



1 **A statistically based seasonal precipitation forecast model with**
2 **automatic predictor selection and its application to Central and**
3 **South Asian headwater catchments**

4

5 Lars Gerlitz, Heiko Apel, Abror Gafurov, Katy Unger-Shayesteh, Sergiy Vorogushyn, & Bruno Merz

6 GFZ German Research Center for Geosciences, Section 5.4: Hydrology, Telegrafenberg, 14473 Potsdam, Germany

7

8 *Correspondence to:* Lars Gerlitz (lars.gerlitz@gfz-potsdam.de)

9

10

11 **Abstract.** The study presents a statistically based seasonal precipitation forecast model, which automatically identifies
12 suitable predictors from globally gridded SST and climate variables by means of an extensive data mining procedure and
13 explicitly avoids the utilization of typical large-scale climate indices. This leads to an enhanced flexibility of the model
14 and enables its automatic calibration for any target area without any prior assumption concerning adequate predictor
15 variables. Potential predictor variables are derived by means of a cellwise correlation analysis of precipitation anomalies
16 with gridded global climate variables under consideration of varying lead times. Significantly correlated grid cells are
17 subsequently aggregated to predictor regions by means of a variability based cluster analysis. Finally for every month and
18 lead time, an individual random forest based forecast model is constructed, by means of the preliminary generated
19 predictor variables. Monthly predictions are aggregated to running three-months periods in order to generate a seasonal
20 precipitation forecast.

21 The model is exemplarily applied and evaluated for selected headwater catchments in Central and South Asia. Particularly
22 for winter and spring precipitation (which is associated with a westerly flow in the entire target domain) the model shows
23 solid results with correlation coefficients up to 0.7 between forecasted and observed precipitation, although the variability
24 of precipitation rates is strongly underestimated. Likewise for the monsoonal precipitation amounts in the South Asian
25 target areas a promising skill of the model could be detected. The skill of the model for the dry summer season in Central
26 Asia and the transition seasons over South Asia is found to be low.

27 A sensitivity analysis by means of well-known climate indices (such as the El Nino Southern Oscillation and the North
28 Atlantic Oscillation) reveals the major large-scale controlling mechanisms of the seasonal precipitation climate for each
29 target area. For the Central Asian target areas, both, ENSO and NAO are identified as important controlling factors for
30 precipitation totals during moist spring season. Drought conditions are found to be triggered by a warm ENSO phase in
31 combination with a positive phase of the NAO. For the monsoonal summer precipitation amounts over Southern Asia,
32 the model suggests a distinct negative response to El Nino events.

33

34 **1 Introduction**

35

36 Seasonal precipitation prediction is a crucial task in the field of applied climatology, particularly due to the manifold
37 ecological, economic and social consequences of abnormal weather conditions, such as droughts and flood events.
38 Especially in regions, characterized by a large inter-annual precipitation variability, a seasonal forecast hydro-
39 climatological parameters is required by governmental and non-governmental stakeholders in order to develop and
40 implement adequate adaption strategies e.g. for water resource management and flood protection (Chiew et al., 2003).



1 In general precipitation is a result of complex and interacting atmospheric phenomena at different spatial and temporal
2 scales and is highly variable in space and time. Thus its precise prediction more than several days ahead is illusive.
3 However regional climate conditions are actively involved in large-scale atmospheric patterns, which are (1) occasionally
4 persistent and (2) influenced by lateral boundary conditions, such as sea surface temperatures, land cover and soil moisture
5 and by external factors, e.g. variations of the solar radiation and volcanic eruptions (Palmer and Anderson, 1994; Smith
6 et al., 2012). The fact that the boundary conditions are often characterized by a low frequency variability leads to a certain
7 predictability of medium range climate conditions in many regions of the world.

8 Operational seasonal forecasts are usually based on dynamical Atmosphere Ocean General Circulation Models
9 (AOGCMs). These process-based models enable the prediction of large-scale climate conditions and variations at various
10 temporal scales (Saha et al., 2014; Smith et al., 2012). Based on the fundamental fluid dynamic equations these models
11 are designed to simulate large-scale characteristics of the climate system in a physically consistent manner. With regard
12 to exponentially increasing computing demands, the equations are numerically solved on a coarse regular grid. Small
13 scale processes, such as convective precipitation or the turbulent transport of energy and motion are only indirectly
14 considered by means of empirically based parameterizations (Smith et al., 2012). In order to utilize AOGCMs for seasonal
15 climate forecasts, the models are forced with real time initial and boundary conditions. Especially tropical sea surface
16 temperatures, but also snow covered areas and soil moisture have been identified as important influencing factors for the
17 global circulation (Brands et al., 2012; Douville and Chauvin, 2000; van den Hurk et al., 2010; Orsolini et al., 2013). Best
18 results of process-based seasonal climate forecasts are usually found in the tropics, where large-scale wind fields and
19 associated moisture fluxes are highly influenced by sea surface temperature variations. Skill for the temperate climate
20 zones are mostly lower (Kumar et al., 2013). In general, dynamical climate models are prone to biases due to uncertainties
21 in the initial conditions and are particularly reliable, when large model ensembles are available (Eden et al., 2015; Suárez-
22 Moreno and Rodríguez-Fonseca, 2015). Due to their high computing requirements, dynamical seasonal forecasts are
23 reserved to a few research centers and are not suitable for application in hydro-meteorological and environmental offices,
24 particularly in developing and transition countries.

25 As an efficient alternative, statistical forecast models are widely applied in order to derive suitable input data for climate
26 impact investigations. Based on the assumption that seasonal climate anomalies are triggered by variations of nearby or
27 remote atmospheric, oceanic or terrestrial conditions, these models attempt to find robust statistical relationships between
28 observed climate anomalies and the state of adequate predictor variables during the previous months. Since near surface
29 temperature and precipitation are the most decisive variables for the hydrological budget and exhibit the strongest impact
30 on climate sensitive environments, these variables are frequently used as predictants.

31 Particularly the state of the El Niño Southern oscillation (ENSO) is known to influence the large-scale precipitation
32 patterns almost everywhere on the globe (Dai and Wigley, 2000; Mason and Goddard, 2001; Stone et al., 1996).
33 Particularly the precipitation variability in the tropical regions is directly determined by ENSO due to its impact on the
34 tropical Walker Circulation. During El Niño events positive SST anomalies occur over the eastern tropical Pacific as a
35 result of weakened easterly trade winds. A common consequence is the occurrence of drought periods in South East Asia,
36 especially over Indonesia, and the simultaneous presence of long lasting precipitation events over the arid regions of the
37 western slopes of the South American Andes (Julian and Chervin, 1978; Wang, 2002). However several studies
38 demonstrate a statistically significant correlation of El Niño-Indices (usually derived from SST-observations in the El
39 Niño core regions or from associated pressure gradients between Darwin and Tahiti) with seasonal precipitation time series
40 in other parts of the tropics and also in temperate climate zones. E.g. various studies detected a robust statistical
41 relationship between Australian monsoonal precipitation and the ENSO state during previous months (Cai et al., 2011;
42 Ummenhofer et al., 2009). A significant influence of El Niño events was found for monsoonal precipitation amounts in



1 Eastern and Southern Africa (Liebmann et al., 2014; Ratnam et al., 2014) and the Sahel region (Parhi et al., 2015). For
2 the South Asian monsoon a negative response to El Nino events has been frequently perceived (Krishnaswamy et al.,
3 2014; Lau and Wu, 2001; Surendran et al., 2015). For the semi-arid regions of Central Asia and for the Mediterranean
4 region a positive relationship of winter and spring precipitation to El Nino events during previous autumn was found e.g.
5 by Barlow et al. (2002); Hoell et al. (2013); Roghani et al. (2015) and Syed et al. (2006). Moreover, Fraedrich (1994)
6 and Wu and Lin (2012) detected a statistically significant influence on extra tropical circulation anomalies such as the
7 position of large-scale rossby waves and the associated the North Atlantic Oscillation. This subsequently leads to a certain
8 impact of El Nino events on the European winter climate, although correlations are in general less robust compared with
9 tropical regions. Other tropical SST modes frequently used in seasonal forecasts include the Indian Ocean Dipole (IOD),
10 the Atlantic Multidecadal Oscillation (AMO) and the Pacific Decadal Oscillation (PDA), which have a significant
11 predictive skill for their adjacent coastal regions (Eden et al., 2015).

12
13 Numerous studies additionally use customized SST indices as predictor variables for seasonal precipitation forecasts. For
14 example Hartmann et al. (2016) tested the predictive skill of mean SSTs from various ocean basins surrounding the Asian
15 continent for the precipitation variability in the arid Tarim basin in North Western China. Hertig and Jacobeit (2010)
16 investigated the predictive skill of EOF-derived SST patterns of the Northern Atlantic, in order to forecast winter
17 precipitation amounts in the Mediterranean. Seibert et al., (2016) recently demonstrated that customized SST indices from
18 the Indian and Southern Atlantic Ocean improve the quality of statistical seasonal forecasts for the Limpopo-basin in
19 Southern Africa. Suárez-Moreno and Rodríguez-Fonseca (2015) showed that particularly for coastal regions, adjacent
20 Sea surface temperatures can significantly improve the seasonal forecast of precipitation.

