APPLICATION PAPER

Precipitation prediction over the upper Indus Basin from large-scale circulation patterns using Gaussian processes

Kenza Tazi[1,2] ⓘ, Andrew Orr[2] ⓘ, J. Scott Hosking[2,3] ⓘ and Richard E. Turner[1,3]

[1]Department of Engineering, University of Cambridge, Cambridge, UK
[2]British Antarctic Survey, Cambridge, UK
[3]The Alan Turing Institute, London, UK
Corresponding author: Kenza Tazi; Email: kt484@cam.ac.uk

Abstract

Water resources from the Indus Basin sustain over 270 million people. However, water security in this region is threatened by climate change. This is especially the case for the upper Indus Basin, where most frozen water reserves are expected to decrease significantly by the end of the century, leaving rainfall as the main driver of river flow. However, future precipitation estimates from global climate models differ greatly for this region. To address this uncertainty, this paper explores the feasibility of using probabilistic machine learning to map large-scale circulation fields, better represented by global climate models, to local precipitation over the upper Indus Basin. More specifically, Gaussian processes are trained to predict monthly ERA5 precipitation data over a 15-year horizon. This paper also explores different Gaussian process model designs, including a non-stationary covariance function to learn complex spatial relationships in the data. Going forward, this approach could be used to make more accurate predictions from global climate model outputs and better assess the probability of future precipitation extremes.

Impact Statement

Future precipitation in the upper Indus Basin and, by extension, the water security of downstream communities remains difficult to predict. New approaches quantifying the probabilities of extreme precipitation events, which lead to floods or droughts, are needed to inform more sustainable policy and climate change adaptation. This paper presents one such method.

1. Introduction

The Indus is one of the longest rivers in Asia. From the Tibetan Plateau, it meanders through northern India and along the entire length of Pakistan before flowing into the Arabian Sea. Over its 3.000 km journey shown in Figure 1, it sustains the livelihoods of more than 270 million people, including the world's largest contiguous irrigation system (Shrestha et al., 2019). Of all the rivers originating in the Hindu-Kush Karakoram Himalaya (HKKH) region, the Indus depends the most on high-altitude water resources (Immerzeel et al., 2010; Lutz et al., 2014). Over 60% of its total flow is attributed to glacier and snow
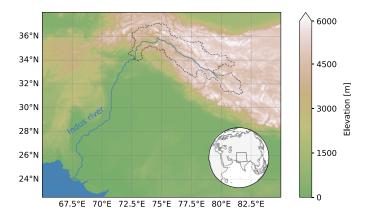

Figure 1. *Map of Indus Basin showing elevation (TROPOMI, 2019), the Indus River (Kelso and Patterson, 2010, in blue) and the UIB (NASA JPL, 2013, watershed boundary shown with dashed line). The inset shows the watershed's location with respect to the Asian continent.*

melt in an area known as the Upper Indus Basin (UIB) (Immerzeel and Bierkens, 2012; Lutz et al., 2016). The Indus River is therefore one of the most sensitive to climate change. However, there is considerable uncertainty around the future amount, seasonality, and composition of the Indus' flow (Krishnan et al., 2019; Lutz et al., 2016; Orr et al., 2022; Sabin et al., 2020). This includes predictions of extremes such as floods and droughts. Coupled with an increasing demand for freshwater, future water security in this area is at risk. This security is predominantly undermined by the river's largest source of current and future hydrological uncertainty: precipitation (Immerzeel et al., 2020; Orr et al., 2022; Smolenaars et al., 2021).

In the UIB, precipitation largely originates from two major atmospheric events: the Indian summer monsoon (ISM) and the western disturbances (WDs). The ISM brings precipitation from June to September through the displacement of humid air from the Indian Ocean, and in particular the Bay of Bengal (Bookhagen and Burbank, 2010). The strength of the ISM is primarily modulated by the El Niño-Southern Oscillation (Ding and Wang, 2005; Di Capua et al., 2020). WDs are extratropical storms that originate over the Mediterranean region. They travel as frontal systems, collecting moisture from the Mediterranean, Black, and Caspian seas until they intensify upon reaching South Asia (Midhuna et al., 2020). They are primarily modulated by the North Atlantic Oscillation and El Niño-Southern Oscillation (Hunt and Zaz, 2023; Bharati et al., 2025) and are strongest in winter from December to April, peaking in March over the UIB (Filippi et al., 2014; Dahri et al., 2016; Bharati et al., 2025). Consequently, the eastern UIB receives up to 70% of its annual precipitation in the summer, whereas the western UIB receives 40–60% of its precipitation during winter (Dahri et al., 2016; Immerzeel et al., 2015). The complex orography of the UIB also strongly influences precipitation, resulting in large localized spatial–temporal variations, including extremes (Bookhagen and Burbank, 2006; Lutz et al., 2016).

Global climate models (GCMs) often struggle to accurately simulate synoptic and local precipitation patterns (Stephens et al., 2010; Rocheta et al., 2014). Their coarse spatial resolution and limited physics cannot resolve complex hydro-meteorological dynamics, including over the HKKH region (Palazzi et al., 2013; Panday et al., 2015). To overcome these issues, regional climate models (RCMs) are used to dynamically downscale GCM outputs. These models better represent precipitation variability, including extremes. However, RCMs assume that large-scale driving fields, such as temperature and geopotential, are reliably simulated (Chokkavarapu and Mandla, 2019). RCMs are also computationally expensive, lack error estimates, propagate considerable model-dependent uncertainty (Hawkins and Sutton, 2009), and are generally not optimized for mountainous regions, especially the HKKH (Cannon et al., 2017; Norris et al., 2017, 2019; Orr et al., 2017). For example, RCMs from the Coordinated Regional Climate Downscaling Experiment (CORDEX) for South Asia regularly report seasonal errors of over 100% for present-day precipitation when compared to gauge-based gridded precipitation (Sanjay et al., 2017).

On the other hand, large-scale circulation variables, such as horizontal wind, temperature, and pressure fields, are typically better simulated by GCMs (Tapiador et al., 2019; Arias et al., 2023). Furthermore, these large-scale features have previously been used to statistically downscale monthly precipitation and seasonal precipitation forecasts over the Tian Shan and Pamir mountains in Central Asia using support vector regression (Umirbekov et al., 2022) and to downscale monthly cumulative and daily extreme precipitation over High Mountain Asia using neural networks (Gerlitz et al., 2015).

This paper explores the feasibility of using variables better represented by GCMs to predict future precipitation over the UIB. For this study, monthly fifth-generation ECMWF reanalysis (ERA5) precipitation data (Hersbach et al., 2020) is used to train a machine learning model for temporal prediction over a 15-year horizon. The following experiments serve as a proof-of-concept for applying this approach to predict future precipitation from GCMs. If ERA5 precipitation is assumed to be close to reality, this application can be considered a proof-of-concept for heterogeneous model output statistics downscaling and bias-correction (Maraun and Widmann, 2017). If ERA5 precipitation is assumed to be different from reality, this application can simply be considered an emulation of ERA5.

A large variety of methods are applicable to this task. Traditional methods such as multiple linear regression or distribution shift methods are more interpretable (Schoof and Pryor, 2001; Gudmundsson et al., 2012; Maraun and Widmann, 2017). However, scalable machine learning methods, such as convolutional neural networks, recurrent neural networks, or transformers (Wang et al., 2021; Zhong et al., 2024), and in particular probabilistic machine learning methods, such as generative adversarial networks (Harris et al., 2022; Kumar et al., 2023; Leinonen et al., 2021) or diffusion models (Ling et al., 2024; Srivastava et al., 2024; Mardani et al., 2025), have become strong choices due to their accuracy in modeling complex spatiotemporal distributions.

In this work, we choose to instead explore Gaussian processes (GP) (Rasmussen and Williams, 2005). This Bayesian machine learning method retains a high level of interpretability while providing a mechanism for incorporating domain knowledge into the model and returning uncertainty estimates through predictive probability distributions of future precipitation. These distributions can then be used to quantify the probability of extreme precipitation events. Different GP covariance functions are explored, including a non-stationary Gibbs kernel, building on work from Lalchand et al. (2022). The paper is structured as follows. The data are first presented in Section 2 and the methodology in Section 3. Model performance is presented in Section 4. Finally, the study's limitations and further work are discussed in Section 5.

