basin-wide modeling (43.90%) despite the presence of some dry months in time series of some RR stations (R3 and R4). Thermodynamic contributions still largely dominate the PMS precipitation simulations (Table 5), where the lower level relative humidity appears as the most influential predictor (i.e., best SPM) in three out of four sub-regional models.

Similarly, the average modeling performance in the MS across its three different precipitation regions also exhibits nearly similar prediction skills (30.60%,  $\sim$  2% less). Still, the dynamic predictors largely dominated these HA-MS simulations as shown by the final statistical models (Table 5), although the model for the northern Himalayan region (R1) also includes the contribution of specific humidity.

Overall, the HA-UIB regionalization significantly improves the spatio-altitudinal representation over the UIB in all seasons, without losing much of the predictability power. This increased spatial coverage helps to infer precipitation dynamics over the hydrologically important yet highly uncertain HA regions within the UIB. The resulting inferences can be significantly trusted for the wetter southern Himalayan regions during all seasons and the northwestern Hindukush and the Karakorum regions during the PMS and WS due to the

stronger representations by the respective RR stations. However, caution is required during the MS particularly over the northwestern Hindukush region (R2) and the Karakoram-central region (R1) due to relatively poor representation by the respective RR stations. The weaker MS penetration into northwestern and Karakoram region could be one of the reasons. Overall, thermodynamic contributions increase in the trans-Himalayan regions, which reflect the role of atmospheric moisture buildup around the UIB from different moisture sources, as demonstrated by various centers of variation for humidity-related predictors (not shown).

## 5 Physical consistency of the statistical relationships: composite analysis

Composites can help to identify atmospheric conditions during different precipitation regimes to evaluate and explain the physical mechanisms of statistical models. The choice of model(s) to be presented in the following section is based on the relative importance for the season (i.e., the core precipitation seasons), the significance of the regional scale, prediction performance of the statistical models, and a distinct domination of a single PC in the final regression models. Using these multiple considerations, we select one MS and one WS region to discuss the atmospheric characteristics underlying the statistical models.

#### 5.1 MS synoptic analysis using composite

During the MS, a larger region (R3) that contains 12 stations and receives a higher amount of precipitation (181 mm/month) is selected for the composite analysis. We use upper level meridional winds (va200) to model the precipitation variations at its RR station, which exhibits a high statistical skill during validation (52.41%). We use PC1, which has a correlation (Spearman) of 0.62 with the observations of RR station to construct composites. The positive loading pattern of PC1 (Fig. 7) stretches over a larger region (from Bay of Bengal to the East African coast) with its center over the Arabian Sea. Positive PC scores indicate drier conditions and vice versa. We select a threshold of  $\pm$  1.5 PC scores, which helps to clearly define these contrasting patterns (Table 6) with ten positive cases (average 69 mm per month) and six negative cases (average 344 mm per month).

The monsoon circulation is characterized by an anticyclone with its center along the southern boundary of the Tibetan Plateau, a tropical easterly jet in the upper troposphere, a westerly jet over the Arabian Sea, and the southwesterly flow over the Bay of Bengal in the lower troposphere (e.g., Krishnamurti 1973; Ashfaq et al. 2017; Duan et al. 2012). More recently, the analysis of extratropical influences on regional MS dynamics has identified an upper level high over



 Table 5
 Same as Table 4, but for HA-UIB regions

WS M	odels	., out for fire-0	12 regions										
M2 IAI	oueis												
Region	RR (Corr)	Reg. Alt	g. Alt GLM Dry Mon Predictors		PCs Mean Preci (mm/mon)			SE m)	MSESS (%)				
		(m)		(Nos)	(Nos)	(Name)	(Nos)	Obs	Mod	Cal	Val	Cal	Val
R1	Kakul (0.95)	1174 (308-2744)	Tweedie	7	1	va850	7	93	93	49.88	50.54	51.92	49.57
R2	Gupis (0.65)	3201 (2156-4030)	Tweedie	56	2	ua700+hus1000	6	12	11	14.11	15.35	39.87	28.39
R3	Chilas (0.78)	2379 (1251-3146)	Tweedie	39	2	va850+zg700	8	17	18	17.67	18.21	33	26.1
R4	Gilgit (1)	1460 (1460)	Tweedie	33	2	va700+hus1000	6	7	7	8.79	8.81	20.85	18.42
Avg.UIB				34	1.8		6.8	32	32	22.60	23.20	36.41	31.12
PMS N	Models												
R1	Astore (0.78)	2896 (1251-4030)	Gamma	0	2	hur700+va850	7	57	57	32.6	34.02	55.44	49.37
R2	Ghari Duputta (0.94)	956 (702-1308)	Gamma	0	2	va700 + ua850	8	101	101	44.74	48.29	41.14	30.83
R3	Gilgit (1)	2874 (1460-3200)	Tweedi	8	2	hur700+ua200	8	21	21	17.21	18.59	43.66	29.54
R4	Chitral (0.91)	2014 (752-3000)	Tweedi	12	2	hur1000+va500	6	43	43	23.28	23.97	68.68	63.74
Avg.UIB				5	2		7.3	56	55	29.46	31.22	52.23	43.37
MS Models													
R1	Sakardu (0.68)	2788 (1251-4730)	Tweedie	16	2	va850+hus700	7	13	13	11.32	12.77	44.09	28.57
R2	Darosh (0.67)	2618 (752-3719)	Tweedie	11	1	va200	6	21	21	16.20	16.79	30.82	20.71
R3	GDP (0.90)	2218 (1251-4030)	Gamma	5	1	ua500	7	186	186	75.31	80.19	52.06	44.72
Avg.UIB				11	1.3		6.7	73	73	34.28	36.58	42.32	30.60



the west central Asia (WCA) as a stronger MS precursor (e.g., Kazmi et al. 2016). In our composite analyses, we analyze the relevant dynamical circulations in the upper and lower tropospheres to explain the associated physical mechanisms.

During the wet MS composite of zg200 (Fig. 8a), a stronger high-pressure anomaly over the WCA and a low-pressure anomaly over the Indian Peninsula reflect amplification of the upper level anticyclone and tropical easterly jet. Strong upper level easterlies (Fig. 8a) point towards a stronger than normal local Hadley circulation. All of these conditions should favor a stronger than normal cross-equatorial moisture flow. Additionally, the circulation anomalies in Fig. 9a exhibit stronger than normal westerlies over the Arabian Sea and south-westerlies over the Bay of Bengal. These upper and lower tropospheric anomalies (Figs. 8c and 9c) in monsoon dynamics point towards a stronger than normal monsoon due to an enhanced moisture supply through oceanic sources and an increased dynamic and orography forcing (Hunt et al. 2018; Ashfaq et al. 2017; Curio and Scherer 2016; Mei et al. 2015; Saeed et al. 2010).