21 Few studies utilize large-scale atmospheric pressure modes for seasonal climate predictions. However, e.g. Wu et al. (2009)
22 report, that the winter state of the North Atlantic Oscillation (defined as the pressure gradient between the Iceland low
23 and the Azores high pressure cell) influences the SST pattern of the Northern Atlantic during spring season and affects
24 the intensivity of the subsequent East Asian Summer Monsoon via cross Eurasian teleconnections. Hasson et al. (2014)
25 found a statistical significant influence of the North Atlantic Oscillation on winter precipitation amounts in the Indus
26 basin. Likewise Hartmann et al. (2016) tested the predictive skill of pressure patterns over Europe and Asia (such as the
27 North Atlantic Oscillation or the Siberian High Index) on precipitation anomalies in the Tarim Basin.

28 Eventually local land cover characteristics are frequently applied in statistical seasonal forecast models. For example
29 Cohen and Entekhabi (1999) and Cohen and Barlow (2005) show that the snow cover over Eurasia during autumn and
30 spring alters the large-scale atmospheric circulation over the Northern hemisphere with wide implications on precipitation
31 patterns during subsequent months. Brands et al. (2012) report a statistical significant relationship between late autumn
32 snow cover over Eurasia and winter time precipitation over Europe. Tian and Fan (2015) argue, that the state of the NAO
33 and the associated precipitation patterns over Europe are influenced by both, Atlantic SSTs and snow cover rates over
34 Eurasia. Likewise some studies indicate a negative response of the South Asian monsoon to higher snow cover rates over
35 Eurasia, most likely due to a delayed surface heating of the Asian continent (Wu and Qian, 2003; Zhang et al., 2004).
36 Recently some studies also included local soil moisture or previous rainfall into statistical forecasting models in order to
37 capture water recycling due to autochthonous weather conditions and persistent circulation characteristics (Eden et al.,
38 2015; van den Hurk et al., 2010).

39 As shown, most statistical forecast applications utilize either well-known climate indices or expert knowledge based
40 customized indices from SSTs or land cover characteristics. Customized indices are frequently included, since typical
41 climate indices do not cover regional scale anomalies of SST or pressure patterns which might be important predictor
42 variables for seasonal climate forecasts in certain regions. However these customized indices are usually calibrated with



1 regard to specific target areas and thus are not transferable to any other regions. Hence state of the art seasonal climate
2 forecast models are either based on a fixed number of climate indices (and thus might not consider important predictor
3 variables) or are highly site specific and barely transferable to other regions. Recently some advances towards an
4 automatic predictor selection were made by Suárez-Moreno and Rodríguez-Fonseca (2015), who used gridded SST fields
5 as potential predictors in order to automatically identify SST patterns, which are relevant for the seasonal precipitation
6 forecast in selected target areas.

7 With the aim of developing an operational seasonal forecast model, which is easily transferable to any region in the world,
8 we present a generic data mining approach which automatically selects potential predictors from gridded SST
9 observations and large-scale atmospheric circulation patterns derived from reanalysis data. Subsequently the approach
10 generates robust statistical relationships with posterior precipitation anomalies for user selected target regions. Therefore
11 the statistical package R (R Development Core Team, 2008) as well as the scripting environment of the free and open
12 source GIS system SAGA (Conrad et al., 2015) are utilized. The precipitation forecast model is based on a cell-wise
13 correlation analysis of various gridded variables with regional precipitation estimates, which identifies grid cells with
14 potential predictive skill for a specific target area with different time lags. Grid cells which significantly correlate with
15 precipitation anomalies during subsequent months are aggregated to predictor regions by means of an automatic cluster
16 analysis for every variable and time lag. Thus, for every target area, specific predictor variables are automatically derived.
17 The cluster regions are afterwards utilized as potential predictors in a non-parametric and non-linear random forest based
18 modelling approach. Based on an independent period, the model performance for the selected target area is evaluated
19 before an operational forecast is generated based on real time predictor fields.

20 In the following section we provide a detailed overview about the utilized data sets and the main model components, such
21 as predictor selection, model calibration and evaluation. Subsequently we exemplarily provide some applications of the
22 model for selected target areas in Central and South Asia. In order to make the individual modelling steps more
23 comprehensible, we already provide some major interim results for one target area within the following section. The target
24 area of the Naryn headwater catchment is located in Kyrgyzstan and is mostly influenced by westerly air masses.
25 Precipitation observations show a clear annual cycle with maximum precipitation amounts during spring season and a
26 large inter-annual variability due to its location in continental Central Asia. A detailed description of the climatic
27 characteristics of the Naryn basin is provided in Sec. 3.

29 **2 Methods and Data**

31 **2.1 Modelling Structure**

32
33 The major objective of the presented model is to derive suitable predictor variables from global oceanic and atmospheric
34 fields and to develop robust statistical relationships which enable a seasonal precipitation forecast for user selectable
35 target regions. The underlying data sets as well as the major model components are summarized in Fig. 1. In order to
36 analyze the precipitation variability in selected target areas the model is based on the CRU TS 2.0 precipitation data set,
37 which provides monthly precipitation estimates for the 20th century on a global grid with a resolution of 0.5° lat./long.
38 (Harris et al., 2014; New et al., 1999). The data set is based on a dense network of observations for the period from 1961
39 to 1990, which were used for the spatialization of monthly mean precipitation amounts, and a compilation of station
40 records with longer time series available, which were used for the calculation of anomalies and were subsequently
41 spatially interpolated based on inverse distances. New et al. (1999) show that this approach is suitable for the resolution



1 of 0.5° since it combines a climatic baseline, which is highly influenced by the underlying topography with simple
2 interpolated anomalies, which are mainly driven by large-scale weather conditions.

3

4 [Fig. 1]

5

6

7 Areal mean monthly precipitation sums for the selected target region are extracted from the CRU data set. Since monthly
8 time series of precipitation are usually positively skewed, which might not compromise the assumptions of the subsequent
9 correlation analysis, the actual values are converted into the Standardized Precipitation Index (SPI) (Guttman, 1998;
10 McKee et al., 1993) for every single month of the year. Therefore the precipitation distribution for each month is fitted to
11 a gamma distribution with suitable shape and scale parameters. The exceedance probability of observed precipitation
12 amounts is then converted into z-values of the normal distribution. The SPI values, which are normally distributed by
13 definition, are subsequently cell-wise correlated with gridded global SST and climate data with lead times ranging from
14 1 to 6 months. For every variable and lead time grid cells are identified, which significantly correlate with the mean
15 monthly SPI time series. These grid cells are subsequently aggregated to predictor regions with similar variability by
16 means of a Hill-Climbing based k-means cluster analysis. For every large-scale variable and time lag the areal mean
17 anomalies of those cluster regions are considered as potential predictor variables for a random forest based precipitation
18 forecast model. All data sets (predictants and predictor variables) are automatically processed for the period from 1948 to
19 2004. In order to find robust predictor variables for monthly precipitation amounts and to exclude incidental correlations,
20 the data set is randomly partitioned into two subsets with 27 years of observations each. One of those subsets is utilized
21 for the cell-wise correlation analysis, the second one is employed for the subsequent calibration of a random forest based
22 forecast model. Since precipitation usually shows a rather random temporal variability at a monthly time scale, results of
23 the monthly precipitation forecast are in general unreliable. Thus, modelling results are aggregated to running three-
24 months precipitation totals. The period from 2005 to 2014 is eventually used for model evaluation.

25

26 **2.2 Predictor selection**

27

28 As briefly reviewed in the introductory section, seasonal precipitation anomalies in many regions of the world can be
29 statistically forecasted by means of large-scale atmospheric and oceanic indices or under consideration of customized
30 parameters. With the aim of automatically deriving adequate predictor variables for monthly precipitation anomalies from
31 large-scale atmospheric and oceanic conditions an expansive correlation and data-mining procedure is conducted by the
32 presented seasonal forecast model. A brief summary of global gridded variables which are used for the identification of
33 potential predictor variables is given in Tab. 1.

34

35 In order to reveal the influence of nearby or remote SST anomalies on precipitation characteristics, we make use of the
36 NOAA Extended global Sea Surface Temperature ERSST V3b (Smith et al., 2008; Smith and Reynolds, 2003), which is
37 available at a resolution of 2°x2° for the period from 1854 onwards. The data set is based on in situ sea surface temperature
38 observations only, which are regionalized by means of statistical methods, considering both, low and high frequency
39 oceanic modes. With the aim of avoiding statistical artefacts, resulting from the variability of the sea ice extent in polar
40 oceans, we restricted the analysis of SST patterns to the geographical region between 65° N and 65° S. Further we utilize
41 variables representing the state of the large-scale atmospheric circulation from the NCAR-NCEP reanalysis (Kalnay et
42 al., 1996). The reanalysis, which is published by the National Center for Environmental Prediction (NCEP) and the
National Center for Atmospheric Research (NCAR), is a near-real-time gridded data set which combines atmospheric



1 observations with climate modelling results by means of a 4d-assimilation system for the period from 1948 onwards. As
2 potential atmospheric predictors we used monthly aggregated values of sea level pressure (SLP), the 500 hPa geopotential
3 height (GPH500) as well as the geopotential thickness between the 500 hPa and 200 hPa pressure level (GPH500-200).
4 In order to investigate the land surface characteristics and their subsequent effects, we additionally utilize monthly
5 aggregated global grids of near surface temperature (TEMP), antecedent precipitation amounts (PREC) and snow water
6 equivalent (SWE) from the NCAR-NCEP reanalysis as potential predictor variables. While the intrinsic pressure related
7 variables are provided at a spatial resolution of $2.5^{\circ} \times 2.5^{\circ}$ lat./long., the diagnostic land surface variables are available at
8 a resolution of approximately $1.875^{\circ} \times 1.904^{\circ}$. All predictor fields are cell-wise normalized for every month respectively.
9 Since the utilized large-scale predictor variables are updated regularly and freely downloadable they are suitable for the
10 development of an operational seasonal precipitation forecast system.