2. ERA5 precipitation and predictive variables

Previous research has found that ERA5 climate reanalysis, although generally exhibiting a wet bias for the HKKH region (Baudouin et al., 2020; Chen et al., 2021; Kumar et al., 2021), offers precipitation estimates close to observations, in terms of amounts, seasonality, and variability, from a daily to multi-annual scale, compared to other reanalysis and RCM products over the UIB (Baudouin et al., 2020; Dahri et al., 2021; Khanal et al., 2023). In this work, monthly averages are chosen to learn the UIB's climatology rather than using daily averages that capture its more stochastic weather patterns. Smoother monthly data also avoids dealing with consecutive days of zero precipitation, which would require the more involved modeling of a Bernoulli-gamma distribution (Martinez-Villalobos and Neelin, 2019; Girona-Mata et al., 2024).

Figure 2 shows ERA5 monthly precipitation data over the UIB from 1970 to 2004. The precipitation has an approximately log-normal distribution with a mean of $1.83\,\text{mm}\,\text{d}^{-1}$, a standard deviation of $1.84\,\text{mm}\,\text{d}^{-1}$ and values ranging from 0.00 to $23.40\,\text{mm}\,\text{d}^{-1}$. Soft k-means clustering is applied to the precipitation data using different time intervals and cluster numbers, excluding latitude and longitude. Three distinct precipitation regimes, shown in Figure 3, are identified and subsequently referred to as the Gilgit, Ngari, and Khyber regimes. Figure 2 shows that the Ngari regime is characterized by a clear periodic pattern, while the Gilgit and Khyber regimes have more stochastic time series (Figure 2).

Potential variables for predicting future precipitation are compiled from existing literature. They are summarized in Table 1. Large-scale geopotential fields are first considered. The strength and variability of
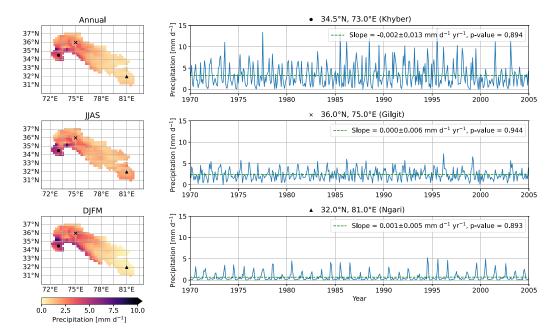
Figure 2. *ERA5 monthly precipitation over the UIB between 1970 and 2004. The average annual and seasonal precipitation over the UIB (left) and time series for three locations in the UIB (right) are shown. The time series shows the typical precipitation regimes in the basin (Figure 3) and are fit with a linear regression model (dotted green lines) to identify trends over the studied time period.*
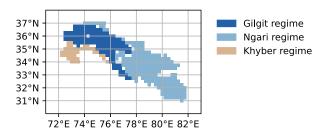

Figure 3. *UIB precipitation regimes. Clusters are generated by applying soft k-means to ERA5 monthly precipitation data from 1970 to 2004. The figure shows the clusters for a weighting threshold of 0.6, following the approach of Lalchand et al. (2022).*

the ISM and WDs over the UIB have previously been linked to large-scale geopotential fields outside the UIB via empirical orthogonal function (EOF) analysis (Filippi et al., 2014; Mölg et al., 2014). EOF analysis, also known as geographically weighted principal component analysis, is the decomposition of a signal or dataset in terms of its orthogonal basis functions. EOFs reduce the dimensionality and noise of large-scale fields and are therefore commonly used to determine atmospheric patterns (Weare and Nasstrom, 1982). Here, monthly EOFs are generated from daily ERA5 geopotentials at 200 hPa, 500 hPa, and 850 hPa over the UIB. Spatially averaged EOFs over the Bay of Bengal (10–16°N, 83–93°E) and east of the Caspian Sea (35–40°N, 60–70°E) are also calculated to capture large-scale pressure systems entering the basin.

In addition to geopotential fields, monthly historical climate indices from NOAA (2011) are investigated to train the model. Climate indices are numerical measures used to describe specific patterns or anomalies in the Earth's atmosphere and oceans. These indices are typically derived from observational data. The NAO, Niño 3.4, and Niño 4 indices, which provide a representation of the North Atlantic and El Niño-Southern oscillations, are considered. The NAO index is calculated by performing a rotated principal component analysis (RPCA) to the geopotential anomaly at 500 hPa (National Weather Service Climate Prediction

Table 1. *Considered model features with dimensions and abbreviations. These monthly variables are or are derived from ERA5 (Hersbach et al., 2020), with the exception of the atmospheric indices, which are calculated by NOAA (2011) using historical observations*

Feature	Unit	Abbreviation
Time	Years	t
Latitude	°N	lat
Longitude	°E	lon
Altitude	m	z
Slope of sub-grid orography	–	slor
Angle of sub-grid orography	deg	anor
Total column water vapour	kg/m^2	tcwv
m dew-point temperature	K	d2m
UIB-averaged 2 m temperature	K	t2m
NAO index	–	NAO
El Nino 3.4 index	K	N34
El Nino 4 index	K	N4
EOFs of geopotential at 200 hPa over UIB	–	EOF200U1, EOF200U2
Averaged EOFs of geopotential at 200 hPa over Bay of Bengal	–	EOF200B1, EOF200B2
Averaged EOFs of geopotential at 200 hPa east of Caspian Sea	–	EOF200C1, EOF200C2
EOFs of geopotential at 500 hPa over UIB	–	EOF500U1, EOF500U2
Averaged EOFs of geopotential at 500 hPa over Bay of Bengal	–	EOF500B1, EOF500B2
Averaged EOFs of geopotential at 500 hPa east of Caspian Sea	–	EOF500C1, EOF500C2
EOFs of geopotential at 850 hPa over UIB	–	EOF850U1, EOF850U2
Averaged EOFs of geopotential at 850 hPa over Bay of Bengal	–	EOF850B1, EOF850B2
Averaged EOFs of geopotential at 850 hPa east of Caspian Sea	–	EOF850C1, EOF850C2

Center, 2008). The Niño 3.4 and 4 indices measure anomalous sea surface temperatures over overlapping areas in the tropical Pacific Ocean. Finally, ERA5 surface-level variables are studied. These include static fields such as latitude, longitude, altitude, slope of sub-grid orography, and angle of sub-grid orography as well as local meteorological variables such as 2 m dew point temperature and total column water vapour. Although the latter two variables are not large-scale fields, they can be more faithfully modelled by GCMs. This is a result of the variables being more directly related to the thermodynamic properties of the atmosphere and less influenced by the complex microphysical processes that affect precipitation, especially over mountains (Houze, 2012; Tapiador et al., 2019). The monthly 2 m temperature averaged over 20–40°N 65–85°E is also investigated as a proxy for seasonality with annual periodicity similar to precipitation.

3. Method

This section first introduces GPs and methods to address non-stationary and non-Gaussian data. We then present the experimental setup, including the validation protocol and the specific models used.

3.1. Gaussian processes

Consider the data set $\{y_i\}_{i=1}^N$ representing N monthly ERA5 precipitation values where $y_i \in \mathbb{R}$. The values are given at inputs $\{x_i\}_{i=1}^N \in \mathbb{R}^D$, where D is the number of input dimensions (e.g., time, latitude, longitude, or elevation). We assume y_i can be generated by a function f at x_i:

$$y_i = f(x_i) + \varepsilon_i, \tag{1}$$

where ε_i is a noise term and is assumed to be normally distributed with zero mean and standard deviation σ_n, $\varepsilon_i \sim \mathcal{N}(0; \sigma_n^2)$.

The function f can be modeled with a GP. We refer the reader to Rasmussen and Williams (2005) for an introduction to GPs and follow their notation. A GP is a stochastic process for which any finite collection of its random variables is distributed according to a multivariate normal distribution. Similarly to a multivariate normal distribution, a GP is defined by a mean function $\mu(\boldsymbol{x})$ and a covariance function or kernel $k(\boldsymbol{x},\boldsymbol{x}')$:

$$f(\boldsymbol{x}) \sim GP(\mu(\boldsymbol{x}), k(\boldsymbol{x},\boldsymbol{x}')), \tag{2}$$

where $\boldsymbol{x}$ and $\boldsymbol{x}'$ are two arbitrary inputs. Practically, the mean function $\mu(\boldsymbol{x})$ is used to model behaviour common to all data, such as linear trends. The covariance function $k(\boldsymbol{x},\boldsymbol{x}')$ captures the correlation between the outputs and encodes properties such as smoothness and periodicity. The mean and covariance function can therefore be used to sample representative functions from the GP. The spread of these functions is interpreted as the uncertainty of the model.