In contrary, the composite during dryer MS (Fig. 8b) exhibits an eastward shift in the upper level anticyclone and a weakening of the tropical easterly jet. The weakening of the easterly jet (Fig. 8b) points towards a weaker than normal local Hadley circulation, which reduces the cross-equatorial flow. Similarly, an eastward shift of the anticyclone is linked with drier than normal conditions over western South Asia and the UIB, and wetter than normal conditions over eastern South Asia (Ashfaq et al. 2009; Soman and Kumar 1993). In the lower levels, both, westerlies over the Arabian Sea and south-westerlies over the Bay of Bengal, exhibit a weakening. It should be noted that the Arabian Sea and Bay of Bengal are important moisture sources for northern and northwestern South Asia (Mei et al. 2015). Therefore, anomalies in the lower level circulations should (Fig. 8c) lead to a decrease in moisture supply from these two oceanic sources, which, when combined with upper level circulation anomalies, lead to a decrease in the MS precipitation particularly over western South Asia that includes the UIB (Wang et al. 2019; Ashfaq et al. 2017; Ahmad et al. 2012).

The composite differences (Fig. 8c) reflect the relative strength of upper level circulations (WCA high, tropical easterly jet) to restrict the interaction of mid-latitude MS currents around UIB and their lower level dynamical signatures (Fig. 9c) to produce an excessive MS precipitation around UIB.

#### 5.2 WS synoptic analysis using composite

Similar to the MS, we construct composites for the WS to analyze the circulation characteristics during the wet and dry precipitation regimes over a selected UIB region. We select region R3 in the WS for the construction of composites, which contains 16 stations and receives a high regional precipitation

(70 mm per month). A single predictor (va850) effectively models the observed precipitation at its RR station with a validation skill of 53%. PC2, which has the highest absolute regression coefficient (0.51) and high co-variability with the RR station (correlation coefficient of 0.65) is used to subsample the composites. The higher positive loadings of PC2 extend over a larger region that stretches from southern Pakistan to Africa and is centered over the Arabian Sea off the East African coast (Fig. 10). Due to a positive regression coefficient and the loading pattern, the higher positive PC scores (Table 7) support above-normal precipitation (12 cases, 187 mm per month), whereas the higher negative PC scores support below-normal precipitation (4 cases, 8 mm per month).

During the WS, the extratropical storms originate mainly at the surface level around the North Atlantic, Europe, and the Mediterranean region, and propagate eastward by prevailing westerlies at mid and upper tropospheric levels to produce precipitation over the UIB (e.g., Martyn 1992; Barlow et al. 2005). Therefore, in the following discussion, we analyze composites of relevant dynamical predictors to describe the anomalous WS precipitation regimes over the UIB.

During a wetter WS, the sea level pressure composites (Fig. 11a) exhibit a higher pressure anomaly over northern Siberia and Europe and a low-pressure anomaly over northern Africa, which supports a strong eastward propagation of the extratropical cyclones, commonly known as the Western Disturbances (Ridley et al. 2013). Such an anomaly is analogous to the pattern that emerges during the negative phase of the North Atlantic Oscillation (NAO) and facilitates cyclogenesis activities around the Mediterranean, Europe, and the Middle East region. These storm tracks are further intensified over the WCA region due to a low-pressure anomaly (Fig 11a), which favors a positive precipitation anomaly over the UIB and surrounding areas (Syed et al. 2006). Moreover, a low sea level pressure anomaly over the Bay of Bengal, eastern Tibetan Plateau and a low-pressure anomaly over the southern Arabian Sea also support a stronger moisture transport from the Arabian Sea towards the UIB—a condition well known during the positive or warm phase of the El Niño-Southern Oscillation (ENSO), as shown by Syed et al. (2006, 2010) and Wang and Xu (2018). A cyclonic circulation anomaly in Fig. 12a and the eastward circulation anomaly over the Mediterranean Sea provide further evidence of excessive precipitation around the UIB. On the other hand, the dryer WS composite (Fig. 11b) exhibits a southward shift of the Siberian high and produces high sea level pressure anomalies over the UIB, Central Asia, and the Mediterranean, and a low sea level pressure anomaly over northern Europe. These conditions are analogous to the positive phase of the NAO and favor a northward movement of the extratropical cyclones tracks during the WS. Consequently, the lack of cyclogenesis activity reduces the strength and moisture transport towards the UIB. Additionally, the moisture supply from the Arabian Sea also



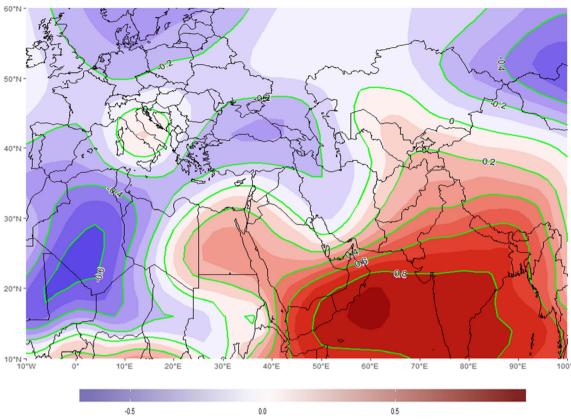


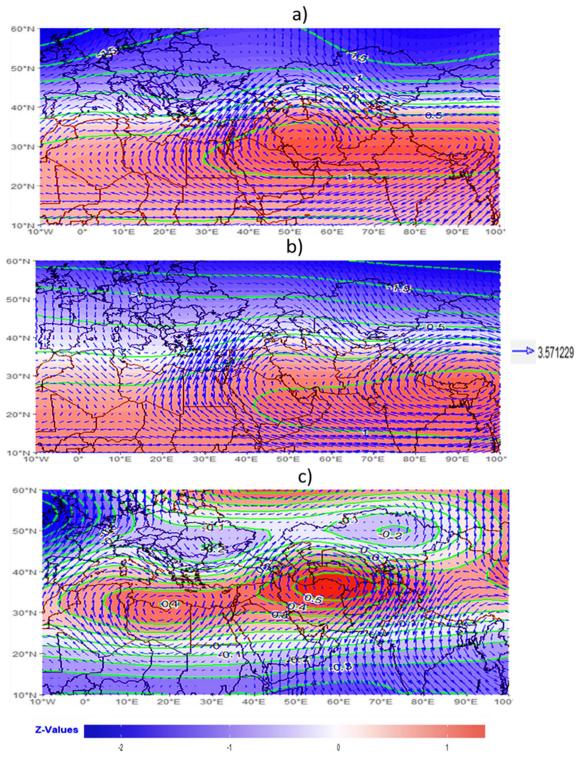
Fig. 7 Loading pattern of PC1-va200 (most influential PC in the final regression model due to the highest absolute regression coefficient) during the MS season. The green contours show labeled loadings over different regions