11

12 [Tab. 1]

13 We assume that typical atmospheric and oceanic indices are determined by large-scale pressure patterns or SST modes
14 and thus are inherently included in those global gridded data sets. Likewise, additional predictor variables, which might
15 be specific for a particular target region (e.g. SSTs at adjacent coasts, regional snow cover rates or enhanced water
16 availability due to high precipitation amounts during previous months) are expected to be covered by the predictor fields
17 and will be identified as relevant predictors by means of the following correlation and data-mining procedure.

18 Primarily, based on the first random sample of 27 years, a pearson correlation analysis of the monthly SPI values is
19 conducted for each gridded large-scale variable and each grid cell. The correlation analysis is separately executed for
20 every month of the year and for lead times of 1 to 6 months. Thus the identification of relevant predictor variables and
21 regions is specific for every month and lead time. Particularly for temperature related predictor variables, the time series
22 might include statistically significant trends, due to anthropogenic greenhouse gas emissions, which frequently exceed
23 the magnitude of natural variability. However, there is evidence, that seasonal precipitation anomalies in specific target
24 regions are in fact highly influenced by SST anomalies of nearby or remote oceans, but do not show a distinct response
25 to global warming during recent decades (Hoerling et al., 2010). Thus the time series of potential predictor grid cells are
26 detrended prior the correlation analysis. For every variable, each grid cell, which correlates significantly ($\alpha=0.1$) with
27 the SPI time series is subsequently labeled as potentially predictive for the monthly precipitation forecast. This
28 comparably low level of statistical significance is deliberately chosen in order to detect second order correlations and
29 conditional statistical relationships. Overall, the correlation analysis generates a data set of 504 correlation grids, each of
30 them for a specific predictor variable, month of the year and time lag. Fig. 2 (A1 and A2) exemplarily shows the results
31 of the correlation analysis for the standardized precipitation values of the Naryn region with global gridded SSTs for
32 March and September with a lag time of two months.

33

34 [Fig. 2]

35

36 During March (representing the wet season in Naryn) monthly precipitation shows a clear positive response to January
37 SST variations in the El Nino core region – a result which has been frequently reported for Central Asia – but also SST
38 anomalies in the Arabian Sea are highly (positively) correlated. During September (representing rather dry climate
39 conditions), the ENSO-precipitation relationship is less pronounced and the spatial distribution of potential predictive
40 SST-regions is rather scattered, indicating less robust statistical relationships.



1 In a subsequent step each of the correlation grids (which usually contain a large number of potentially predictive grid
2 cells) is aggregated to a distinct number of correlation regions, by means of a SAGA-GIS based Hill-Climbing k-means
3 Cluster Analysis (Friedman and Rubin, 1967). Therefore (for every specific month, time lag and predictor variable) the
4 complete normalized time series of all potentially predictive grid cells within the 56 years of model calibration is
5 considered. The iterative and unsupervised classification technique firstly randomly allocates every grid cell to one of k
6 clusters. The error sum of squares is calculated as the sum of Euclidian distances of all associated grid cells from the
7 cluster centroid and displays the quality of the cluster estimation. Every grid cell is subsequently reallocated to the nearest
8 cluster and cluster centroids and error terms are recalculated. This procedure is iteratively conducted, until the error sum
9 of squares converges to its minimum value. Basically the clustering algorithm minimizes the error sum of squares within
10 the cluster groups and maximizes the error sum of squares among them. This leads to definition of regions with similar
11 temporal variability during the calibration period and thus identifies important large-scale patterns of the considered
12 predictor variable with high predictive potential for the seasonal precipitation forecast.

13 As default, the number of clusters for every correlation grid is set to 12, which has been found to adequately identify
14 typical large scale oceanic and atmospheric features (see e.g. Fig. 2, B1 and B2). An excessive number of clusters might
15 result in a disjunction of predictor regions, which reduces the predictive skill. On the contrary an insufficient number of
16 clusters will lead to an aggregation of large regions which might still be characterized by a large inhomogeneity and thus
17 are not suitable for the derivation of potential predictor variables.

18 As shown in Fig. 2, the El Nino core region in January (yellow cluster in B1) is identified as one important region for the
19 forecast of monthly precipitation amounts in March for the Naryn basin, e.g. the positive precipitation anomalies of the
20 years 1958, 1966, 1969, 1987 and 2004 are all associated with El Nino events. Likewise dry periods usually coincide with
21 La Nina events, characterized by negative SST anomalies. The nearby clusters, which are characterized by negative
22 correlations represent compensating currents and are associated with the El Nino Southern Oscillation. The January SST
23 of the Arabian Sea is identified as an independent predictor variable for the precipitation amounts in March. For the
24 precipitation variability in September the majority of predictive SST clusters is located in the Pacific Ocean.

25 The areal mean time series for every cluster are eventually used as potential predictors in the seasonal forecast model.
26 For all 7 gridded variables the cluster analysis with $k=12$ clusters is conducted resulting in an overall a number 84 potential
27 predictors for every month and lead time.

28

29 **2.3 Forecast Model Calibration**

30

31 For every month of the year and every lead time, one separate statistical forecast model is established based on the
32 potential predictor variables derived from the correlation and cluster analysis. In order to avoid overfitting and to develop
33 robust regression relationship, the model calibration is based on the second random sample and thus is independent from
34 the predictor selection procedure. Some of the potential predictor variables are highly correlated due to their association
35 to the same phenomenon, e.g. the El Nino Southern Oscillation is manifested in various SST regions and significantly
36 influences the large-scale pressure and precipitation patterns in many regions of the world. Additionally the distribution
37 of potential predictor variables is unknown, e.g. precipitation or snow water equivalent are most likely extremely skewed
38 and not normally distributed. Thus a reliable forecasting approach requires a non-parametric statistical technique, without
39 any assumption concerning the distribution and statistical independence of predictor variables. We make use of a random
40 forest based approach (Breiman, 2001), a widely utilized data mining technique, which stands out due to its flexibility
41 concerning the characteristics of predictant and predictor variables and due to its ability to detect non-linear and
42 conditional statistical relationships. Basically random forest models represent an advancement of regression tree



1 algorithms (Breiman et al., 1984) which automatically classify large data sets by means of adequate predictor variables
2 in order to identify statistical structures in the predictor space, which are highly associated with a response variable
3 (Gerlitz, 2014; Zorita et al., 1995).
4 Classification is conducted by means of an iterative procedure. In every processing step one predictor variable and one
5 split value are identified, which classify the learning sample into two sub groups, characterized by a maximal homogeneity
6 (i.e. a minimum variance) of the predictant variable. However, since the recursive regression tree approach tends to
7 considerably overfit the predictor-predictant relationships and does not only classify important structures within the
8 feature space but also the inherent noise of the predictant variable, the predictive skill of single regression trees is
9 frequently insufficient. Therefore, random forest applications consider an ensemble of various trees, which are based on
10 a subset of the complete data set respectively. By means of this bagging approach a large number of trees is constructed.
11 Prediction values are eventually calculated as the mean of predictions from all single trees. The bagging approach and the
12 ensemble composition of the final random forest model avoid overfitting and additionally provide an internal error and
13 confidence estimation (Chen et al., 2012).
14 The specific forecast models for every month and lead time are constructed based on random forests with 500 realizations.
15 Regression trees are recursively constructed until the final leaves include 3 observations or less. For the determination of
16 each splitting criterion, a randomly selected bagging-sample 2/3 of the entire learning sample is utilized.
17 As the predictant variable the absolute amount of monthly precipitation is used. This allows the subsequent additive
18 aggregation of the monthly forecast values to seasonal precipitation amounts and the evaluation of the model at different
19 temporal scales. Fig. 3 shows exemplarily the results of the monthly precipitation forecast with varying lead times for the
20 Naryn catchment for the evaluation period from 2005 to 2014.

[Fig. 3]

21 Values are converted to the monthly standardized precipitation index based on observations from the entire model
22 calibration period from 1948 to 2003. Obviously the variability of precipitation amounts is highly underestimated by the
23 random forest based precipitation forecast models, which is a typical feature of regression based statistical models,
24 particularly if the predictant variable is characterized by a large non-predictable noise. Furthermore, the correlation of
25 forecasted and observed precipitation is low with values below 0.2 for most months and lead times. The rather poor results
26 at the monthly scale certainly reflect the non-predictable noise of monthly precipitation amounts and thus lead to the
27 assumption that modelling results should not be evaluated based on discrete monthly values due to the high frequency
28 variability of precipitation events. This is confirmed by the aggregation of observations and modelling results to three-
29 months running totals, which leads to a significant increase of correlation and a decrease of variance underestimation.
30 Fig. 4 shows the time series of SPI values for running three-months total precipitation amounts. Particularly the wet winter
31 and spring seasons of 2004/2005 and 2009/2010 in Naryn are satisfactory predicted by the statistical forecast models with
32 varying lead times, but also the negative precipitation anomalies of the 2008 and 2010 spring seasons are well captured.
33 The best predictive skill is found for a lead time of 1 to 3 months ($r^2 > 0.3$), however, even for a lead time of 6 month a
34 certain skill is detected ($r^2 = 0.13$).