The covariance function makes GPs well-suited for highly correlated geophysical datasets. If the covariance function is stationary, the correlation of the outputs depends only on the distance between the inputs $\boldsymbol{x}$ and $\boldsymbol{x}'$. The Gibbs kernel (1998) introduces non-stationarity through input-dependent covariance lengthscales in each dimension. This flexibility should allow the function samples to adapt to the varying precipitation patterns across the UIB. Instead of a single point estimate of the lengthscale ℓ_d per dimension d, we can consider $\ell_d(\boldsymbol{x})$, a lengthscale function. The Gibbs kernel for multi-dimensional inputs is given by,

$$k_{\text{Gibbs}}(\boldsymbol{x},\boldsymbol{x}') = \prod_{d=1}^{D} \sqrt{\frac{2\ell_d(\boldsymbol{x})\ell_d(\boldsymbol{x}')}{\ell_d^2(\boldsymbol{x}) + \ell_d^2(\boldsymbol{x}')}} \exp\left\{ -\sum_{d=1}^{D} \frac{(x_d - x_d')^2}{\ell_d^2(\boldsymbol{x}) + \ell_d^2(\boldsymbol{x}')} \right\}. \tag{3}$$

We model these lengthscale functions non-parametrically with latent GPs, building on work from Lalchand et al. (2022). In this paper, our aim is to design a covariance function using knowledge of precipitation and large-scale circulation patterns and enhance it by learning non-stationary spatial relationships between the data.

The distribution of GP is, by definition, Gaussian. Making the distribution of the target variable (i.e., precipitation) more normal can therefore help with inference. Warped GPs (Snelson et al., 2003) allow us to handle non-Gaussian distributions by transforming the GP inputs and outputs. For precipitation in the UIB, a Box–Cox function is used as the warping function:

$$y_i(\lambda) = \begin{cases} \dfrac{y_i^{\lambda} - 1}{\lambda}, & \text{if } \lambda \neq 0 \\ \log y_i, & \text{if } \lambda = 0 \end{cases} \tag{4}$$

where y_i is the ith observation and is assumed to be positive and λ is the scaling factor found through maximum likelihood estimation. We note that the inverse transformation of the predicted mean yields the median of the resulting distribution.

3.2. Experimental setup

The design and validation of the following models adhere to the guidelines proposed by Tazi et al. (2023). To train and evaluate the GPs, ERA5 monthly precipitation data are divided chronologically. The training set represents 35 years between 1970 and 2004. The model is evaluated on the next 15 years of precipitation data between 2005 and 2019. Ideally, the model would be trained on all the data in the UIB. However, to overcome the high computational complexity of GPs for large datasets, the data are sampled. Unless otherwise stated, the training set contains 10,000 samples from the training period, while the validation and test sets are each comprised of 3500 samples from the validation period. The sampling for each set is random, stratified by location, and without replacement. The training set represents 6% of the total available data over the training period. Performance is evaluated by computing the root mean square error (RMSE), the coefficient of determination R^2, and the mean log loss (MLL) as defined in Appendix A.

For the first model investigated, the temporal distribution of the precipitation is captured by fitting a GP to the time series of each grid point within the UIB. This collection of GPs is spatially independent and is subsequently referred to as the single-location model. A constant mean function is used here and in the following models, as no obvious trends are shared between the data, as shown in Figure 2. For precipitation in the UIB, an approximately annual periodicity is expected. We therefore use an additive kernel structure with local periodicity defined as

$$k_{\text{SLM}} = k_{\text{Per}}(x_t) \cdot k_{\text{Mat32}}(x_t) + \sum_{d=1}^{D} k_{\text{Mat32}}(\boldsymbol{x}_{\text{other}}), \tag{5}$$

where k_{Per} is the periodic kernel and k_{Mat32} is the Matérn 3/2 kernel. We use a Matérn kernel rather than a smooth squared exponential kernel to better model discontinuities in time and space. The periodic and Matérn kernels follow their standard definitions as presented by Duvenaud (2014). The other variables, excluding time and spatial information such as elevation, are represented by $\boldsymbol{x}_{\text{other}}$. The $\boldsymbol{x}_{\text{other}}$ inputs are determined through greedy feature selection as detailed in Appendix B and could include local or large-scale features from Table 1. To establish a baseline, we first train the single-location model with all data at each location for the full training period (full model). A data-limited model is then trained using the previously described training set, approximately 34 points per location.

For the second model, we apply a GP to each cluster shown in Figure 3. The cluster model uses a similar kernel design:

$$k_{\text{STAT}} = k_{\text{Per}}(x_t) \cdot k_{\text{Mat32}}(x_t) + k_{\text{Mat32}-\text{ARD}}(\boldsymbol{x}_{\text{lat,lon}}) + \sum_{d=1}^{D} k_{\text{Mat32}}(\boldsymbol{x}_{\text{other spatial}}), \tag{6}$$

where the subscript 'STAT' refers to the stationarity of the kernel, $k_{\text{Mat32-ARD}}$ is the Matérn 3/2 kernel with automatic relevance determination (MacKay, 1996), $\boldsymbol{x}_{\text{lat,lon}}$ are the latitude and longitude inputs, and $\boldsymbol{x}_{\text{other spatial}}$ are the other inputs, excluding time, latitude, and longitude, but including other spatial variables such as elevation or slope. These inputs are again chosen through greedy feature selection (Appendix B).

Finally, GPs are fit to the entire UIB. Two whole-basin models are applied. One model is defined using the stationary kernel from Equation 6. The other model uses a non-stationary kernel with the Gibbs function described in Equation 3 for the spatial inputs $\boldsymbol{x}_{\text{lat,lon}}$. The non-stationary kernel is defined as

$$k_{\text{NONSTAT}} = k_{\text{Per}}(x_t) \cdot k_{\text{Mat32}}(x_t) + k_{\text{Gibbs}}(\boldsymbol{x}_{\text{lat,lon}}) + \sum_{d=1}^{D} k_{\text{Mat32-ARD}}(\boldsymbol{x}_{\text{other spatial}}). \tag{7}$$

4. Results

4.1. Single-location model

The single-location model is created by training a GP at each gridpoint. Following the results shown in Figure B1, the inputs from Table 1 for this model are time, total column water vapour, 2 m dew point temperature, UIB-averaged 2 m temperature, and the first EOF components of the geopotential at 200 and 500 hPa over the UIB. The outcome of the selection is consistent with features that have a high linear correlation with precipitation. Although large-scale circulation is known to influence precipitation, this link is not established for many of the single-location models during the feature selection process.

For the full single-location model, the GPs have an average predictive skill of $R^2 = 0.50$, RMSE $= 0.71\,\text{mm}\,\text{d}^{-1}$ and MLL $= 0.87$ for the test set as shown in Table 2. The performance of the full model with respect to location, evaluated on all data between 2005 and 2019 is plotted in Figure 4. The southeast UIB presents the best R^2 scores. In these cases, the GPs faithfully predict most precipitation values but sometimes struggle to match the monsoon highs. Several GPs in the centre and west of the basin capture precipitation patterns with less accuracy. These regions roughly correspond to the boundaries between the clusters shown in Figure 3. In these more difficult cases, the GPs increase their uncertainty

Table 2. *Performance of the single-location(SLM), cluster (Khyber, Gilgit, Ngari), and whole-basin models (WBM). The RMSE is given in* mmd^{-1}

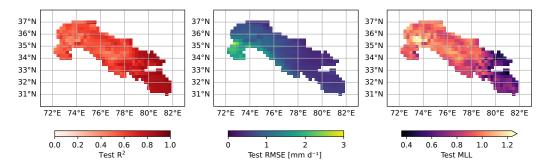|	Training			Validation			Test		
	R^2	RMSE	MLL	R^2	RMSE	MLL	R^2	RMSE	MLL
SLM (full)	0.88	0.35	-0.75	0.50	0.70	0.86	0.50	0.71	0.87
SLM (data lim.)	0.99	0.03	-4.69	-0.33	1.20	4.53	-0.29	1.15	2.46
Khyber	0.91	0.84	0.26	0.49	1.82	1.03	0.49	2.05	1.05
Gilgit	0.82	0.61	0.51	0.36	1.14	1.16	0.33	1.13	1.16
Ngari	0.76	0.45	0.56	0.58	0.66	0.95	0.58	0.64	0.93
Area weighted avg.	0.80	0.55	0.51	0.49	0.96	1.03	0.48	0.97	1.02
WBM stat.	0.74	0.92	0.63	0.51	1.21	0.91	0.49	1.25	0.92
WBM non-stat.	0.75	0.91	1.94	0.29	1.46	1.80	0.27	1.50	1.67

Figure 4. *Test R^2 (left), RMSE (centre), and MLL (right) for the full single-location model across the UIB evaluated on all precipitation data between 2005 and 2019.*

bounds accordingly, indicated by a higher MLL. The RMSE values scale with the average annual precipitation (Figure 2). Spatial differences in R^2 scores could suggest that one or more important variables are missing. The differences could also point to an inappropriate kernel design, such as constraining the precipitation to follow a periodic pattern. The data-limited model yields predictive skill of $R^2 = -0.29$, RMSE $= 1.15\,\mathrm{mmd}^{-1}$ and MLL $= 2.46$ over the 15-year horizon (Table 2). However, including information from neighbouring locations and tailoring the input features improves performance, as shown with the cluster model in the next section.