**Table 6** A statistical summary of the MS composite analysis of a selected UIB precipitation region (R3). The table provides information about the association between observed and modeled precipitation, used GLM framework to identify governing circulations (predictor) and the regression coefficients of different PCs including the intercept term. The

last column gives the significance level of the estimated model coefficients. The information about the most influential PC, its correlation with observations of the RR, threshold of PC scores used to subset contrasting precipitation regimes, and their statistical summaries are also provided in the table

Parameters	Value/description	Model significance level (alpha, $\alpha$ )
Mean observed precipitation rate of the RR (mm/month)	181	NA
Adapted GLM framework	Gamma	
Governing predictor	va200	
Mean modeled precipitation rate of the RR (mm/month)	180	
Correlation between observed and modeled precipitation (Spearman)	0.80	
Regression coefficient (intercept term)	5.05	0.001
Regression coefficient (PC1)	- 0.43	0.001
Regression coefficient (PC2)	-0.27	0.001
Regression coefficient (PC4)	-0.11	0.05
Regression coefficient (PC5)	0.08	1
Regression coefficient (PC7)	-0.10	0.1
Regression coefficient (PC9)	-0.09	0.1
Regression coefficient (PC10)	0.14	0.001
Most influential PC (maximum absolute regression coefficient)	PC1	NA
Correlation between PC1 and RR observations	0.62	
PC scoring threshold for composites	1.50	
No. of positive cases (PC scores $= > 1.50$ )	10	
Mean precipitation per positive event (mm/month)	69	
Dominant month during positive cases	September	
No. of negative cases (PC scores = $< -1.50$ )	6	
Mean precipitation per negative event (mm/month)	344	
Dominant month during negative cases	July	





**Fig. 8** Standardized composite means of geopotential heights and atmospheric winds at 200 hPa (wind vector, m/s) during the MS precipitation season for a sample UIB region (R3). The panel (a) shows the mean state of the large-scale circulations during above-normal

precipitation and (b) represents the conditions during below-normal precipitation over R3. Panel (c) reflects the difference between these two circulations (a minus b)

greatly reduces due to the anomalous northerly winds (Fig. 12b) as also pointed out by Hunt et al. (2018). Wind anomalies over the Mediterranean Sea and land (Fig. 12b) also favor a reduced

moisture flow, eventually resulting in lower than normal precipitation over the UIB as reflected in our corresponding composite. The difference plots (Figs. 11c and 12c) reflect the



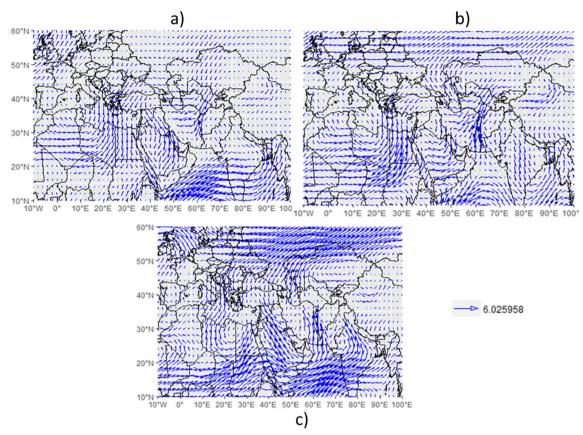


Fig. 9 (a-c) Same as Fig. 8 but for lower tropospheric wind composites at 850 hPa atmospheric level (wind vector, m/s) during the MS precipitation season

influence of relative dynamical forcing supporting a positive precipitation around the UIB.

We also constructed similar composites for the geopotential heights at 700 hPa, 500 hPa, and 200 hPa atmospheric levels to analyze the corresponding circulations as the low-pressure systems travel at high altitudes towards the UIB (Fig. 13). The corresponding difference plots (Fig. 13a-c) confirm the presence of previously described mechanisms and their barotropic nature across the troposphere. It is important to note that some of the earlier studies relate a positive NAO signal with increased winter precipitation over the UIB (e.g., Syed et al. 2010; Ahmad et al. 2015), which is not consistent with our findings. Instead, we see evidence that perhaps the negative NAO supports positive winter precipitation over the UIB. Some other studies such as Archer and Fowler (2004) also point towards a mixed precipitation response during the positive and negative phases of the NAO. Therefore, further investigation is needed to establish a robust understanding of the NAO influence on winter precipitation variability over the UIB.

#### 5.3 Sources of uncertainty

In this study, we have attempted to improve seasonal precipitation assessments within the UIB by addressing many of the common uncertainty sources through methodological improvements. Still, our estimates are based on certain assumptions. For instance, despite significantly improving the HA representation in our station ensemble, altitudes beyond 4730 m (see Table 1) still have no representation in our analysis due to the unavailability of reliable time series data. This lack of representation of higher altitudes has the potential to influence the regionalization process and reliability of precipitation inferences over the UIB. However, it is worth mentioning that the bulk of the UIB precipitation falls between 3000 and 4500 m altitudes (Hewitt 2014), and its magnitude starts decreasing rapidly above 5000 m (Hewitt 2014; Winiger et al. 2005). Thus, it is expected that the regions of maximum precipitation are well represented in our station ensemble (see Section 3.1 and Table 1). Moreover, the mean equilibrium line altitude (ELA) is estimated between 4500 and 5500 m (e.g., Khan et al. 2015b) in the UIB and any precipitation beyond the ELA does not contribute to the runoff. We assume that precipitation variation between 4730 m and the ELA may be represented by our identified UIB regions, and thus, inferences can be extended to fill gaps beyond the available observations. Curio and Scherer (2016) classified the region around the UIB into at least four different precipitation classes to define



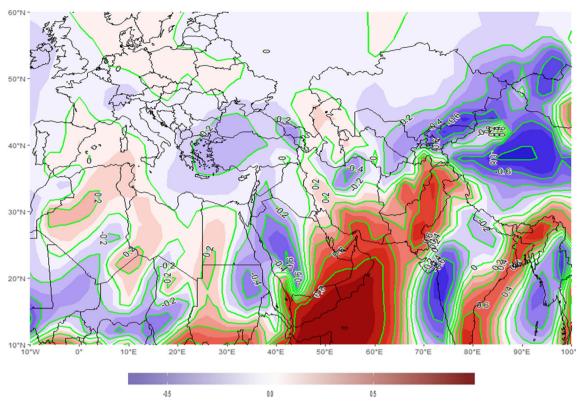


Fig. 10 Loading pattern of PC2-va850 (most influential PC in the final regional model due to the highest absolute regression coefficient) during the WS precipitation. The green contours show labeled loadings over different regions

Table 7 Same as Table 6, but provide information about a selected UIB region (R3) during the WS