35

36 [Fig. 4]

37



1 With this in mind we define two forecast periods with a length of three-months respectively. For every particular month,
2 one forecast model is established based on the sum of monthly modelling results of the upcoming three-months, a second
3 forecast is conducted based on lead times of 4 to 6 months:

4

$$5 \quad F[1:3] = \sum_{l=1}^3 RF(m,l) \quad \& \quad F[4:6] = \sum_{l=4}^6 RF(m,l)$$

6

7 where $RF(m,l)$ is the specific Random Forest forecast model based on predictor variables of the month m and precipitation
8 anomalies occurring after a lead time of l months.

9 The evaluation of the seasonal forecast model is automatically conducted for the running three monthly precipitation
10 totals.

11

12 **2.4 Model Evaluation**

13

14 Since the skill of the automatic forecast model is likely to vary depending on the target area and the associated
15 precipitation regimes during different seasons, an evaluation of the automatic seasonal forecast model performance is
16 necessary in order to assess the reliability of the forecast and to interpret the results. Based on the independent period
17 from 2005 to 2014, the deterministic forecasts of three-months running totals are automatically evaluated. Although, the
18 analysis based on 10 years only might be insufficient for the precise quantification of statistical model skill, we assume
19 that a general assessment of the model quality is feasible. We abstained from the implementation of a cross-validation
20 procedure due to the high computational demands of the predictor selection routine. For each of the running three-months
21 periods, traditional performance parameters such as correlation, bias and root mean square error (RMSE) are computed,
22 which enables the assessment of the model performance for various seasons. In order to achieve a maximal comparability
23 of different target areas, bias and RMSE are specified as the percentage of the long time precipitation totals for each three
24 month period respectively.

25 Moreover, since stakeholders often require robust predictions of anomalous periods, the ability of the forecast model to
26 forecast drought and moist conditions is evaluated by means of receiver operating characteristics (ROC) for each three-
27 months period and areas under the curve (AUC) are provided. Therefore the running three-months precipitation totals are
28 converted to the associated standardized precipitation indices, based on observations of the entire model calibration period.
29 The deterministic SPI forecast is then converted into a probabilistic prediction by means of a simple residual based
30 approach. Assuming that SPI residuals are normal distributed for each three-months period respectively, we use the 10
31 years of independent observations to assess the specific standard deviation for each of the three-months periods, which is
32 subsequently utilized to transform the deterministic forecast into a normalized probability distribution. ROC curves are
33 then constructed for SPI threshold values of -0.5, representing moderate drought, and +0.5, indication moist conditions.
34 For various probability thresholds positive hit rates (defined as the number of correctly identified droughts divided by the
35 overall number of drought events) are plotted against the false negative rate (defined as the coefficient of the number of
36 false alarms and the number on non-drought conditions). ROC curves for moist conditions are equivalently constructed.
37 Eventually the area under the curve is interpreted as a performance measure of the seasonal forecast model. AUC-values
38 near 1 indicate a perfect predictive skill considering the forecast of droughts or moist periods, values of 0.5 or less indicate
39 no predictive skill at all.

40



1

2

3 **3 Model Application to headwater catchments in Central and South Asia**

4

5 With regard to an increasing demand of climatological and hydrological forecasts in this vulnerable region, we applied
6 the presented model to four headwater catchments covering different climatic settings in Central and South Asia (see Fig.
7 4). The Naryn catchment is located in the Central Tian Shan region, drains towards west and is one of the major tributaries
8 of the Syr Darya river. Kokcha is situated in the Hindu Kush region in North Eastern Afghanistan and contributes to the
9 Amu Darya river system. As presented by the mean 850 hPa wind field of the NCAR reanalysis (Fig. 5), both catchments
10 are mainly controlled by extratropical westerly circulation patterns (with contributions from south during winter and from
11 high latitudes during summer) and receive a precipitation maximum during spring season. Particularly Kokcha is
12 characterized by a large inter-annual precipitation variability as reflected by the coefficient of variation (defined as the
13 standard deviation of precipitation amounts divided by the long-term mean for every month) above $c=1$ during dry season.
14 Due to the high elevation, precipitation during moist season mainly falls as snow and is released during warm and dry
15 summer season (Barlow and Tippett, 2008; Dixon and Wilby, 2015; Schär et al., 2004). Thus winter and spring
16 precipitation amounts in the mountainous areas provide a vast share of the main Central Asian river flow during the
17 vegetation period and form the basis for the irrigation dependent agriculture of the riparian countries, which are
18 characterized by semi-arid to arid climate conditions throughout the year.

19 In contrast the catchments of Karnali and Arun are both located in the Nepal Himalayas and represent two important
20 tributaries of the Ganges river. During winter, both catchments are under influence of westerly winds and receive a certain
21 amount of precipitation due to the passage of westerly disturbances, however the maximum of precipitation is associated
22 with the Indian Summer Monsoon, which transports moist air masses from the Arabian Sea and the Bay of Bengal into
23 the target areas. A correlation analysis of seasonal SPI values for the selected target areas (not shown) reveals their
24 association to the described pluviometric regimes. In general significant correlations ($\alpha < 0.1$) are found for Naryn
25 with Kokcha and Karnali with Arun throughout the year, which indicates their association to the continental Central Asian
26 and the Monsoonal influenced subdomain respectively.

27

28 [Fig. 5]

29

30 The model application to the selected headwater catchments with different climatic characteristics enables the
31 identification of important predictor variables and the analysis of the model performance for the varying pluviometric
32 regimes of the Central and South Asian domain. In the following section we briefly introduce the large-scale atmospheric
33 processes which lead to a spatial and seasonal differentiation of precipitation amounts in this vast target domain and
34 present some influencing factors which have been frequently linked to the inter-annual precipitation variability.
35 Subsequently we discuss the modelling results with regard to major large-scale atmospheric forcing mechanisms and
36 provide a sensitivity analysis which uncovers important influencing factors on the precipitation variability under
37 consideration of different seasons of the year.

38

39 **3.1 Pluviometric regimes and precipitation variability over Central and South Asia**

40

41 In general the climate of Central and South Asia is influenced by two major pluviometric regimes, which are related to
42 westerly and monsoonal circulation systems. During boreal cold season the entire region is influenced by westerly



1 circulation patterns and precipitation is mainly associated with mid-latitude disturbances originating over the Atlantic
2 Ocean and the Mediterranean (Bohner, 2006; Bothe et al., 2011; Gerlitz et al., 2015; Maussion et al., 2014). Since the
3 track of westerly disturbances is mainly determined by the position of the 200 hPa westerly Jetstream at the polar frontal
4 zone, a clear seasonal cycle of precipitation is distinctly defined. Particularly the western part of the Himalayas receive a
5 considerable amount of winter precipitation associated with the uplift of westerly air masses, which reaches up to 60 %
6 of the annual precipitation total (Bohner, 2006; Gerlitz et al., 2015; Wulf et al., 2010). During spring the zone of westerly
7 precipitation migrates towards north, reaches the Hindu Kush region and the Pamir in March and continues to the Tien
8 Shan region in May. Mariotti (2007) show that during winter season, a northward current over the Arabian countries
9 transports tropical air masses into Central Asia, which represents an important moisture source for the westerly air masses.
10 While the continental Central Asian countries remain under influence of extratropical westerly air masses throughout the
11 year, the tropical monsoon circulation is established over South Asia during summer season (Bohner, 2006; Bookhagen
12 and Burbank, 2006; Gerlitz et al., 2015). Due to a declining strength of the monsoonal moisture fluxes towards west, clear
13 gradient of precipitation totals from east to west has been observed (Bohner, 2006; Wulf et al., 2010).

14 Investigations of the inter-annual variability of precipitation rates over Central and South Asia have frequently been
15 conducted. Most studies (Li and Yanai, 1996; Peings and Douville, 2009; Prodhomme et al., 2014) show evidence that
16 the intensity of the Indian Summer monsoon is associated with the magnitude of pressure gradients between the Indian
17 Ocean and the Asian continent, which has been linked to the extent of the snow cover over the Asian mainland and the
18 SST of the Indian Ocean (Wu and Qian, 2003). Moreover, many studies highlight the importance of the Southern
19 Oscillation for the intensity of monsoonal precipitation. Studies by Pokhrel et al. (2012) and Sigdel and Ikeda (2013)
20 indicate that El Niño events lead to reduced moisture fluxes into South Asia. Ashok et al. (2001) further identified the
21 Indian Ocean Dipole as an important predictor for the Indian Summer Monsoon. Some studies illustrate that the
22 correlation of the Southern Oscillation index (SOI) and the Indian Summer Monsoon precipitation is non-stationary and
23 weakened during recent decades (Kumar et al., 1999; Wang and He, 2012). However, Yim et al. (2013) detected a recovery
24 of the negative ENSO-Monsoon relationship during the 1990s. Chang et al. (2001) suggested that the breakdown of robust
25 relationships is due to changes in the North Atlantic climate. Rajeevan et al. (2006) detected a statistically significant
26 correlation of Western Europe winter temperatures and subsequent monsoonal precipitation amounts.