4.2. Cluster model

Next, GPs are trained on the precipitation clusters. Input features are selected for each precipitation regime in Figure B2. For the cluster models, 2 m dew point temperature is favoured over total column water vapour for most of the basin in terms of local climatic variables. The cluster models are also able to better leverage relationships between precipitation and large-scale features. The predictive skill of the models is also minimally improved (and even occasionally worsened) when including orographic features such as slope, elevation, or the angle of the sub-grid orography, despite their well-established influence on precipitation (Houze, 2012; Dahri et al., 2016; Wolvin et al., 2024).

The features' contributed variance is then compared in Table C1. The 2 m dew point temperature is a strong predictor for all precipitation regimes, in particular for Ngari model. The total column water vapour and the El Nino 3.4 index contribute strongly to the Khyber model, whereas orography is more important for the Gilgit model. Large-scale features generally contribute more to the Gilgit and Khyber regimes.

These results are consistent with known precipitation drivers for the UIB. The Khyber cluster is located at a lower altitude but is more exposed to frontal systems driven by large-scale circulation patterns. On the other hand the Ngari cluster is more sheltered and mainly receives precipitation from the ISM. The Gilgit cluster sits between the two and is characterized by the largest elevation gradients (Wolvin et al., 2024).

In this setting, the cluster model benefits from far less data than the full single-location model, but is still able to infer information about the distribution of precipitation. Using an area weighted average for this model, Table 2 shows $R^2 = 0.49$, RMSE $= 0.96\,\mathrm{mm\,d^{-1}}$ and MLL $= 1.03$ for the test set. The uncertainty calibration for the cluster model therefore improves on the full single-location model. However, the GP distribution medians generally underestimate high precipitation values at both training and test time.

4.3. Whole-basin models

GPs with stationary and non-stationary kernels are then applied to the whole basin. Following the results from Figure B3, the stationary model is trained with time, latitude, longitude, 2 m dew point temperature, UIB-averaged 2 m temperature, the first EOF components of the geopotential at 500 and 850 hPa over the UIB, the averaged first EOF component of the geopotential at 200 hPa over the Bay of Bengal and the averaged first and second EOF components of the geopotential at 500 hPa east of the Caspian Sea. These results are consistent with the trends observed in the previous section, with the stationary whole-basin model selecting the most important features from the cluster models.

For the stationary model, the 2 m dew point temperature and the UIB-averaged 2 m temperature contribute the most to the prediction of precipitation (Table C1). The same features are also used to train the non-stationary GP. The learnt lengthscales for latitude and longitude are presented in Figure 5. The optimized lengthscale values match the annual distribution of precipitation shown in Figure 2 with on average longer lengthscales for longitude and shorter lengthscales for latitude.

The whole-basin model with a stationary kernel has a similar R^2 (0.49), a worse RMSE ($1.25\,\mathrm{mm\,d^{-1}}$) but a better MLL (0.91) compared to the cluster model (Table 2). On the other hand, the non-stationary model struggles to generalize as well to new times. This more flexible model yields large predictive variances and thus MLL scores, including at training time. This behaviour suggests that this kernel design is not appropriate and may be more suitable for low-dimensional interpolation tasks such as downscaling. More data or simpler lengthscale functions could also improve model accuracy and uncertainty calibration.

5. Discussion

In this paper, we have shown that GPs can generate uncertainty-aware predictions of precipitation with moderate skill. With further analysis, the stochastic nature of these models can help us better capture the climate change signal in the means, extremes, and seasonal cycles. By producing interpretable outputs similar to a perturbed-physics ensemble for a single RCM, GPs set themselves apart from other traditional downscaling and bias-correction methods.
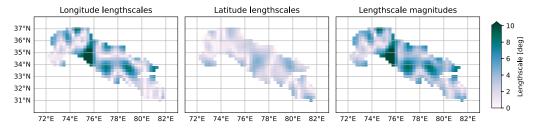

Figure 5. *Learnt non-stationary kernel lengthscales over the UIB for longitude, latitude and their magnitude.*

However, this study suffers from three main limitations. First, the application of the cluster and whole-basin models is restricted by the scalability of GPs. For these models, 6% of the training data is used for training. This amount of data requires 60 GB of RAM to train the stationary whole-basin model, which is generally not possible without a high-memory GPU (e.g., NVIDIA A100 or V100 with 80 GB HBM2 memory) to load and optimize the covariance matrix. On a single Intel Xeon Gold 5317 CPU, training the model would take 62 hours. Therefore, methods to select representative data samples (Liu et al., 2004) and scale GPs such as products of experts, low-rank approximations, and inducing points (Liu et al., 2020; Tazi et al., 2023) will need to be implemented to keep future inference and hyperparameter learning tractable. Moving to even larger sets ($>10^6$ data points) would require probabilistic deep learning, such as neural processes (Vaughan et al., 2022; Andersson et al., 2023) or hybrid deep learning-GP approaches (Putra et al., 2022; Sun et al., 2022).

Second, our knowledge of past precipitation in the UIB is incomplete. The limited in situ observations make capturing the true underlying distribution of precipitation and evaluating models challenging. Accessibility and harsh weather conditions mean that most precipitation gauges are placed in relatively low-altitude locations (Dahri et al., 2016). Alternatives, such as remote sensing, exist, but these tools alone cannot determine precipitation at a fine enough spatiotemporal scale to resolve its distribution accurately (Yin et al., 2008; Andermann et al., 2011; Meng et al., 2014; Shukla et al., 2019).

Third, most features, including large-scale circulation fields and indices, do not provide clear enough signals to improve the GP predictions. This behaviour is due to the features being heavily cross-correlated: only a few variables are needed for efficient inference. Principal component analysis could be a useful tool to improve feature selection, model skill, and the GP's numerical stability during inference, but at the expense of losing model interpretability. There is also growing evidence that long-distance atmospheric interactions, including those studied in this paper, are prone to analytical artefacts and are not as strong or simple climate predictors as once thought (Runge et al., 2014; McGraw and Barnes, 2018).

The studied temporal resolution and timescales could also limit the establishment of correlations between precipitation and these large-scale variables, and thus model accuracy. Previous work has shown that correlations with atmospheric oscillations are highly seasonal and sometimes predictive up to a year in advance (Umirbekov et al., 2022). Building more seasonal structure into the covariance function could therefore improve the viability of using large-scale features. Furthermore, weekly or daily fluctuations in the ISM and WDs cannot be resolved on a monthly scale. Including these fluctuations could allow the model to capture intra-seasonal variability crucial to predicting extreme precipitation events in the UIB. Future work could also investigate meteorological conditions during these extreme events in more depth and test more climate variables, such as large-scale humidity at different pressure levels.

Going forward, the GP models could also be refined by explicitly incorporating physics linked to precipitation over the UIB. Multiple GPs could be trained to learn the relationships between different variables, e.g., temperature and elevation. The outputs of these models could then be used to predict precipitation. For this purpose, a setup similar to the GP auto-regressive (GPAR) framework by Requeima et al. (2019) could be considered. More faithful results could also be achieved by incorporating training data from other sources, such as remote sensing or in situ precipitation gauges using multi-fidelity GPs (Tazi et al., 2024).

Additionally, a better scoring function could be used for the feature selection experiments. The MLL is calculated in the warped precipitation space and likely underestimates the loss for large precipitation values. This score also does not penalize for model complexity, which could lead to overly complicated models with too many features. Future model outputs could also be assessed with respect to long-term variability to better reflect their sensitivity to climate change.

Finally, the eventual aim of this work is to generate local precipitation predictions from GCM outputs. To achieve this goal, a GP model could be re-trained on coarsened ERA5 input data and make predictions from GCM projections. This procedure should be straightforward, as the ensemble members of the

Coupled Model Intercomparison Project (CMIP) report all the variables needed to derive the input features presented in this paper (Eyring et al., 2016; Hersbach et al., 2020). The results of this process could enable more accurate predictions of local precipitation distributions over the UIB.