Parameters	Value/description	Model significance level (alpha, $\alpha$ )				
Mean observed precipitation rate of the RR (mm/month) Adapted GLM framework	93 Tweedie	NA				
Governing predictor	Va850					
Mean modeled precipitation rate of the RR (mm/month)	93					
Correlation between observed and modeled precipitation (Spearman)	0.78					
Regression coefficient (intercept term)	4.33	0.001				
Regression coefficient (PC2)	0.51	0.001				
Regression coefficient (PC6)	- 0.21	0.001				
Regression coefficient (PC7)	- 0.16	0.01				
Regression coefficient (PC9)	0.23	0.001				
Regression coefficient (PC10)	0.18	0.001				
Regression coefficient (PC13)	- 0.12	0.05				
Regression coefficient (PC15)	- 0.11	0.05				
Most influential PC (maximum absolute regression coefficient) Correlation between PC1 and RR observations	PC2 0.65	NA				
PC scoring threshold for composites	1.50					
No. of positive cases (PC scores $= > 1.50$ )	12					
Mean precipitation per positive event (mm/month)	187					
Dominant month during positive cases	March					
No. of negative cases (PC scores = $< -1.50$ )	4					
Mean precipitation per negative event (mm/month)	8					
Dominant month during negative cases	December					



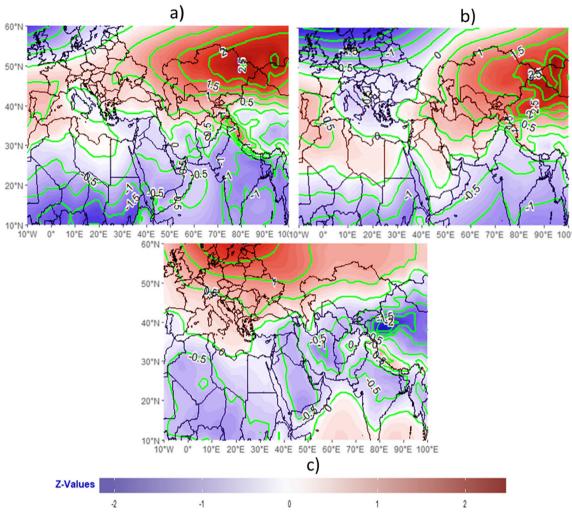


Fig. 11 (a-c) Same as Fig. 9 but for sea level pressure composites during the WS

spatiotemporal precipitation variation using HAR data that covers the entire orography of the UIB. We also identify on average four different regions based upon observations, which are in line with their findings and support our assumption regarding the adequacy of our regionalization in explaining effective orography of the entire UIB. Our study is also limited to only those station observations that are located on the Pakistan side of the UIB. While there are some online data sources (e.g., NOAA-NCDC's data available from https://www.ncdc. noaa.gov) that provide observations of few stations on the Indian side of the UIB, those are mostly located in the low-altitude areas and do not overlap recent time period covered by the HA stations used in this study. Therefore, while it is understandable that additional station observations should improve the spatial coverage of the UIB, substantial effort is needed for the spatiotemporal homogenization of these datasets for drawing meaningful inferences, and particularly to infer orography, which is the focus of our study. We rather assume that precipitation variability explained by our regions may also represent the conditions in surroundings of the UIB of Pakistan, as the same large-scale forcing governs the regional precipitation. Already some studies (e.g., Archer and Fowler 2004) have pointed out that some of the Indian stations show stronger covariation with precipitation of adjacent stations of Pakistan.

Moreover, we have employed ERA-Interim-based predictors in our study, as the more recent ERA5 (Hersbach et al. 2018) was not publicly available at the time when this study was begun. Since statistical downscaling is based on the use of large-scale circulations, we do not expect that the improvements in the horizontal resolution in ERA5 will have a significant impact on the downscaling results. In any case, we evaluate the robustness of ERA-Interim-based final predictors by performing a similar PCA on ERA5 predictors and use Taylor diagrams (Taylor 2001) for the comparison of loading patterns (centers of variation) of the two datasets. We find a high correspondence among dynamical predictors that dominate our final regression models (Section 4.2.1). Thus, the strong



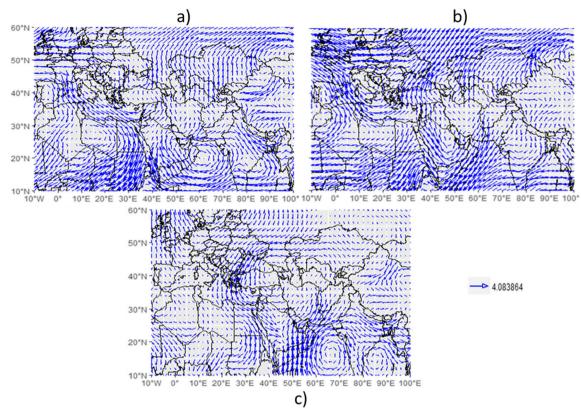
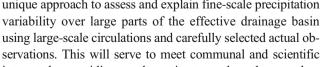


Fig. 12 (a-c) Same as Fig. 9 but during the WS precipitation

correspondence of governing PCs among these different datasets (not shown) provides confidence in the use of ERA-Interim as predictors.

Another shortcoming is related to the RMSEs in Tables 4 and 5, which are sometimes higher than the corresponding precipitation rates. However, the relatively higher RMSEs are associated with a higher number of zeros in the observed time series, which increases the total errors due to the addition of smaller errors in simulating exact zeros in the regression models. In the case of gamma or those Tweedie models that have very few number of months with zero precipitation, the mean errors are within an acceptable range (e.g., in Table 4, PMS-R3 and WS-R3). Therefore, despite relatively higher error rates during some (relatively dry) winter cases, the ability of our models to simulate a large fraction of observed seasonal variance is still significant.

Despite some of these uncertainties, our study presents a unique approach to assess and explain fine-scale precipitation interests by providing an alternative to analyze the complex, climate-sensitive, and yet data-scarce region, within the available observational framework.





#### 6 Summary and conclusions

Complex processes govern spatiotemporal characteristics of regional precipitation over the UIB. An inadequacy of observational coverage within the UIB and regional limitations of various de facto precipitation products further compound the challenge of a reliable estimation of precipitation at varying time scales. In this study, we employed atmospheric circulations within a statistical modeling framework to resolve the observed (seasonal) precipitation over the Indus basin of Pakistan with a focus on the UIB. By taking a distinct advantage of data from recent HA observatories, we adapted a Kmeans cluster analysis to demonstrate the effectiveness of relatively low-altitude stations with historic data to provide precipitation inferences over more uncertain, but hydrologically important, HA of the UIB. This precipitation regionalization scheme also captured the spatial heterogeneity of precipitation over the UIB by identifying more sub-regions within the UIB, when compared with the ones identified over the LI. We argue that the precipitation regionalization is effective in filling out the spatial gaps over the UIB despite the lack of observations at very high altitudes. Furthermore, the RR stations for each of the precipitation regions were carefully identified to serve as predictands. We used different dynamical and thermodynamic variables from the ERA-Interim reanalysis as predictors.