27 In contrast for the variability of winter and spring precipitation (associated with westerly weather patterns over Central
28 and South Asia) a positive relationship with the El Niño Southern Oscillation has been observed. Severe droughts have
29 been linked to the El Niño cold phase (La Niña) (Barlow et al., 2002, 2015; Hoell et al., 2013). Roghani et al., 2015 and
30 Shirvani and Landman, 2015 found statistically significant correlations of the Southern Oscillation index during summer
31 and autumn with precipitation amounts over Iran in subsequent winter. Likewise, a significant positive correlation of the
32 ENSO state with winter precipitation amounts over the Southern Himalayan slopes has been detected (Dimri, 2013; Yadav
33 et al., 2010). Mariotti (2007) show that the moisture fluxes originating over the Arabian Sea are enhanced during ENSO
34 warm phase due to the strengthening of the southwesterly current over the Arabian countries. Beside tropical SST modes,
35 the impact of Northern Atlantic climatic conditions on the winter climate of Central Asia have been frequently
36 investigated. Bothe et al. (2011) show that drought and moist winter seasons over Central Asia are dominated by different
37 wave patterns over the Eurasian sector. Schiemann et al. (2008) report that an anomalous location or a decreasing strength
38 of the westerly Jetstream result in drought conditions over parts of Central Asia due to modified tracks and intensities of
39 westerly disturbances. Dimri (2013) found that a distinct southward shift of the westerly Jetstream is associated with wet
40 winter conditions over the Himalayas. Syed et al. (2006, 2010) indicate that positive winter precipitation anomalies over
41 Afghanistan, Pakistan and Tajikistan are usually associated with El Niño events combined with a positive state of the
42 North Atlantic Oscillation. Negative correlations between the NAO index and observed precipitation anomalies were



1 found for Kyrgyzstan and Northern Uzbekistan. Yin et al., (2014) further showed that the positive phase of the Eastern-
2 Atlantic/Western-Russia and the Polar/Eurasian patterns lead to enhanced moisture fluxes into Central Asia.
3 Eventually, Hartmann et al. (2016) suggest that beside of well-known atmospheric modes, the sea surface temperatures
4 of the main moisture sources might influence the precipitation climate of the Tarim basin.
5

6 **3.2 Modelling results**

7
8 The seasonal precipitation model, including the automatic predictor selection routine, has been applied to each of the
9 selected target regions and the results have been evaluated with regard to different seasons and the accompanying
10 precipitation regimes. Fig. 6 shows the time series of observed three-months running totals (blue bars) and the aggregated
11 results (red lines) of the F[1:3] and the F[4:6] forecast model for the evaluation period from 2005 to 2014. In order to
12 keep the annual cycle, values are displayed at the center of each three month period. The date of forecast generation is
13 1.5 months earlier for F[1:3] and 4.5 months earlier for F[4:6]. The corresponding SPI values for each of the running
14 three-months periods are presented and the 90% confidence interval of the residual based probabilistic forecast is
15 illustrated. Tab. 3 summarizes the modelling results in terms of correlation, bias, RMSE and AUC for moderate drought
16 and moist conditions. The performance measures are provided for each of the running three-months periods, respectively.
17 Particularly for the Naryn catchment in Kyrgyzstan, drought and moist conditions during the evaluation period are well
18 captured by the statistical model. Moist spring seasons in 2005, 2010 and 2013 are adequately predicted by both, the
19 F[1:3] and the F[4:6] forecast models. Also the drought years of 2008 and 2012 are accurately predicted by the forecast
20 model, although the severity of the extreme 2008 drought is highly underestimated. Correlations between observed and
21 modelled precipitation totals are high (>0.4) for the prediction of winter and spring precipitation. AUC values >0.7
22 indicate that the model is capable to forecast moderate drought and moist conditions during winter and spring. For the
23 dry summer season the model fails to adequately capture the inter-annual variability of precipitation amounts, which is
24 evidenced by negative correlations and RMSE values up to 44 % of the long term mean.

25 The headwater catchment of Kokcha in Northern Afghanistan shows a similar variability of precipitation forecast with
26 positive SPI values during 2010 and negative precipitation anomalies during 2008 and 2011, which indicates a common
27 large-scale climatic forcing of these Central Asian catchments. The drought conditions during boreal cold seasons of
28 2007/2008 and 2010/2011 are evident in both the observational and the modelled time series. However the variability of
29 precipitation rates in Kokcha is highly underestimated by the statistical model, which leads to higher RMSE values
30 throughout the year. Accordingly the AUC values seldom exceed 0.7 and indicate a limited skill of the model to forecast
31 drought or moist seasons for Kokcha during the evaluation period.

32 For the monsoonal influenced catchment Karnali in Western Nepal maximum correlations were achieved during winter
33 and summer. Particularly for the monsoon season, high correlations and AUC values above 0.7 result in both the F[1:3]
34 and the F[4:6] forecasts, which indicates the ability of the model to predict monsoonal drought periods several months in
35 advance. The negative precipitation anomalies during summer monsoon seasons of 2007 and 2009 are captured by the
36 forecast model. However the severe monsoonal drought in 2009 is strongly underestimated. Likewise the extreme winter
37 drought of 2008/2009 in Karnali and Arun is only rudimentarily captured by the F[1:3] forecast model. For the transition
38 seasons, negative correlations, high RMSE values of up to 60% of the long-term mean and AUC values below 0.5 indicate
39 a poor performance of the statistical model. The performance of the F[4:6] model is limited for the Karnali catchment and
40 shows moderate correlations and AUC values only for the summer monsoon season. Eventually for the catchment of Arun
41 in Eastern Nepal notable correlations of observed precipitation amounts were only detected for F[1:3] based prediction
42 of the monsoon season. For winter and spring season the skill of the model is fairly limited.



1 In overall, the statistical model adequately captures the variability of westerly precipitation amounts for the Naryn,
2 Kokcha and Karnali catchments. For the monsoonal influenced target areas Karnali and Arun a certain predictability of
3 summer precipitation amounts was detected. During dry seasons (summer and autumn in Central Asia and transition
4 seasons in Nepal), the model performance for the evaluation period is generally poor.

5

6 [Fig. 6]

7

8 [Tab. 2]

9

10 **3.3 Sensitivity Analysis**

11

12 In comparison to e.g linear models with a small set of independent predictor variables, the complex structure of the
13 presented random forest based statistical model does not directly reveal physically interpretable input–output relationships.
14 Particularly the fact that the predictor selection procedure generates a large sample of partially highly correlated predictor
15 variables, which basically comprise the same information concerning the large scale climatic variability, impedes a direct
16 interpretation of the predictor importance and variable response. The aggregation of the modelling results to three-months
17 running precipitation totals further complicates the analysis of the inherent model structure. In order to overcome the
18 blackbox character of the statistical model, we conducted a sensitivity analysis for the selected target areas under
19 consideration of well-known atmospheric indices. In order to discover the influence of selected climate indices, individual
20 random forest models were forced with modified input data, containing only those predictor variables, which are highly
21 correlated with the considered index. This enables the estimation of the fractional response of the model to the considered
22 parameter and reveals their influence on the hydroclimatic variability in the selected target areas. The results of the
23 sensitivity analysis further enable the comparison of the modelling results with previous studies which utilize traditional
24 climate indices in order to explain or forecast precipitation anomalies in Central and South Asia (see section 3.1 for a
25 brief summary).

26 In order to estimate the influence of selected climate indices for different seasons, we analyzed the F[1:3] forecast models
27 based on predictor variables from November, February, May and August, which enable the precipitation forecast for
28 December to February, March to May, June to August and September to November respectively. Each F[1:3] forecast
29 model consists of 3 random forests, with lead times between one and three month, which are all based on predictor
30 variables derived from globally gridded variables from one particular month. With the aim of investigating the model
31 response to a selected climate indice, the time series of potential predictor variables, which are significantly correlated
32 with the indice ($\alpha=0.01$) are maintained, while the others were set to zero. The statistical forecast model is then applied
33 with modified predictor time series. The results are converted to three-months SPI values and are plotted for the entire
34 calibration and evaluation period (see Fig 7) in comparison with the evolution of the parameter itself. Particularly the
35 response of the model to selected indices during drought and moist periods (represented by red to blue bars in Fig.7)
36 provide an insight into the internal model structure and enables the identification of important large-scale influencing
37 factors triggering anomalous hydroclimatic conditions during the subsequent season. However, due to the non-linear
38 nature of the statistical model, the response fractions should not be perceived as independent or additive and should rather
39 be interpreted as a general sensitivity of the model.

40 As potentially important large scale climate indices we make use of the index of the North Atlantic Oscillation (NOA)
41 and the El Nino-3 Index, which are frequently mentioned as important influencing factors on the Central and South Asian



1 precipitation climate (Barlow et al., 2002; Hoell et al., 2013; Khidher and Pilesjö, 2014; Syed et al., 2006). The results of
2 the sensitivity analysis for NAO and ENSO are plotted in Fig 7.
3 Further we tested the sensitivity of the model to variations of the Eastern-Atlantic/Western-Russia (EA/WR) and the
4 Polar/Eurasian (P/E) pattern, as proposed by Yin et al., (2014). Additionally we utilized SST anomalies at adjacent oceans
5 as suggested by Hartmann (2015), which were extracted from the ERSST V3b data set. Since the response to most indices
6 was found to be negligible, we refrain from plotting the results of the sensitivity analysis, but will nevertheless highlight
7 some findings.