6. Conclusion

This paper has demonstrated that it is possible to predict certain aspects of precipitation over complex orography using variables that are better represented by GCMs, including large-scale circulation fields. With only 6% of the available data for the training period, the whole-basin model with a stationary covariance function is able to predict monthly precipitation with an average R^2 of 0.49, an RMSE of $1.25\,\text{mmd}^{-1}$ and an MLL of 0.92. The 2 m dew point temperature and the UIB-averaged 2 m temperature were identified as the most predictive features for precipitation. Using a cluster basin model with more tailored input features produces a smaller RMSE ($0.97\,\text{mmd}^{-1}$) but a higher MLL (1.02). These results improve on the data-limited single-location model, but not the full single-location model (RMSE = 0.50, RMSE = $0.71\,\text{mmd}^{-1}$, MLL = 0.87). The stationary whole-basin model also improves on the non-stationary whole-basin model in this data-sparse setting. The non-stationary model could therefore be more suitable for better-sampled or interpolative tasks. In summary, these experiments highlight the trade-offs between learning spatial covariance, incorporating domain knowledge, and increasing computational complexity with GPs. More importantly, this work paves the way for more accurate precipitation predictions with uncertainty distributions and, in turn, more sustainable climate adaptation at a local level.

References

Andermann C, **Bonnet S and Gloaguen R** (2011) Evaluation of precipitation data sets along the Himalayan front. *Geochemistry, Geophysics, Geosystems 12*(7). https://doi.org/10.1029/2011GC003513.

Andersson TR, **Bruinsma WP**, **Markou S**, **Requeima J**, **Coca-Castro A**, **Vaughan A**, **Ellis A-L**, **Lazzara MA**, **Jones D**, **Hosking S**et al (2023) Environmental sensor placement with convolutional Gaussian neural processes. *Environmental Data Science 2*, e32. https://doi.org/10.1017/eds.2023.22

Arias P, **Bellouin N**, **Coppola E**, **Jones R**, **Krinner G**, **Marotzke J**, **Naik V**, **Palmer M**, **Plattner G-K**, **Rogelj J**, **Rojas M**, **Sillmann J**, **Storelvmo T**, **Trewin PTB**, **Rao KA**, **Adhikary B**, **Allan R**, **Armour K**, **Bala G**, … **Zickfeld K** (2023) Technical summary. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 35–144). Cambridge University Press. https://doi.org/10.1017/9781009157896

Baudouin J-P, **Herzog M and Petrie CA** (2020) Cross-validating precipitation datasets in the Indus River basin. *Hydrology and Earth System Sciences 24*(1), 427–450. https://doi.org/10.5194/hess-24-427-2020.

Bharati P, **Deb P**, **Hunt KMR**, **Orr A and Dash MK** (2025) ENSO-induced latitudinal variation of the subtropical jet modulates extreme winter precipitation over the Western Himalaya. *Adavances in Atmospheric Sciences 42*, 1–11. https://doi.org/10.1007/s00376-024-4057-2.

Bookhagen B and Burbank DW (2006) Topography, relief, and TRMM-derived rainfall variations along the Himalaya. *Geophysical Research Letters 33*(8). https://doi.org/10.1029/2006GL02603.

Bookhagen B and Burbank DW (2010) Toward a complete Himalayan hydrological budget: Spatiotemporal distribution of snowmelt and rainfall and their impact on river discharge. *Journal of Geophysical Research: Earth Surface 115*(F3). https://doi.org/10.1029/2009JF001426.

Cannon F, **Carvalho LM**, **Jones C**, **Hoell A**, **Norris J**, **Kiladis GN and Tahir AA** (2017) The influence of tropical forcing on extreme winter precipitation in the western Himalaya. *Climate Dynamics 48*, 1213–1232. https://doi.org/10.1007/s00382-016-3137-0.

Chen Y, **Sharma S**, **Zhou X**, **Yang K**, **Li X**, **Niu X**, **Hu X and Khadka N** (2021) Spatial performance of multiple reanalysis precipitation datasets on the southern slope of central Himalaya. *Atmospheric Research 250*, 105365. https://doi.org/10.1016/j.atmosres.2020.105365.

Chokkavarapu N and Mandla VR (2019) Comparative study of GCMs, RCMs, downscaling and hydrological models: A review toward future climate change impact estimation. *SN Applied Sciences 1*(12), 1698. https://doi.org/10.1007/s42452-019-1764-x.

Dahri ZH, **Ludwig F**, **Moors E**, **Ahmad B**, **Khan A and Kabat P** (2016) An appraisal of precipitation distribution in the high-altitude catchments of the Indus basin. *Science of the Total Environment 548*, 289–306. https://doi.org/10.1016/j.scitotenv.2016.01.001.

Dahri ZH, **Ludwig F**, **Moors E**, **Ahmad S**, **Ahmad B**, **Ahmad S**, **Riaz M and Kabat P** (2021) Climate change and hydrological regime of the high-altitude Indus basin under extreme climate scenarios. *Science of the Total Environment 768*, 144467. https://doi.org/10.1016/j.scitotenv.2020.144467.

Di Capua G, **Kretschmer M**, **Donner RV**, **Van Den Hurk B**, **Vellore R**, **Krishnan R and Coumou D** (2020) Tropical and mid-latitude teleconnections interacting with the Indian summer monsoon rainfall: A theory-guided causal effect network approach. *Earth System Dynamics 11*(1), 17–34. https://doi.org/10.5194/esd-11-17-2020.

Ding Q and Wang B (2005) Circumglobal teleconnection in the northern hemisphere summer. *Journal of Climate 18*(17), 3483–3505. https://doi.org/10.1175/JCL13473.1.

Duvenaud D (2014) *Automatic Model Construction with Gaussian Processes.* Doctoral dissertation, University of Cambridge. https://doi.org/10.17863/CAM.14087.

Eyring V, **Bony S**, **Meehl GA**, **Senior CA**, **Stevens B**, **Stouffer RJ and Taylor KE** (2016) Overview of the coupled model Intercomparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development 9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016.

Filippi L, **Palazzi E**, **von Hardenberg J and Provenzale A** (2014) Multidecadal variations in the relationship between the NAO and winter precipitation in the Hindu Kush–Karakoram. *Journal of Climate 27*(20), 7890–7902. https://doi.org/10.1175/JCLI-D-14-00286.1.

Gerlitz L, **Conrad O and Böhner J** (2015) Large-scale atmospheric forcing and topographic modification of precipitation rates over high Asia–a neural-network-based approach. *Earth System Dynamics 6*(1), 61–81. https://doi.org/10.5194/esd-6-61-2015.

Gibbs MN (1998) *Bayesian Gaussian Processes for Regression and Classification.* Doctoral dissertation, University of Cambridge. (accessed 11 Nov 2024) https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=b5a0c62c8d7cf51137bfb079947b8393c00ed169

Girona-Mata M, **Orr A**, **Widmann M**, **Bannister D**, **Dars GH**, **Hosking S**, **Norris J**, **Ocio D**, **Phillips T**, **Steiner J and Richard ET** (2024) Probabilistic precipitation downscaling for ungauged mountain sites: A pilot study for the Hindu Kush Karakoram Himalaya. *EGUsphere 2024*, 1–33. https://doi.org/10.5194/egusphere-2024-2805.

Gudmundsson L, **Bremnes JB**, **Haugen JE and Engen-Skaugen T** (2012) Techinal note: Downscaling RCM precipitation to the station scale using statistical transformations and a comparison of methods. *Hydrology and Earth System Sciences 16*(9), 3383–3390. https://doi.org/10.5194/hess-16-3383-2012.

Harris L, **McRae ATT**, **Chantry M**, **Dueben PD and Palmer TN** (2022) A generative deep learning approach to stochastic downscaling of precipitation forecasts. *Journal of Advances in Modeling Earth Systems 14*(10), e2022MS003120. https://doi.org/10.1029/2022MS003120.

Hawkins E and Sutton R (2009) The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society 90*(8), 1095–1108. https://doi.org/10.1175/2009BAMS2607.1.

Hersbach H, **Bell B**, **Berrisford P**, **Hirahara S**, **Horányi A**, **Muñoz-Sabater J**, **Nicolas J**, **Peubey C**, **Radu R**, **Schepers D**, **Simmons A**, **Soci C**, **Abdalla S**, **Abellan X**, **Balsamo G**, **Bechtold P**, **Biavati G**, **Bidlot J**, **Bonavita M**, **de Chiara G**, **Dahlgren P**, **Dee D**, **Diamantakis M**, **Dragani R**, **Flemming J**, **Forbes R**, **Fuentes M**, **Geer A**, **Haimberger L**, **Healy S**, **Hogan RJ**, **Hólm E**, **Janisková M**, **Keeley S**, **Laloyaux P**, **Lopez P**, **Lupu C**, **Radnoti G**, **de Rosnay P**, **Rozum I**, **Vamborg F**, **Villaume S and Thépaut J-N** (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society 146*(730), 1999–2049. https://doi.org/10.1002/qj.3803.