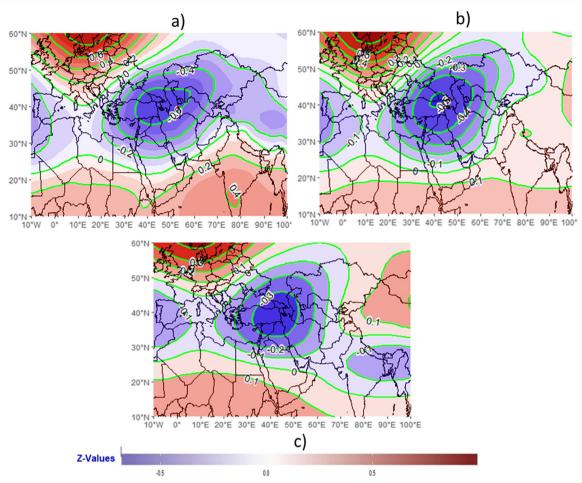


Fig. 13 Composites of geopotential height differences across the troposphere during WS. (a) geopotential heights at 700 hPa, (b) geopotential heights at 500 hPa, and (c) geopotential heights at 200 hPa

We adapted a GLM framework with gamma and Tweedie distributions to model the typical predictorpredictand relationships across the Indus River basin during the major precipitation seasons. The final selection of precipitation models was based on minimizing the MSE (maximizing the MSESS) by duly considering the multicollinearity among different predictors within a robust cross-validation framework. Overall, the precipitation models exhibited a high performance over the wetter southern Himalayans and various LI regions by mainly using different dynamical predictors. The typical modeling framework also demonstrated an adequate performance in resolving the seasonal precipitation of cryospheredominated and topographically complex trans-Himalayans regions, which largely govern the hydrological regime of the Indus River system. Knowing that highly localized processes are difficult to model and also govern some part of the mountainous precipitation variability, the modeled skills over different parts of the UIB are quite reasonable to explain the precipitation dynamics using large-scale atmospheric circulations.

A skillful modeling of the trans-Himalayan regions, particularly during the westerly dominated circulations (PMS and WS) which mainly nourish the seasonal snowpacks, should help to improve our limited understanding about the spatial characteristics of regional precipitation and to assess the future stability of associated cryosphere. However, relatively complex models containing both dynamic and thermodynamic predictors were required to simulate such cryosphere-dominated trans-Himalayan regions. Moreover, we also separately modeled the HA part of the UIB due to its importance in the basin hydrology and associated greater uncertainty. We also demonstrated the robustness of ERA-Interim circulations against the latest ERA5 reanalysis, and the insensitivity of the final precipitation models against different precipitation measurement losses.

Furthermore, we constructed composites for two major wet seasons (MS and WS) to analyze and explain the large-scale atmospheric structures during the wet and dry precipitation anomalies over the UIB. These composites provided additional confidence regarding the reliability of the statistical models by explaining the physical processes that govern precipitation



dynamics over the complex UIB. We showed that the strength and location of an upper level high over the WCA (Siberian high) influence the anomalous behavior of UIB precipitation by triggering different dynamical and thermodynamic instabilities during the MS (WS).

We have also simultaneously modeled the precipitation distribution over different LI regions in a similar way to have a coherent perspective about the water supply (UIB) and its demand (LI). Such investigations will further help in devising appropriate strategies for water resources planning at the basin scale. These statistically skillful and physically consistent models also offer an opportunity to investigate the sub-regional response of the Indus basin under different climate scenarios by downscaling relevant circulations from the GCMs. Given that the latest GCMs tend to exhibit a relatively better skill in representing large-scale circulations when compared with their raw precipitation output (e.g. Mueller and Seneviratne 2014), future precipitation changes, derived through such predictors, may provide more realistic policy feedback to support the regional adaptations.

However, considering melt-dominated hydrological regime of the Indus basin, the downscaling of precipitation alone cannot fulfill the requirements for understanding the basin hydrology. Therefore, similar efforts are needed to downscale regional temperatures. Thus, we plan to apply similar statistical downscaling techniques on the temperature fields in addition to the use of these models for precipitation projections amid different climate scenarios.

**Acknowledgments** The authors would like to thank PMD and WAPDA for sharing the meteorological data. The authors also acknowledge the ECMWF for the provision of the ERA-Interim and ERA5 datasets. This manuscript has been co-authored by an employee of Oak Ridge National Laboratory, managed by UT Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy.

Funding information Open Access funding provided by Projekt DEAL. The Himalayan Adaptation, Water and Resilience (HI-AWARE) consortium through PARC under the Collaborative Adaptation Research Initiative in Africa and Asia (CARIAA) under financial support from UK-ID and IDRC, Canada, mainly funded this work. The DAAD Germany and the University of Augsburg provided additional funding, which also need due appreciation. M. A. was supported by the National Climate-Computing Research Center, which is located within the National Center for Computational Sciences at the Oak Ridge National Laboratory (ORNL) and supported under a Strategic Partnership Project, 2316-T849-08, between DOE and NOAA. E.H. was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under project number 408057478.

Data availability The reanalysis data used in our study can be assessed through <a href="https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim">https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim</a>. However, the precipitation data of the stations used in our analysis can be requested directly from the WAPDA and PMD, as this is not publically available. The US department of Energy will provide

public access to these results in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

#### **Compliance with ethical standards**

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a>.

#### References

- Ahmad S, Kazuaki N, Tamura T, Ohta T, Ikoma E, Kitsuregawa M, Koike T (2012) Characteristics of climatological troposheric conditions during pre-monsoon and matured phases of Pakistan summer monsoon. Ann J Hydraulic Eng 56(4):I\_157–I\_162. https://doi.org/10.2208/jscejhe.68.I 157
- Ahmad I, Ambreen R, Sun Z, Deng W (2015) Winter-spring precipitation variability in Pakistan. Am J Clim 4:115–139. https://doi.org/10.4236/ajcc.2015.41010
- Akaike H (1973) Information theory as an extension of the maximum likelihood principle. Second International Symposium on Information Theory, Akademiai Kiado, pp. 267-281
- Akhtar M, Ahmad N, Booij MJ (2008) The impact of climate changes on the water resources of Hindukush-Karakorum-Himalayaregionunder different glacier coverage scenarios. J Hydrol 355, pp. 148–163. Retrieved from https://doi.org/10.1016/j.jhydrol.2008.03.015
- Alexandersson H (1986) A homogeneity test applied to precipitation data. Int J Climatol. 6, pp. 661–675. Retrieved from https://doi.org/10.1002/joc.3370060607
- Ali S, Li D, Congbin F, Khan F (2015) Twenty first century climatic and hydrological changes over Upper Indus Basin of Himalayan region of Pakistan. Environ Res Lett 10(014007):1–20. https://doi.org/10.1088/1748-9326/10/1/014007
- Anderson TW, Darling DA (1952) Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes. Ann Math Stat 23:193–212
- Archer DR, Fowler HJ (2004) Spatial and temporal variations in precipitation in the Upper Indus Basin, global teleconnections and hydrological implications. Hydrol Earth Syst Sci 8(1):47–61. https://doi.org/10.5194/hess-8-47-2004
- Ashfaq M, Shi Y, Tung WW, Trapp RJ, Gao XJ, Pal JS, Diffenbaugh NS (2009) Suppression of south Asian summer monsoon precipitation in the 21st century. Geophys Res Lett 36. https://doi.org/10.1029/2008gl036500
- Ashfaq M, Rastog D, Mei R, Toumag D, Leung LR (2017) Sources of errors in the simulation of south Asian summer monsoon in the CMIP5 GCMs. Clim Dyn 49:193–223. https://doi.org/10.1007/s00382-016-3337-7