8
9 The plotted time series (Fig. 7) clearly indicate that the state of the El Nino Southern Oscillation determines the
10 precipitation variability in the entire target area. For winter and spring season (represented by the F[1:3] forecast models
11 for November and February) a positive response of the model to predictors related to the ENSO-3 index is evident for all
12 target areas, indicating an intensification of moisture fluxes and associated westerly disturbances over the entire domain
13 during ENSO warm phase and a reversed effect during cold phase of El Nino, which is consistent with previous studies
14 on the variability of cold season precipitation totals in the vast target domain (Barlow et al., 2002, 2015; Dimri, 2013;
15 Mariotti, 2007; Syed et al., 2006). The model response is strongest for the moist seasons, which is Mar-Apr-May for the
16 Central Asian and Dec-Jan-Feb for the Himalayan catchments. This coincides with a moderate to high model performance
17 for Naryn and Kokcha during spring and Karnali during winter season and emphasizes the relevance of ENSO for the
18 winter and spring moisture fluxes into Central and South Asia. It is obvious that particularly during winter and spring
19 drought conditions frequently occur simultaneously in Naryn and Kokcha and are mostly associated with an ENSO cold
20 phase. Winter drought conditions in the Himalayan headwater catchments likewise occur more often subsequent to La
21 Nina events. For example the frequently investigated prolonged winter and spring drought between 1999 and 2001 in the
22 entire target domain is obviously associated with a cold phase of the El Nino Southern Oscillation. Likewise the recent
23 negative precipitation anomalies during the 2007/2008 winter for the entire target domain and the severe 2008/2009 winter
24 drought over Nepal, are both related to a cold phase of ENSO, which is well captured by the winter and spring forecast
25 models.

26 Moist winter and spring conditions (e.g. during 2002, 2003 and 2005 in the entire target) occur frequently under ENSO
27 warm phase.

28

29 [Fig. 7]

30

31 In addition, the winter and spring precipitation forecast models for Naryn and to a lesser extent for Kokcha distinctly
32 responds to variations of predictor variables related to the North Atlantic Oscillation. Particularly during spring, which
33 represents the moist season in both headwater catchments the model obviously shows a negative reaction to variations of
34 the NAO index, which is of the same magnitude as the response to ENSO related predictors. Again this is confirmed by
35 previous studies of Syed et al. (2006) who found a negative correlation between the state of the NAO and winter precip-
36 itation over the Northern Central Asian countries and a reverse relationship for a band covering Iran, southern Afghanistan
37 and Pakistan. Thus the combination of a negative NAO phase with a warm phase of the El Nino Southern Oscillation is
38 likely to trigger drought conditions over Central Asia in spring. During winter and particularly for the target areas in the
39 Himalayan region, the response to NAO related predictor variables is distinctly less defined. For the Naryn basin a slight
40 positive response of winter precipitation amounts to variations of the NAO index during autumn exists. Further a positive
41 response to the EA-WR is evident for winter precipitation amounts in the Naryn catchment, however the magnitude of
42 the model response is substantially lower in comparison with ENSO and NAO (not shown).



1 Variations of adjacent Sea Surface Temperatures (not shown) tend to play a less significant role for the precipitation
2 variability in all target areas during winter and spring season. Solely the Mar-Apr-May forecast model for Naryn shows
3 a distinct positive response to SSTs of the Bay of Bengal. However, the simultaneous occurrence with ENSO related
4 anomalies suggests connection of both parameters, most likely due to the statistical dependence of the El Nino Southern
5 Oscillation and the Indian Ocean Dipole, which dominates the SST variations in the Indian Ocean. During summer and
6 autumn season, the positive response of the F[1:3] model to El Nino events remains constant for the target areas Naryn
7 and Kokcha, which are under influence of westerly circulation patterns throughout the year. Although model performance
8 for the Central Asian headwater catchments is poor during summer season, the results of the sensitivity study are con-
9 sistent with findings by Mariotti (2007), who proposed seasonal independent enhanced south easterly moisture fluxes into
10 Central Asia during the ENSO warm phase.

11 For the monsoonal influenced catchments Karnali and Arun, a distinct negative relationship between ENSO variations
12 and summer and autumn precipitation variability is evident in the modelling results, which again is confirmed by previous
13 studies (Rajeevan and Pai, 2007; Sigdel and Ikeda, 2013; Wu et al., 2009). Also the analysis suggested response of the
14 summer monsoon to SST variations in the Indian Ocean (not shown) which, albeit, is of significant smaller magnitude.
15 No response of the monsoonal precipitation forecast was found to the extratropical NAO, EA/WR and P/E signals.

16
17

18 **4 Summary and Outlook**

19

20 We presented a statistically based modelling framework, which automatically identifies suitable predictors from globally
21 gridded climate variables by means of an extensive data mining procedure and explicitly avoids the utilization of typical
22 large-scale climate indices. This leads to an enhanced flexibility of the model and enables its automatic calibration for
23 any target area without any prior assumption concerning adequate predictor variables. Potential predictor variables are
24 derived by means of a cellwise correlation analysis of precipitation anomalies within a user selectable target area with
25 global climate variables. The correlation analysis is conducted for monthly values with lead times ranging from one to
26 six months. For each potential predictor variable, month and lead time, significantly correlated grid cells are aggregated
27 to predictor regions by means of a variability based cluster analysis. Finally for every month and lead time, an individual
28 random forest based forecast model is constructed, by means of the preliminary generated predictor variables. In order to
29 reduce the risk of overfitting, predictor selection and model calibration are based on independent samples. Due to the
30 large noise of observed precipitation amounts at a monthly time scale, the random forest based forecasts for every month
31 with lead times of one to three-months and four to six months are aggregated to running three-months predictions. These
32 are automatically evaluated based on an independent period and modelling performance measures are provided for each
33 of the running three-months predictions, which enables the assessment of the model performance for different seasons of
34 the year.

35 The model was exemplarily applied to selected headwater catchments in Central and South Asia. While the Central Asian
36 catchments are under influence of westerly air masses throughout the year, the target areas in Southern Asia receive
37 moisture fluxes from westerly winds during winter and are under influence of the South Asian Monsoon during summer
38 season

39 Particularly for the catchment of Naryn in the Tien Shan region (Kyrgyzstan) but also for Kokcha in the Hindu Kush
40 (Northern Afghanistan) the model showed a solid performance, especially for the moist spring season. The capability of
41 the model to predict moderate drought events or anomalous moist conditions up to six months in advance is reflected by
42 high correlations with observations and AUC values >0.7 for the evaluation period. Due to the fact that precipitation in



1 the high elevations mainly falls as snow and is released during dry summer season, the irrigated agriculture of the down-
2 stream countries is highly vulnerable to drought events during winter and spring. Some studies indicate, that the natural
3 summer discharge of the tributaries of the major Central Asian rivers can be accurately forecasted by means of winter
4 precipitation amounts or snow cover rates, which are usually available in spring (Barlow and Tippett, 2008; Dixon and
5 Wilby, 2015). A modelling chain including statistical precipitation forecasting und runoff prediction could extent the
6 forecast range and enable the initiation of adequate adaption strategies.

7 For the South Asian target areas the model performance was found to be slightly lower, but particularly for the econom-
8 ically important monsoonal precipitation amounts an adequate skill could be detected.

9 A sensitivity analysis of the complex statistical model by means of well-known climate indices shows, that the model
10 automatically finds relevant predictor variables, among others, those which are associated with typical climatic modes,
11 such as the North Atlantic or the El Nino Southern Oscillation. Further the sensitivity analysis enables the estimation of
12 the model response to specified climatic modes and thus reveals the major influencing factors for the observed precipita-
13 tion variability. The winter and spring precipitation amounts in the entire target were found to be highly determined by
14 the state of the El Nino Southern Oscillation with positive precipitation anomalies during El Nino events in the entire
15 target domain. Additionally for the Central Asian catchments the state of the North Atlantic Oscillation has been identified
16 as an important controlling factor. The sensitivity analysis of the model suggests, that drought events are frequently trig-
17 gered by a positive NAO phase in combination with an ENSO cold phase. Concerning the forecast of summer precipita-
18 tion amounts for the monsoonal influenced regions Karnali and Arun, the model shows a distinct negative response to El
19 Nino events.

20 In general, the statistical model is characterized by a large underestimation of variance, but the forecast of a certain
21 drought risk appears feasible. The accurate prediction of severe drought periods however remains difficult by means of
22 statistical techniques. Therefore the atmospheric and oceanic patterns which trigger extreme drought or moist conditions
23 and the interaction of potential influencing factors (such as the state of the North Atlantic Oscillation, the El Nino Southern
24 Oscillation or SST anomalies at adjacent coasts) need to be further investigated.

26 Acknowledgements

27 This work was carried out within the framework of the CAWa (Water in Central Asia) project (<http://www.cawa-pro->
28 [ject.net](http://www.cawa-project.net), contract no. AA7090002), funded by the German Federal Foreign Office as part of the “Berlin Process”.

31 References

- 33 Ashok, K., Guan, Z. and Yamagata, T.: Impact of the Indian Ocean dipole on the relationship between the Indian
34 monsoon rainfall and ENSO, *Geophys. Res. Lett.*, 28(23), 4499–4502, doi:10.1029/2001GL013294, 2001.
- 35 Barlow, M., Cullen, H. and Lyon, B.: Drought in Central and Southwest Asia: La Niña, the Warm Pool, and Indian
36 Ocean Precipitation, *J. Climate*, 15(7), 697–700, doi:10.1175/1520-0442(2002)015<0697:DICASA>2.0.CO;2, 2002.
- 37 Barlow, M., Zaitchik, B., Paz, S., Black, E., Evans, J. and Hoell, A.: A Review of Drought in the Middle East and
38 Southwest Asia, *J. Climate*, doi:10.1175/JCLI-D-13-00692.1, 2015.
- 39 Barlow, M. A. and Tippett, M. K.: Variability and Predictability of Central Asia River Flows: Antecedent Winter
40 Precipitation and Large-Scale Teleconnections, *J. Hydrometeor.*, 9(6), 1334–1349, doi:10.1175/2008JHM976.1, 2008.
- 41 Bohner, J.: General climatic controls and topoclimatic variations in Central and High Asia, *Boreas*, 35(2), 279–295,
42 doi:10.1111/j.1502-3885.2006.tb01158.x, 2006.