Hersbach H, **Bell B**, **Berrisford P**, **Biavati G**, **Horányi A**, **Sabater MJ**, **Nicolas J**, **Peubey C**, **Radu R**, **Rozum I**, **Schepers D**, **Simmons A**, **Soci C**, **Dee D and Thépaut J-N** (2023a) *ERA5 Monthly Averaged Data on Pressure Levels from 1940 to Present [Data Set].* https://doi.org/10.24381/cds.6860a573

Hersbach H, **Bell B**, **Berrisford P**, **Biavati G**, **Horányi A**, **Sabater MJ**, **Nicolas J**, **Peubey C**, **Radu R**, **Rozum I**, **Schepers D**, **Simmons A**, **Soci C**, **Dee D and Thépaut J-N** (2023b) *ERA5 Monthly Averaged Data on Single Levels from 1940 to Present [Data Set].* https://doi.org/10.24381/cds.f17050d7

Houze RA (2012) Orographic effects on precipitating clouds. *Reviews of Geophysics 50*(1). https://doi.org/10.1029/2011RG000365.

Hunt KM and Zaz SN (2023) Linking the North Atlantic Oscillation to winter precipitation over the Western Himalaya through disturbances of the subtropical jet. *Climate Dynamics 60*(7), 2389–2403. https://doi.org/10.1007/s00382-022-06450-7.

Immerzeel W and Bierkens M (2012) Asia's water balance. *Nature Geoscience 5*(12), 841–842. https://doi.org/10.1038/ngeo1643.

Immerzeel WW, **Van Beek LP and Bierkens MF** (2010) Climate change will affect the Asian water towers. *Science 328*(5984), 1382–1385. https://doi.org/10.1126/science.1183188.

Immerzeel W, **Wanders N**, **Lutz A**, **Shea J and Bierkens M** (2015) Reconciling high-altitude precipitation in the upper Indus basin with glacier mass balances and runoff. *Hydrology and Earth System Sciences 19*(11), 4673–4687. https://doi.org/10.5194/hess-19-4673-2015.

Immerzeel WW, **Lutz AF**, **Andrade M**, **Bahl A**, **Biemans H**, **Bolch T**, **Hyde S**, **Brumby SP**, **Davies BJ**, **Elmore AC**, **Emmer A**, **Feng M**, **Fernández A**, **Haritashya UK**, **Kargel JS**, **Koppes MN**, **Kraaijenbrink PDA**, **Kulkarni AV**, **Mayewski PA**, **Nepal S**, **Pacheco P**, **Painter TH**, **Pellicciotti F**, **Rajaram H**, **Rupper S**, **Sinisalo A**, **Shrestha AB**, **Viviroli D**, **Wada Y**, **Xiao C**, **Yao T and Baillie JEM** (2020) Importance and vulnerability of the world's water towers. *Nature 577*(7790), 364–369. https://doi.org/10.1038/s41586-019-1822-y.

Kelso NV and Patterson T (2010) Introducing natural earth data -naturalearthdata.com. *Geographia Technica 5*(82–89), 25. Accessed 11 November 2024. https://technicalgeography.org/pdf/sp_i_2010/12_introducing_natural_earth_data_naturaleart.pdf

Khanal S, **Tiwari S**, **Lutz A**, **Hurk B and Immerzeel W** (2023) Historical climate trends over High Mountain Asia derived from ERA5 reanalysis data. *Journal of Applied Meteorology and Climatology 62*(2), 263–288. https://doi.org/10.1175/JAMC-D-21-0045.1.

Krishnan R, **Shrestha AB**, **Ren G**, **Rajbhandari R**, **Saeed S**, **Sanjay J**, **Syed MA**, **Vellore R**, **Xu Y**, **You Q and Ren Y** (2019) Unravelling climate change in the Hindu Kush Himalaya: Rapid warming in the mountains and increasing extremes. In *The Hindu Kush Himalaya Assessment: Mountains, Climate Change, Sustainability and People*. Springer, pp. 57–97. https://doi.org/10.1007/978-3-319-92288-1_3

Kumar M, **Hodnebrog Ø**, **Daloz AS**, **Sen S**, **Badiger S and Krishnaswamy J** (2021) Measuring precipitation in Eastern Himalaya: Ground validation of eleven satellite, model and gauge interpolated gridded products. *Journal of Hydrology 599*, 126252. https://doi.org/10.1016/j.jhydrol.2021.126252.

Kumar B, **Atey K**, **Singh BB**, **Chattopadhyay R**, **Acharya N**, **Singh M**, **Nanjundiah RS and Rao SA** (2023) On the modern deep learning approaches for precipitation downscaling. *Earth Science Informatics 16*(2), 1459–1472. https://doi.org/10.1007/s12145-023-00970-4.

Lalchand V, **Tazi K**, **Cheema TM**, **Turner RE and Hosking S** (2022) Kernel learning for explainable climate science. In *16th Bayesian Modelling Applications Workshop at UAI*. https://doi.org/10.48550/arXiv.2209.04947

Leinonen J, **Nerini D and Berne A** (2021) Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network. *IEEE Transactions on Geoscience and Remote Sensing 59*(9), 7211–7223. https://doi.org/10.1109/TGRS.2020.3032790.

Ling F, **Lu Z**, **Luo J-J**, **Bai L**, **Behera SK**, **Jin D**, **Pan B**, **Jiang H and Yamagata T** (2024) Diffusion model-based probabilistic downscaling for 180-year East Asian climate reconstruction. *npj Climate and Atmospheric Science 7*(1), 131. https://doi.org/10.1038/s41612-024-00679-1.

Liu H, **Motoda H and Yu L** (2004) A selective sampling approach to active feature selection. *Artificial Intelligence 159*(1–2), 49–74. https://doi.org/10.1016/j.artint.2004.05.009.

Liu H, **Ong Y-S**, **Shen X and Cai J** (2020) When Gaussian process meets big data: A review of scalable GPs. *IEEE Transactions on Neural Networks and Learning Systems 31*(11), 4405–4423. https://doi.org/10.1109/TNNLS.2019.2957109.

Lutz A, **Immerzeel W**, **Shrestha A and Bierkens M** (2014) Consistent increase in High Asia's runoff due to increasing glacier melt and precipitation. *Nature Climate Change 4*(7), 587–592. https://doi.org/10.1038/nclimate2237.

Lutz AF, **Immerzeel WW**, **Kraaijenbrink PD**, **Shrestha AB and Bierkens MF** (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.

MacKay DJC (1996) Bayesian non-linear modeling for the prediction competition. In *Maximum Entropy and Bayesian Methods: Santa Barbara, California, USA, 1993*. Springer, pp. 221–234. https://doi.org/10.1007/978-94-015-8729-7_18

Maraun D and Widmann M (2017) *Statistical Downscaling and Bias Correction for Climate Research*. Cambridge University Press. https://doi.org/10.1017/9781107588783.

Mardani M, **Brenowitz N**, **Cohen Y**, **Pathak J**, **Chen C-Y**, **Liu C-C**, **Vahdat A**, **Nabian MA**, **Ge T**, **Subramaniam A**, **Kashinath K**, **Kautz J and Pritchard M** (2025) Residual corrective diffusion modeling for km-scale atmospheric downscaling. *Communications Earth & Environment 6*(1), 124. https://doi.org/10.1038/s43247-025-02042-5.

Martinez-Villalobos C and Neelin JD (2019) Why do precipitation intensities tend to follow gamma distributions? *Journal of the Atmospheric Sciences 76*(11), 3611–3631. https://doi.org/10.1175/JAS-D-18-0343.1.

McGraw MC and Barnes EA (2018) Memory matters: A case for granger causality in climate variability studies. *Journal of Climate 31*(8), 3289–3300. https://doi.org/10.1175/JCLI-D-17-0334.1.

Meng J, **Li L**, **Hao Z**, **Wang J and Shao Q** (2014) Suitability of TRMM satellite rainfall in driving a distributed hydrological model in the source region of Yellow River. *Journal of Hydrology 509*, 320–332. https://doi.org/10.1016/j.jhydrol.2013.11.049.

Midhuna T, **Kumar P and Dimri A** (2020) A new Western disturbance index for the Indian winter monsoon. *Journal of Earth System Science 129*, 1–14. https://doi.org/10.1007/s12040-019-1324-1.

Mölg T, **Maussion F and Scherer D** (2014) Mid-latitude westerlies as a driver of glacier variability in monsoonal High Asia. *Nature Climate Change 4*(1), 68–73. https://doi.org/10.1038/nclimate2055.

NASA JPL (2013) *NASA Shuttle Radar Topography Mission Global 1 Arc Second.* Data Set. https://doi.org/10.5067/MEASURES/SRTM/SRTMGL1.003.