- Barlow M, Wheeler M, Lyon B, Cullen H (2005) Modulation of daily precipitation over southwest Asia by the Madden-Julian oscillation. Mon Weather Rev 133:3579–3594
- Bashir F, Zeng X, Guptta H, Hazenberg P (2017) A Hydrometeorological perspective on the Karakoram anomaly using unique valley-based synoptic weather observations. Geophys Res Lett 44(20):1–9. https://doi.org/10.1002/2017GL075284
- Bhambri R, Bolch T, Kawishwar P, Dobhal DP, Srivastava D, Pratap B (2013) Heterogeneity in glacier response in the upper Shyok valley, northeast Karakoram. Cryosphere 7:1385–1398. https://doi.org/10. 5194/tc-7-1385-2013
- Cannon F, Carvalho LMV, Jones C, Norris J (2015) Winter westerly disturbance dynamics and precipitation in the western Himalaya and Karakoram: a wave-tracking approach. Theor Appl Climatol 125:27–44. https://doi.org/10.1007/s00704-015-1489-8
- Cogley JG (2011) Present and future states of Himalaya and Karakoram glaciers. Ann Glaciol 52(59):69–73
- Curio J, Scherer D (2016) Seasonality and spatial variability of dynamic precipitation controls on the Tibetan Plateau. Earth Syst Dynam (7): 767–782. https://doi.org/10.5194/esd-7-767-2016
- Dahri ZH, Ludwig F, Moors E, Ahmad B, and Kabat P (2016) An appraisal of precipitation distribution in in the high altitude catchments of the Indus basin. Sci Total Environ. 548-549, pp. 289-306. Retrieved from https://doi.org/10.1016/j.scitotenv.2016.01.001, 2016
- Dahri ZH, Moors E, Ludwig F, Ahmad S, Khan A, Ali I, Kabat P (2018) Adjustment of measurement errors to reconcile precipitation distribution in the high-altitude Indus basin. Int J Climatol 2018(38): 3842–3860. https://doi.org/10.1002/joc.553
- Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer A (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorol Soc 137:553–597. https://doi.org/10.1002/qj.828,2011. 4757,4759,4775
- Duan AM, Wu GX, Liu YM, Ma YM, Zhao P (2012) Weather and climate effects of the Tibetan Plateau. Adv Atmos Sci 29:978–992
- Dunn PK (2004) Occurrence and quantity of precipitation can be modelled simultaneously. Int J Climatol 24:1231–1239
- Dunn PK (2010) Tweedie: Tweedie exponential family models. R package. Version 2.0.5.
- FAO (2011) Irrigation in Southern and Eastern Asia in figures AQUASTAT Survey 2011. Retrieved from http://www.fao.org/nr/Water/aquastat/basins/indus/indus-CP eng.pdf
- Filippi L, Palazzi E, von Hardenberg J, Provenzale A (2014) Multidecadal variations in the relationship between the NAO and winter precipitation in the Hindu-Kush Karakoram. *J Clim* Retrieved from https://doi.org/10.1175/JCLI-D-14-00286.1, 27, 7890, 7902
- Hasan MM, Dunn PK (2011) Two Tweedie distributions that are nearoptimal for modelling monthly rainfall in Australia. Int J Climatol 31:1389–1397
- Hasson S (2016) Future water availability from Hinukush-Karakoram-Himalya Upper Indus Basin under conflicting climate change signals. Climate 4(3):40. https://doi.org/10.3390/cli4030040
- Hasson S, Böhner J, Chishtie F (2019) Low fidelity of CORDEX and their driving experiments indicates future climatic uncertainty over Himalayan watersheds of Indus basin. Clim Dyn 52:777–798. https://doi.org/10.1007/s00382-018-4160-0
- Hersbach H, de Rosnay P, Bell B, Schepers D, Simmons A, Soci C, Abdalla S, Alonso-Balmaseda M, Balsamo G, Bechtold P, Berrisford P, Bidlot JR, de Boisséson E, Bonavita M, Browne P, Buizza R, Dahlgren P, Dee, Dragani R, Diamantakis M, Flemming J, Forbes R, Geer A, Haiden T, Hólm E, Haimberger L, Hogan R, Horányi A, Janiskova M, Laloyaux P, Lopez P, Munoz-Sabater J,