- 1 Bookhagen, B. and Burbank, D. W.: Topography, relief, and TRMM-derived rainfall variations along the Himalaya,
2 Geophys. Res. Lett., 33(8), L08405, doi:10.1029/2006GL026037, 2006.
- 3 Bothe, O., Fraedrich, K. and Zhu, X.: Precipitation climate of Central Asia and the large-scale atmospheric circulation,
4 Theor Appl Climatol, 108(3-4), 345–354, doi:10.1007/s00704-011-0537-2, 2011.
- 5 Brands, S., Manzanas, R., Gutiérrez, J. M. and Cohen, J.: Seasonal Predictability of Wintertime Precipitation in Europe
6 Using the Snow Advance Index, J. Climate, 25(12), 4023–4028, doi:10.1175/JCLI-D-12-00083.1, 2012.
- 7 Breiman, L.: Random Forests, Machine Learning, 45(1), 5–32, doi:10.1023/A:1010933404324, 2001.
- 8 Breiman, L., Friedman, J., Stone, C. J. and Olshen, R. A.: Classification and Regression Trees, Taylor & Francis., 1984.
- 9 Cai, W., van Rensch, P., Cowan, T. and Hendon, H. H.: Teleconnection Pathways of ENSO and the IOD and the
10 Mechanisms for Impacts on Australian Rainfall, J. Climate, 24(15), 3910–3923, doi:10.1175/2011JCLI4129.1, 2011.
- 11 Chang, C.-P., Harr, P. and Ju, J.: Possible Roles of Atlantic Circulations on the Weakening Indian Monsoon Rainfall–
12 ENSO Relationship, J. Climate, 14(11), 2376–2380, doi:10.1175/1520-0442(2001)014<2376:PROACO>2.0.CO;2,
13 2001.
- 14 Chen, J., Li, M., Wang, W., Chen, J., Li, M. and Wang, W.: Statistical Uncertainty Estimation Using Random Forests
15 and Its Application to Drought Forecast, Statistical Uncertainty Estimation Using Random Forests and Its Application to
16 Drought Forecast, Mathematical Problems in Engineering, Mathematical Problems in Engineering, 2012, 2012,
17 e915053, doi:10.1155/2012/915053, 10.1155/2012/915053, 2012.
- 18 Chiew, F. H. S., Zhou, S. L. and McMahon, T. A.: Use of seasonal streamflow forecasts in water resources management,
19 Journal of Hydrology, 270(1–2), 135–144, doi:10.1016/S0022-1694(02)00292-5, 2003.
- 20 Cohen, J. and Barlow, M.: The NAO, the AO, and Global Warming: How Closely Related?, J. Climate, 18(21), 4498–
21 4513, doi:10.1175/JCLI3530.1, 2005.
- 22 Cohen, J. and Entekhabi, D.: Eurasian snow cover variability and northern hemisphere climate predictability, Geophys.
23 Res. Lett., 26(3), 345–348, doi:10.1029/1998GL900321, 1999.
- 24 Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V. and Böhner, J.:
25 System for Automated Geoscientific Analyses (SAGA) v. 2.1.4, Geosci. Model Dev., 8(7), 1991–2007,
26 doi:10.5194/gmd-8-1991-2015, 2015.
- 27 Dai, A. and Wigley, T. M. L.: Global patterns of ENSO-induced precipitation, Geophys. Res. Lett., 27(9), 1283–1286,
28 doi:10.1029/1999GL011140, 2000.
- 29 Dimri, A. P.: Interannual variability of Indian winter monsoon over the Western Himalayas, Global and Planetary
30 Change, 106, 39–50, doi:10.1016/j.gloplacha.2013.03.002, 2013.
- 31 Dixon, S. G. and Wilby, R. L.: Forecasting reservoir inflows using remotely sensed precipitation estimates: a pilot study
32 for the River Naryn, Kyrgyzstan, Hydrological Sciences Journal, 0(0), 1–16, doi:10.1080/02626667.2015.1006227,
33 2015.
- 34 Douville, H. and Chauvin, F.: Relevance of soil moisture for seasonal climate predictions: a preliminary study, Climate
35 Dynamics, 16(10-11), 719–736, doi:10.1007/s003820000080, 2000.
- 36 Eden, J. M., van Oldenborgh, G. J., Hawkins, E. and Suckling, E. B.: A global empirical system for probabilistic
37 seasonal climate prediction, Geosci. Model Dev., 8(12), 3947–3973, doi:10.5194/gmd-8-3947-2015, 2015.
- 38 Fraedrich, K.: An ENSO impact on Europe?, Tellus A, 46(4), 541–552, doi:10.1034/j.1600-0870.1994.00015.x, 1994.
- 39 Friedman, H. P. and Rubin, J.: On Some Invariant Criteria for Grouping Data, Journal of the American Statistical
40 Association, 62(320), 1159–1178, doi:10.2307/2283767, 1967.
- 41 Gerlitz, L.: Using fuzzified regression trees for statistical downscaling and regionalization of near surface temperatures
42 in complex terrain, Theor Appl Climatol, 122(1-2), 337–352, doi:10.1007/s00704-014-1285-x, 2014.



- 1 Gerlitz, L., Conrad, O. and Böhner, J.: Large-scale atmospheric forcing and topographic modification of precipitation
2 rates over High Asia – a neural-network-based approach, *Earth Syst. Dynam.*, 6(1), 61–81, doi:10.5194/esd-6-61-2015,
3 2015.
- 4 Guttman, N. B.: Comparing the Palmer Drought Index and the Standardized Precipitation Index I, *JAWRA Journal of*
5 *the American Water Resources Association*, 34(1), 113–121, doi:10.1111/j.1752-1688.1998.tb05964.x, 1998.
- 6 Harris, I., Jones, P. d., Osborn, T. j. and Lister, D. h.: Updated high-resolution grids of monthly climatic observations –
7 the CRU TS3.10 Dataset, *Int. J. Climatol.*, 34(3), 623–642, doi:10.1002/joc.3711, 2014.
- 8 Hartmann, H., Snow, J. A., Stein, S., Su, B., Zhai, J., Jiang, T., Krysanova, V. and Kundzewicz, Z. W.: Predictors of
9 precipitation for improved water resources management in the Tarim River basin: Creating a seasonal forecast model,
10 *Journal of Arid Environments*, 125, 31–42, doi:10.1016/j.jaridenv.2015.09.010, 2016.
- 11 Hasson, S., Lucarini, V., Khan, M. R., Petitta, M., Bolch, T. and Gioli, G.: Early 21st century snow cover state over the
12 western river basins of the Indus River system, *Hydrol. Earth Syst. Sci.*, 18(10), 4077–4100, doi:10.5194/hess-18-4077-
13 2014, 2014.
- 14 Hertig, E. and Jacobeit, J.: Predictability of Mediterranean climate variables from oceanic variability. Part II: Statistical
15 models for monthly precipitation and temperature in the Mediterranean area, *Clim Dyn*, 36(5-6), 825–843,
16 doi:10.1007/s00382-010-0821-3, 2010.
- 17 Hoell, A., Funk, C. and Barlow, M.: The regional forcing of Northern hemisphere drought during recent warm tropical
18 west Pacific Ocean La Niña events, *Clim Dyn*, 42(11-12), 3289–3311, doi:10.1007/s00382-013-1799-4, 2013.
- 19 Hoerling, M., Eischeid, J. and Perlwitz, J.: Regional Precipitation Trends: Distinguishing Natural Variability from
20 Anthropogenic Forcing, *J. Climate*, 23(8), 2131–2145, doi:10.1175/2009JCLI3420.1, 2010.
- 21 Hurk, B. van den, Doblas-Reyes, F., Balsamo, G., Koster, R. D., Seneviratne, S. I. and Jr, H. C.: Soil moisture effects on
22 seasonal temperature and precipitation forecast scores in Europe, *Clim Dyn*, 38(1-2), 349–362, doi:10.1007/s00382-
23 010-0956-2, 2010.
- 24 Julian, P. R. and Chervin, R. M.: A Study of the Southern Oscillation and Walker Circulation Phenomenon, *Mon. Wea.*
25 *Rev.*, 106(10), 1433–1451, doi:10.1175/1520-0493(1978)106<1433:ASOTSO>2.0.CO;2, 1978.
- 26 Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J.,
27 Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C.,
28 Wang, J., Jenne, R. and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, *Bull. Amer. Meteor. Soc.*, 77(3),
29 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.
- 30 Khidher, S. A. and Pilesjö, P.: The effect of the North Atlantic Oscillation on the Iraqi climate 1982–2000, *Theor Appl*
31 *Climatol*, 122(3-4), 771–782, doi:10.1007/s00704-014-1327-4, 2014.
- 32 Krishnaswamy, J., Vaidyanathan, S., Rajagopalan, B., Bonell, M., Sankaran, M., Bhalla, R. S. and Badiger, S.: Non-
33 stationary and non-linear influence of ENSO and Indian Ocean Dipole on the variability of Indian monsoon rainfall and
34 extreme rain events, *Clim Dyn*, 45(1-2), 175–184, doi:10.1007/s00382-014-2288-0, 2014.
- 35 Kumar, A., Chen, M. and Wang, W.: Understanding Prediction Skill of Seasonal Mean Precipitation over the Tropics, *J.*
36 *Climate*, 26(15), 5674–5681, doi:10.1175/JCLI-D-12-00731.1, 2013.
- 37 Kumar, K. K., Rajagopalan, B. and Cane, M. A.: On the Weakening Relationship Between the Indian Monsoon and
38 ENSO, *Science*, 284(5423), 2156–2159, doi:10.1126/science.284.5423.2156, 1999.
- 39 Lau, K.-M. and Wu, H. T.: Principal Modes of Rainfall–SST Variability of the Asian Summer Monsoon: A
40 Reassessment of the Monsoon–ENSO Relationship, *J. Climate*, 14(13), 2880–2895, doi:10.1175/1520-
41 0442(2001)014<2880:PMORSV>2.0.CO;2, 2001.
- 42 Li, C. and Yanai, M.: The Onset and Interannual Variability of the Asian Summer Monsoon in Relation to Land–Sea
43 Thermal Contrast, *J. Climate*, 9(2), 358–375, doi:10.1175/1520-0442(1996)009<0358:TOAIVO>2.0.CO;2, 1996.