National Weather Service Climate Prediction Center (2008) RPCA *Technique for Calculating Monthly Teleconnection Indices.* Accessed 27 July 2020. https://origin.cpc.ncep.noaa.gov/data/teledoc/teleindcalc.shtml.

NOAA (2011) *Climate and Weather Linkage.* Accessed 13 October 2023 https://origin.cpc.ncep.noaa.gov/products/precip/CWlink/MJO/climwx.shtml.

Norris J, **Carvalho LM**, **Jones C**, **Cannon F**, **Bookhagen B**, **Palazzi E and Tahir AA** (2017) The spatiotemporal variability of precipitation over the Himalaya: Evaluation of one-year WRF model simulation. *Climate Dynamics 49*, 2179–2204. https://doi.org/10.1007/s00382-016-3414-y.

Norris J, **Carvalho LM**, **Jones C and Cannon F** (2019) Deciphering the contrasting climatic trends between the central Himalaya and Karakoram with 36 years of WRF simulations. *Climate Dynamics 52*, 159–180. https://doi.org/10.1007/s00382-018-4133-3.

Orr A, **Listowski C**, **Couttet M**, **Collier E**, **Immerzeel W**, **Deb P and Bannister D** (2017) Sensitivity of simulated summer monsoonal precipitation in Langtang Valley, Himalaya, to cloud microphysics schemes in WRF. *Journal of Geophysical Research: Atmospheres 122*(12), 6298–6318. https://doi.org/10.1002/2016JD025801.

Orr A, **Ahmad B**, **Alam U**, **Appadurai A**, **Bharucha ZP**, **Biemans H**, **Bolch T**, **Chaulagain NP**, **Dhaubanjar S**, **Dimri AP**, **Dixon H**, **Fowler HJ**, **Gioli G**, **Halvorson SJ**, **Hussain A**, **Jeelani G**, **Kamal S**, **Khalid IS**, **Liu S**, **Lutz A**, **Mehra MK**, **Miles E**, **Momblanch A**, **Muccione V**, **Mukherji A**, **Mustafa D**, **Najmuddin O**, **Nasimi MN**, **Nüsser M**, **Pandey VP**, **Parveen S**, **Pellicciotti F**, **Pollino C**, **Potter E**, **Qazizada MR**, **Ray S**, **Romshoo S**, **Sarkar SK**, **Sawas A**, **Sen S**, **Shah A**, **Shah MAA**, **Shea JM**, **Sheikh AT**, **Shrestha AB**, **Tayal S**, **Tigala S**, **Virk ZT**, **Wester P and Wescoat Jr JL** (2022) Knowledge priorities on climate change and water in the Upper Indus Basin: A horizon scanning exercise to identify the top 100 research questions in social and natural sciences. *Earth's Future 10*(4), e2021EF002619. https://doi.org/10.1029/2021EF002619.

Palazzi E, **Von Hardenberg J and Provenzale A** (2013) Precipitation in the Hindu-Kush Karakoram Himalaya: Observations and future scenarios. *Journal of Geophysical Research: Atmospheres 118*(1), 85–100. https://doi.org/10.1029/2012JD018697.

Panday PK, **Thibeault J and Frey KE** (2015) Changing temperature and precipitation extremes in the Hindu Kush-Himalayan region: An analysis of CMIP3 and CMIP5 simulations and projections. *International Journal of Climatology 35*(10), 3058–3077. https://doi.org/10.1002/joc.4192.

Putra RI, **Ishikawa T and Tatsubori M** (2022) Spatiotemporal interpolation of ungauged river discharge via deep kernel learning. In *2022 IEEE International Conference on Big Data (Big Data)*. IEEE, pp. 4876–4880. https://doi.org/10.1109/BigData55660.2022.10020965

Rasmussen CE and Williams CK (2005) *Gaussian Processes for Machine Learning*, Vol. *1*. The MIT Press. https://doi.org/10.7551/mitpress/3206.001.0001.

Requeima J, **Tebbutt W**, **Bruinsma W and Turner RE** (2019) The Gaussian process autoregressive regression model (GPAR). In *The 22nd International Conference on Artificial Intelligence and Statistics*, vol. *89*, pp. 1860–1869. https://doi.org/10.48550/arXiv.1802.07182

Rocheta E, **Sugiyanto M**, **Johnson F**, **Evans J and Sharma A** (2014) How well do general circulation models represent low-frequency rainfall variability? *Water Resources Research 50*(3), 2108–2123. https://doi.org/10.1002/2012WR013085.

Runge J, **Petoukhov V and Kurths J** (2014) Quantifying the strength and delay of climatic interactions: The ambiguities of cross correlation and a novel measure based on graphical models. *Journal of Climate 27*(2), 720–739. https://doi.org/10.1175/JCLI-D-13-00159.1.

Sabin TP, **Krishnan R**, **Vellore R**, **Priya P**, **Borgaonkar HP**, **Singh BB and Sagar A** (2020) Climate change over the Himalayas. In Krishnan R, Sanjay J, Gnanaseelan C, Mujumdar M, Kulkarni A and Chakraborty S (eds), *Assessment of Climate Change over the Indian Region: A Report of the Ministry of Earth Sciences (MoES), Government of India*. Springer Singapore, pp. 207–222. https://doi.org/10.1007/978-981-15-4327-2_11.

Sanjay J, **Krishnan R**, **Shrestha AB**, **Rajbhandari R and Ren G-Y** (2017) Downscaled climate change projections for the Hindu Kush Himalayan region using CORDEX South Asia regional climate models. *Advances in Climate Change Research 8*(3), 185–198. https://doi.org/10.1016/j.accre.2017.08.003.

Schoof JT and Pryor S (2001) Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks. *International Journal of Climatology: A Journal of the Royal Meteorological Society 21*(7), 773–790. https://doi.org/10.1002/joc.655.

Shrestha AB, **Wagle N and Rajbhandari R** (2019) Chapter 6 – A review on the projected changes in climate over the Indus Basin. In KhanSI and AdamsTE (eds), *Indus River Basin*. Elsevier, pp. 145–158. https://doi.org/10.1016/B978-0-12-812782-7.00007-2.

Shukla AK, **Ojha CSP**, **Singh RP**, **Pal L and Fu D** (2019) Evaluation of TRMM precipitation dataset over Himalayan catchment: The upper ganga basin, India. *Water 11*(3), 613. https://doi.org/10.3390/w11030613.

Smolenaars WJ, **Lutz AF**, **Biemans H**, **Dhaubanjar S**, **Immerzeel WW and Ludwig F** (2021) From narratives to numbers: Spatial downscaling and quantification of future water, food & energy security requirements in the Indus basin. *Futures 133*, 102831. https://doi.org/10.1016/j.futures.2021.102831.

Snelson E, **Ghahramani Z and Rasmussen C** (2003) Warped Gaussian processes. *Advances in Neural Information Processing Systems 16*. https://proceedings.neurips.cc/paper_files/paper/2003/file/6b5754d737784b51ec5075c0dc437bf0-Paper.pdf (accessed 11 May 2025).

Srivastava P, **Yang R**, **Kerrigan G**, **Dresdner G**, **McGibbon J**, **Bretherton C and Mandt S** (2024) Precipitation downscaling with spatiotemporal video diffusion. *Advances in Neural Information Processing Systems 37*, 56374–56400. https://doi.org/10.48550/arXiv.2309.15214.

Stephens GL, **L'Ecuyer T**, **Forbes R**, **Gettelmen A**, **Golaz J-C**, **Bodas-Salcedo A**, **Suzuki K**, **Gabriel P and Haynes J** (2010) Dreary state of precipitation in global models. *Journal of Geophysical Research: Atmospheres 115*(D24). https://doi.org/10.1029/2010JD014532.

Sun AY, **Jiang P**, **Yang Z-L**, **Xie Y and Chen X** (2022) A graph neural network (GNN) approach to basin-scale river network learning: The role of physics-based connectivity and data fusion. *Hydrology and Earth System Sciences 26*(19), 5163–5184. https://doi.org/10.5194/hess-26-5163-2022.

Tapiador FJ, **Roca R**, **Del Genio A**, **Dewitte B**, **Petersen W and Zhang F** (2019) Is precipitation a good metric for model performance? *Bulletin of the American Meteorological Society 100*(2), 223–233. https://doi.org/10.1175/BAMS-D-17-0218.1.

Tazi K (2024) *Code for 'Precipitation Prediction Over the Upper Indus Basinfrom Large-Scale Circulation Patterns Using Gaussian Processes' [Code]*. https://doi.org/10.5281/zenodo.14170951.