- Peubey C, Radu R, Richardson D, Thépaut JN, Vitart F, Yang X, Zsótér E, and Zuo H (2018) Operational global reanalysis: progress, future directions and synergies with NWP, ERA-report, Serie 27, 10.21957/tkic6g3wm, https://www.ecmwf.int/node/18765.
- Hertig E, Jacobeit J (2015) Considering observed and future nonstationarities in statistical downscaling of Mediterranean precipitation. Theor Appl Climatol 122:667–683
- Hertig E, Merkenschlager C, Jucundus J (2017) Change points in predictors–predictand relationships within the scope of statistical downscaling. Int J Climatol 37:1619–1633. https://doi.org/10. 1002/joc.4801
- Hewitt K (2005) The Karakoram anomaly? Glacier expansion and the 'elevation effect', Karakoram Himalaya. Mt Res Dev (25), pp. 332–340. Retrieved from https://doi.org/10.1659/02764741(2005) 025[0332:TKAGEA]2.0.CO;2
- Hewitt K (2011) Glacier change, concentration, and elevation effects in the Karakoram Himalaya, Upper Indus Basin. Mt Res Dev 31 (3), pp. 188–200. Retrieved from https://doi.org/10.1659/mrd-journal-d-11-00020.1
- Hewitt K (2014) Glaciers of the Karakoram Himalaya. Springer, p 363
  Huffman GJ, Adler RF, Bolvin DT, Gu G, Nelkin EJ, Bowman KP, Hong Y, Stocker EF, Wolff DB (2007) The TRMM multisatellite precipitation analysis (TMPA): quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. J Hydrometeorol 8:38–55. <a href="https://doi.org/10.1175/JHM560.1">https://doi.org/10.1175/JHM560.1</a>
- Hunt KMR, Turner AG, Shaffrey LC (2018) Extreme daily rainfall in Pakistan and north India: scale interactions, mechanisms, and precursors. Amer Meteo Soc 146:1005–1022. https://doi.org/10.1175/ MWR-D-17-0258
- Immerzeel WW, Pellicciotti F, Bierkens MFP (2013) Rising river flows throughout the twenty-first century in two Himalayan glacierized watersheds. Nat Geosci 6:742–745. https://doi.org/10.1038/ NGEO1896
- Immerzeel WW, Wanders N, Lutz AF, Shea JM, Bierkens PFM (2015) Reconciling high-altitude precipitation in the upper Indus basin with glacier mass balances and runoff. Hydrol Earth Syst Sci 19:4673– 4687. https://doi.org/10.5194/hess-19-4673-2015
- Iqbal MF, Athar H (2018) Variability, trends, and teleconnections of observed precipitation over Pakistan. Theor Appl Climatol 134: 613–632. https://doi.org/10.1007/s00704-017-2296-1
- Jacob T, Wahr J, Pfeffer WT, Swenson S (2012) Recent contributions of glaciers and ice caps to sea level rise. Nature 482:514–518
- Kääb A, Berthier E, Christopher N, Gardelle J, Arnaud Y (2012) Contrasting patterns of early twenty-first-century glacier mass change in the Himalayas. Nature 488(7412):495–498
- Kääb A, Treichler D, Nuth C, Berthier E (2015) Brief Communication: contending estimates of 2003–2008 glacier mass balance over the Pamir–Karakoram–Himalaya. Cryosphere 9, pp. 557–564. Retrieved from https://doi.org/10.5194/tc-9-557-2015
- Kapnick SB, Thomas LD, Moetasim A, Segey M (2014) Snowfall less sensitive to warming in Karakoram than in Himalayas due to a unique seasonal cycle. Nat Geosci 7(11):834–840. https://doi.org/ 10.1038/ngeo2269
- Kaspar-Ott I, Elke H, Seerin K, Felix P, Christoph R, Heiko P, Jucunduss J (2019) Weights for general circulation models from CIMP3/CIMP5 in a statistical downscaling framework and the impact on future Mediterranean precipitation. Int J Climatol 39:3639–3654. https://doi.org/10.1002/joc.6045
- Kazmi DH, Li J, Ruan C, Zaho S, Sen Z, Li Y (2016) A statistical downscaling model for summer rainfall over Pakistan. Clim. Dyn. 47, Jianping Li2,3 · Chengqing Ruan4 · Sen Zhao1,5 · Yanjie Li1, 2653-2666. https://doi.org/10.1007/s00382-016-29990-1
- Khan A, Naz BS, Bowling LC (2015a) Separating snow, clean and debris covered ice in the Upper Indus Basin, Hindukush–Karakoram– Himalayas, using Landsat images between 1998 and 2002. J Hydrol 521:46–64



- Khan F, Pilz J, Amjad M, Wiberg D (2015b) Climate variability and its impacts on water resources in the Upper Indus Basin under IPCC climate change scenarios. Int J Global Warming 8(1):46–69. https:// doi.org/10.1504/IJGW.2015.071583
- Khattak MS, Babel MS, Sharif M (2011) Hydro-meteorological trends in the upper Indus River basin in Pakistan. Clim.Res. 46(103):103– 119. https://doi.org/10.3354/cr00957
- Krishnamurti TN (1973) Tibetan high and upper tropospheric tropical circulation during northern summer. Bull Amer Meteor Soc 54: 1234–1249
- Latif Y, Yaoming M, Yaseen M (2018) Spatial analysis of precipitation time series over the Upper Indus basin. Theor Appl Climatol, 131(1-2), pp. 761-775. Retrieved from https://doi.org/10.1007/s00704-016-2007-3
- Lutz AF, Immerzeel WW, Kraaijenbrink PDA, Shrestha AB, Bierkens MFP (2016) Climate change impacts on the upper Indus hydrology: sources, shifts and extremes. PLoS One 11(11):e0165630. https://doi.org/10.1371/journal.pone.0165630
- Mahmood R, Babel MS (2012) Evaluation of SDSM developed by annual and monthly sub-models for downscaling temperature and precipitation in the Jhelum basin, Pakistan and India. Theor Appl Climatol 113:27–44. https://doi.org/10.1007/s00704-012-0765-0
- Martyn D (1992) Climate of the World. Elsevier, New York
- Maussion F, Scherer D, Mölg T, Collier E, Curio J, Finkelnburg R (2014) Precipitation seasonality and variability over the Tibetan Plateauas as resolved by the High Asia Reanalysis. J Clim 27:1910–1927. https://doi.org/10.1175/JCLI-D-13-00282.1
- Mc Cullagh P, Nelder JA (1989) Generalized linear models. Monographs on statistics and applied probability, vol (37). Chapman & Hall, London
- Mei R, Ashfaq M, Rastogi D, Leung LR, Dominguez F (2015) Dominating controls for wetter south Asian summer monsoon in the twenty-first century. J Clim 28(8):3400–3419. https://doi.org/ 10.1175/JCLI-D-14-00355.1
- Minora U, BocchiolaD D, Agata C, Maragno D, Mayer C, Lambrecht A, Mosconi B, Vuillermoz E, Senese A, Compostella C, Simiraglia C, Diolaiuti G (2013) 2001–2010 glacier changes in the Central Karakoram National Park: a contribution to evaluate the magnitude and rate of the "Karakoramanomaly". CryosphereDiscuss 7(3): 2891–2941. https://doi.org/10.5194/tcd-7-2891-2013
- Mishra V (2015) Climatic uncertainty in Himalayan water towers. J Geophys Res-Atmos 120:2689–2705. https://doi.org/10.1002/2014JD022650
- Moberg A, Jones PD, Lister D, Walther A, Brunet M, Jacobeit J, Alexander LV, Della-Marta P, Luterbacher J, Yiou P, Chen D, Klein TAMG, Saladié OSJ, Aguilar E, Alexandersson H, Almarza C, Auer I, Barriendos M, Begert M, Bergström H, Böhm R, Butler CJ, Caesar J, Drebs A, Founda D, Gerstengarbe FW, Micela G, Maugeri M, Österle H, Pandzic K, Petrakis M, Srnec L, Tolasz R, Tuomenvirta H, Werner PC, Linderholm H, Philipp A, Wanner H, Xoplaki E (2006) Indices for daily temperature and precipitation extremes in Europe analyzed for the. J Geophys Res Atmos 111(D22). https://doi.org/10.1029/2006JD007103
- MRI (2015) Elevation-dependent warming in mountain regions of the world. Nat Clim Chang 5:424–430
- Mueller B, Seneviratne SI (2014) Systematic land climate and evapotranspiration biases in CMIP5 simulations. Geophys Res Lett 41: 128–134. https://doi.org/10.1002/2013GL058055
- Nepal S, Shrestha AB (2015) Impact of climate change on the hydrological regime of the Indus, Ganges and Brahmaputra river basins: a review of the literature. Int J Water Resour Dev 627:1–18. https://doi.org/10.1080/07900627.2015.1030494
- Palazzi E, von Hardenberg J, Provenzale A (2013) Precipitation in the Hindu-Kush Karakoram Himalaya: observations and future scenarios. J Geophys Res Atmos 118(1):85–100. https://doi.org/10.1029/ 2012JD018697

- Palazzi E, von Hardenberg J, Terzago S, Provenzale A (2015)
  Precipitation in the Karakoram-Himalaya: a CMIP5 view. Clim
  Dvn 45(1–2):21–45
- Philipp A (2003). Zirkulationsdynamische Telekonnektivität des Sommerniederschlags im südhemisphärischen Afrika. Dissertation, Bayerische Julius-Maximilians-Universität Würzburg
- Preisendorfer R (1988) Principal component analysis in meteorology and oceanography, vol 42. Elsevier, Amsterdam
- Pritchard DM, Forsythe N, Fowler HJ, O'Donnell GM, Li XF (2019) Evaluation of Upper Indus near-surface climate representation by WRF in the high Asia refined analysis. J Hydrometeorol <u>20:467</u>– 487. https://doi.org/10.1175/JHM-D-18-0030.1
- Qiu J (2008) China: the third pole, climate change is coming fast and furious to the Tibetan Plateau. Nature 454:393–396. https://doi.org/10.1038/454393a
- Rasmussen R, Baker B, Kochendorfer J, Meyers T, Landolt S, Fischer AP, Black J, Theriault JM, Kucera P, Gochis D, Smith C, Nitu R, Hall M, Ikeda K, Gutmann E (2012) How well are we measuring snow: the NOAA/FAA/NCAR winter precipitation test bed. Bull Am Meteorol Soc 93:811–829. https://doi.org/10.1175/BAMS-D-11-00052.1
- Rastogi D, Ashfaq M, Leung L, Ghosh S, Saha A, Hodges K, Evans K (2018) Characteristics of Bay of Bengal monsoon depressions in the 21st century. Geophys Res Lett 45(13):6637–6645. https://doi.org/10.1029/2018GL078756
- Ridley J, Wiltshire A, Mathison C (2013) More frequent occurrence of westerly disturbances in Karakoram up to 2100. https://doi.org/10. 1016/j.scitotenv.2013.03.074
- Saeed S, Müller WA, Hagemann S, Daniela J (2010) Circumglobal wave train and the summer monsoon over northwestern India and Pakistan: the explicit role of the surface heat low. Clim Dyn 37: 1045–1060. https://doi.org/10.1007/s00382-010-0888-x
- Smirnov N (1939) On the estimation of discrepancy between empirical curves of distribution for two independent samples. Bulletin Mathematique de 'Universite de Moscou 2:3–14
- Soman MK, Kumar KK (1993) Space-time evolution of meteorological features associated with the onset of Indian summer monsoon. Mon Weather Rev 121(4). https://doi.org/10.1175/1520-0493(1993) 121<1177:STEOMF>2.0.CO;2
- Soncini A, Daniele B, Gabriele C, Alberto B, Renzo R, Christoph M, Astrid L, Elissa P, Claudio S, Guglielmina D (2015) Future hydrological regimes in the upper indus basin: a case study from a high-altitude glacierized catchment. J Hydrometeorol 16(1):306–326. https://doi.org/10.1175/JHM-D-14-0043.1
- Stone M (1979) Comments on model selection criteria of Akaike and Schwarz. J R Stat Soc Ser B Stat Methodol 41:276–278
- Syed F, Giorgi F, Pal JS, King MP (2006) Effect of remote forcings on the winter precipitation of central southwest Asia part 1: observations. Theoret Appl Climatol 86(1–4):147–160. https://doi.org/10.1002/ joc.1887
- Syed FS, Giorgi F, Pal JS, Keay K (2010) Regional climate model simulation of winter climate over Central–Southwest Asia, with emphasis on NAO and ENSO effects. Int J Climatol 30:220–235. https:// doi.org/10.1002/joc.1887
- Syed FS, Iqbal W, Syed AAB, Rasul G (2014) Uncertainties in the regional climate models simulations of south-asian summer monsoon and climate change. Clim Dyn 42(7–8):2079–2097
- Tahir AA, Chevallier P, Arnaud Y, Ahmad B (2011) Snow cover dynamics and hydrological regime of the Hunza River basin, Karakoram Range. Northern Pakistan Hydrol Earth Syst Sci 15(7):2275–2290
- Tahir AA, Chevallier P, Arnaud Y, Ashraf M, Bhatti MT (2014) Snowcover trend and hydrological characteristics of the Astore River basin (Western Himalayas) and its comparison to the Hunza basin (Karakoram region). Sci Total Environ 505, pp. 748–761. Retrieved from https://doi.org/10.1016/j.scitotenv.2014.10.065



- Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res Atmos 106(D7):7183–7192. https://doi.org/10.1029/2000JD900719 (ISSN 2156-2202)
- Turner AG, Annamalai H (2012) Climate change and the South Asian summer monsoon. Nat Clim Chang 2:587–595. https://doi.org/10.1038/nclimate1495
- Wang Y, Xu X (2018) Impact of ENSO on the thermal condition over the Tibetan Plateau. J Meteo Soc Japan 96(3):269–281. https://doi.org/10.2151/jmsj.2018-032
- Wang Z, Yang S, Duan A, Hua W, Ullah K, Liu S (2019) Tibetan Plateau heating as a driver of monsoon rainfall variability in Pakistan. Clim Dyn, (52), pp. 6121–6130. Retrieved from https://doi.org/10.1007/ s00382-018-4507-6
- Wijngaard JB, Klein TAMG, Können GP (2003) Homogeneity of 20th century European daily temperature and precipitation series. Int J

- Climatol (23), pp. 679–692. Retrieved from https://doi.org/10.1002/joc.906
- Wilks S (2006) Statistical methods in the atmospheric sciences. International Geophsical Series, Vol. 91
- Wiltshire AJ (2014) Climate change implications for the glaciers of the Hindu Kush, Karakoram and Himalayan region. Cryosphere 8:941–958. https://doi.org/10.5194/tc-8-941
- Winiger M, Gumpert M, Yamout H (2005) Karakorum–Hindukush—western Himalaya: assessing high-altitude water resources. Hydrol Process 19 (12), pp. 2329–2338. Retrieved from https://doi.org/10.1002/hyp.5887

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