Tazi K, **Lin JA**, **Viljoen R**, **Gardner A**, **John S**, **Ge H and Turner RE** (2023). Beyond intuition, a framework for applying GPs to real-world data. In *ICML Workshop on Structured Probabilistic Inference and Generative Modeling*. https://doi.org/10.48550/arXiv.2307.03093.

Tazi K, **Orr A**, **Hernandez-González J**, **Hosking S and Turner RE** (2024) Downscaling precipitation over high-mountain Asia using multi-fidelity Gaussian processes: Improved estimates from ERA5. *Hydrology and Earth System Sciences 28*(22), 4903–4925. https://doi.org/10.5194/hess-28-4903-2024.

TROPOMI (2019) *GMTED2010 Elevation Data at Different Resolutions [Data Set]*. (accessed 11 Nov 2024) https://www.temis.nl/data/gmted2010/index.php

Umirbekov A, **Peña-Guerrero MD and Müller D** (2022) Regionalization of climate teleconnections across central Asian mountains improves the predictability of seasonal precipitation. *Environmental Research Letters 17*(5), 055002. https://doi.org/10.1088/1748-9326/ac6229.

Vaughan A, **Tebbutt W**, **Hosking JS and Turner RE** (2022) Convolutional conditional neural processes for local climate downscaling. *Geoscientific Model Development 15*(1), 251–268. https://doi.org/10.5194/gmd-15-251-2022.

Wang F, **Tian D**, **Lowe L**, **Kalin L and Lehrter J** (2021) Deep learning for daily precipitation and temperature downscaling. *Water Resources Research 57*(4), e2020WR029308. https://doi.org/10.1029/2020WR029308.

Weare BC and Nasstrom JS (1982) Examples of extended empirical orthogonal function analyses. *Monthly Weather Review 110*(6), 481–485. https://doi.org/10.1175/1520-0493(1982)110%3C0481:EOEEOF%3E2.0.CO;2.

Wolvin S, **Strong C**, **Rupper S and Steenburgh W** (2024) Climatology of orographic precipitation gradients over High Mountain Asia derived from dynamical downscaling. *Journal of Geophysical Research: Atmospheres 129*(20), e2024JD041010. https://doi.org/10.1029/2024JD041010.

Yin Z-Y, **Zhang X**, **Liu X**, **Colella M and Chen X** (2008) An assessment of the biases of satellite rainfall estimates over the Tibetan plateau and correction methods based on topographic analysis. *Journal of Hydrometeorology 9*(3), 301–326. https://doi.org/10.1175/2007JHM903.1.

Zhong X, **Du F**, **Chen L**, **Wang Z and Li H** (2024) Investigating transformer-based models for spatial downscaling and correcting biases of near-surface temperature and wind-speed forecasts. *Quarterly Journal of the Royal Meteorological Society 150*(758), 275–289. https://doi.org/10.1002/qj.4596.

A. Metric definitions

A.1. Root mean square error

The root mean square error (RMSE) represents the typical distance of the model from the data. For multiple input–output pairs $\{X, Y\}$, the RMSE is given by:

$$\mathrm{RMSE} = \sqrt{\langle (Y - f)^2 \rangle} \tag{8}$$

where f is the set of predicted values at X. We use $\langle \cdot \rangle$ here and in the following definitions as a shorthand for the average over the N data points. The RMSE is more sensitive to outliers and systematic errors than the mean absolute error or bias.

A.2. Coefficient of determination

The coefficient of determination or R^2 represents the fraction of the data's variance explained by the model. It is given by:

$$R^2 = 1 - \frac{\mathrm{SS_{res}}}{\mathrm{SS_{tot}}} = 1 - \frac{\sum_i^N (y_i - f_i)^2}{\sum_i^N (y_i - \langle Y \rangle)^2} \tag{9}$$

where $\mathrm{SS_{res}}$ is the sum of the squared residuals and $\mathrm{SS_{tot}}$ the total sum of squares. An R^2 of 1 indicates a perfect fit, while a negative R^2 means the model performs worse than the mean.

A.3. Mean log loss

The mean log loss (MLL) can be interpreted as measure of uncertainty calibration for probabilistic models. Using the predictive distribution at the test inputs X, the probability of the target data Y given the model $\mathcal{M}$, $p(Y|\mathcal{M})$, can be calculated. The log loss is given by taking the negative logarithm of this probability (Rasmussen and Williams, 2005). Taking the mean over all inputs gives the MLL:

$$\text{MLL} = -\langle \log p(Y|\mathcal{M}) \rangle = \langle \frac{1}{2}\log(2\pi\sigma^2) + \frac{(Y-\mu)^2}{2\sigma^2} \rangle, \tag{10}$$

where σ and μ are the predicted means and variance at the inputs X using the model $\mathcal{M}$. Smaller values imply more skill. The MLL is applied to the warped GP predictions.

B. Feature selection experiments

The models input features are selected via the MLL on the validation set. Figures B1, B2, and B3 show the results of greedy feature selection experiments for the single-location, cluster, and stationary whole-basin models trained on the first 5000 points from the training set. Under the assumption that we would like the models to be consistent in time and space, the covariance function are initialized with time for the single-location model and time, latitude and longitude for the cluster and whole-basin models.

Feature rank for the feature selection experiments is established through recursive feature elimination using a random forest ensemble. The recursive feature elimination shows total column water vapour and 2 m dew point temperature to be most important when predicting precipitation. Of the climate indices, the NAO index is most closely linked to precipitation. The first components of the empirical orthogonal functions (EOFs) of the geopotentials over the UIB are the most predictive of the geopotential field variables. The other atmospheric fields are found be to less connected to precipitation than local input features such as slope or elevation. Time-lagged atmospheric indices and large-scale features were also investigated (not shown), but presented no additional correlation with precipitation.
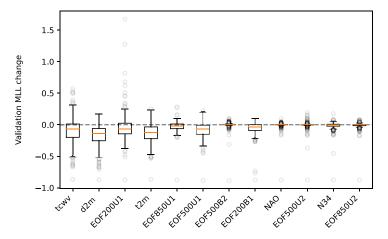

Figure B1. *Model MLL as a function of selected input features for the single-location GPs. The plot illustrates the greedy feature selection process, where the best previous kernel design is used to test the next most likely feature. Negative changes imply the feature improves model predictive skill. Features associated with positive changes decrease model predictive skill and are not included in subsequent covariance functions.*
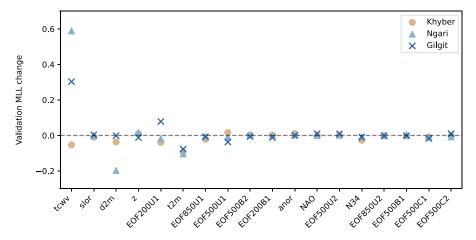
Figure B2. *As Figure B1 for the cluster GPs. Although EOF500C2 reduces the validation MLL of the Ngari GP, we exclude this feature in the final model as it sets the variance of previously selected features to be very low.*
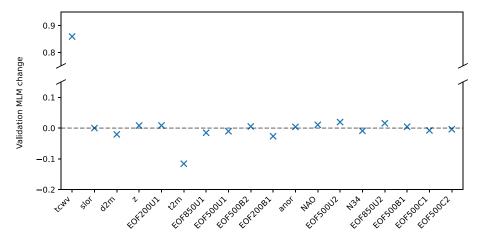

Figure B3. *As Figure B1 for the stationary whole-basin GP.*

C. Feature variance contribution

Table C1 details the contributed variance of the input features for the cluster and whole-basin models.

Table C1. *Feature variance contribution for the cluster (Khyber, Gilgit, Ngari) and stationary whole-basin (UIB) models. The variances are normalized with respect to the training set distribution, with higher values representing larger contributions towards predicting precipitation. Empty entries correspond to features that were not chosen during the feature selection experiments*

Feature	Khyber	Gilgit	Ngari	UIB
time	0.52	0.60	0.39	0.35
lat, lon	2.00	0.12	0.19	0.31
tcwv	30.4	–	–	–
slor	0.01	–	0.04	–
d2m	15.4	22.6	4.23	5.99
z	–	8.79	–	–
EOF200U1	1.95	–	0.07	–
t2m	0.01	17.8	2.14	1.90
EOF850U1	0.01	10^{-3}	0.04	0.21
EOF500U1	1.53	0.07	0.02	0.07
EOF500B2	–	1.84	–	–
EOF200B1	–	0.04	10^{-3}	10^{-3}
anor	–	5.49	–	–
NAO	–	–	10^{-3}	–
N34	2.62	0.05	0.03	0.02
EOF850U2	–	–	0.14	–
EOF500B1	0.32	0.01	–	–
EOF500C1	0.15	0.01	0.12	0.14
EOF500C2	–	–	–	0.15