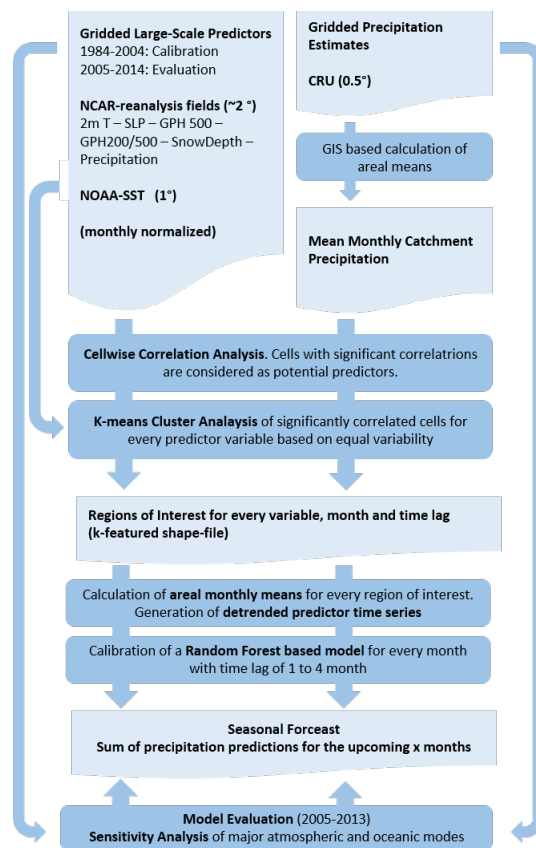
- 1 Liebmann, B., Hoerling, M. P., Funk, C., Bladé, I., Dole, R. M., Allured, D., Quan, X., Pegion, P. and Eischeid, J. K.:
2 Understanding Recent Eastern Horn of Africa Rainfall Variability and Change, *J. Climate*, 27(23), 8630–8645,
3 doi:10.1175/JCLI-D-13-00714.1, 2014.
- 4 Mariotti, A.: How ENSO impacts precipitation in southwest central Asia, *Geophys. Res. Lett.*, 34(16), L16706,
5 doi:10.1029/2007GL030078, 2007.
- 6 Mason, S. J. and Goddard, L.: Probabilistic Precipitation Anomalies Associated with ENSO, *Bull. Amer. Meteor. Soc.*,
7 82(4), 619–638, doi:10.1175/1520-0477(2001)082<0619:PPAAWE>2.3.CO;2, 2001.
- 8 Maussion, F., Scherer, D., Mölg, T., Collier, E., Curio, J. and Finkelnburg, R.: Precipitation Seasonality and Variability
9 over the Tibetan Plateau as Resolved by the High Asia Reanalysis*, *J. Climate*, 27(5), 1910–1927, doi:10.1175/JCLI-D-
10 13-00282.1, 2014.
- 11 McKee, T. B., Doesken, N. J. and Kleist, J.: The Relationship of Drought Frequency and Duration to Time Scales,
12 [online] Available from: <http://ccc.atmos.colostate.edu/relationshipofdroughtfrequency.pdf>, 1993.
- 13 New, M., Hulme, M. and Jones, P.: Representing Twentieth-Century Space–Time Climate Variability. Part I:
14 Development of a 1961–90 Mean Monthly Terrestrial Climatology, *J. Climate*, 12(3), 829–856, doi:10.1175/1520-
15 0442(1999)012<0829:RTCSTC>2.0.CO;2, 1999.
- 16 Orsolini, Y. J., Senan, R., Balsamo, G., Doblas-Reyes, F. J., Vitart, F., Weisheimer, A., Carrasco, A. and Benestad, R. E.:
17 Impact of snow initialization on sub-seasonal forecasts, *Clim Dyn*, 41(7-8), 1969–1982, doi:10.1007/s00382-013-1782-
18 0, 2013.
- 19 Palmer, T. N. and Anderson, D. L. T.: The prospects for seasonal forecasting—A review paper, *Q.J.R. Meteorol. Soc.*,
20 120(518), 755–793, doi:10.1002/qj.49712051802, 1994.
- 21 Parhi, P., Giannini, A., Gentile, P. and Lall, U.: Resolving contrasting regional rainfall responses to El Niño over
22 tropical Africa, *J. Climate*, doi:10.1175/JCLI-D-15-0071.1, 2015.
- 23 Peings, Y. and Douville, H.: Influence of the Eurasian snow cover on the Indian summer monsoon variability in
24 observed climatologies and CMIP3 simulations, *Clim Dyn*, 34(5), 643–660, doi:10.1007/s00382-009-0565-0, 2009.
- 25 Pokhrel, S., Chaudhari, H. S., Saha, S. K., Dhakate, A., Yadav, R. K., Salunke, K., Mahapatra, S. and Rao, S. A.: ENSO,
26 IOD and Indian Summer Monsoon in NCEP climate forecast system, *Clim Dyn*, 39(9-10), 2143–2165,
27 doi:10.1007/s00382-012-1349-5, 2012.
- 28 Prodhomme, C., Terray, P., Masson, S., Boschat, G. and Izumo, T.: Oceanic factors controlling the Indian summer
29 monsoon onset in a coupled model, *Clim Dyn*, 44(3-4), 977–1002, doi:10.1007/s00382-014-2200-y, 2014.
- 30 Rajeevan, M. and Pai, D. S.: On the El Niño-Indian monsoon predictive relationships, *Geophys. Res. Lett.*, 34(4),
31 L04704, doi:10.1029/2006GL028916, 2007.
- 32 Rajeevan, M., Pai, D. S., Kumar, R. A. and Lal, B.: New statistical models for long-range forecasting of southwest
33 monsoon rainfall over India, *Clim Dyn*, 28(7-8), 813–828, doi:10.1007/s00382-006-0197-6, 2006.
- 34 Ratnam, J. V., Behera, S. K., Masumoto, Y. and Yamagata, T.: Remote Effects of El Niño and Modoki Events on the
35 Austral Summer Precipitation of Southern Africa, *J. Climate*, 27(10), 3802–3815, doi:10.1175/JCLI-D-13-00431.1,
36 2014.
- 37 R Development Core Team: R: The R Project for Statistical Computing, R Foundation for Statistical Computing,
38 [online] Available from: <https://www.r-project.org/> (Accessed 17 December 2015), 2008.
- 39 Roghani, R., Soltani, S. and Bashari, H.: Influence of southern oscillation on autumn rainfall in Iran (1951–2011), *Theor*
40 *Appl Climatol*, 1–13, doi:10.1007/s00704-015-1423-0, 2015.
- 41 Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H., Iredell, M., Ek, M.,
42 Meng, J., Yang, R., Mendez, M. P., van den Dool, H., Zhang, Q., Wang, W., Chen, M. and Becker, E.: The NCEP
43 Climate Forecast System Version 2, *J. Climate*, 27(6), 2185–2208, doi:10.1175/JCLI-D-12-00823.1, 2014.



- 1 Schär, C., Vasilina, L., Pertziger, F. and Dirren, S.: Seasonal Runoff Forecasting Using Precipitation from
2 Meteorological Data Assimilation Systems, *J. Hydrometeor.*, 5(5), 959–973, doi:10.1175/1525-
3 7541(2004)005<0959:SRFUPF>2.0.CO;2, 2004.
- 4 Schiemann, R., Lüthi, D., Vidale, P. L. and Schär, C.: The precipitation climate of Central Asia—intercomparison of
5 observational and numerical data sources in a remote semiarid region, *Int. J. Climatol.*, 28(3), 295–314,
6 doi:10.1002/joc.1532, 2008.
- 7 Seibert, M., Apel, H. and Merz, B.: Seasonal forecasting of hydrological drought in the Limpopo basin: A comparison of
8 statistical methods, *Hydrol. Earth Syst. Sci. Discuss.* doi:10.5194/hess-2016-4, 2016, 2016.
- 9 Shirvani, A. and Landman, W. A.: Seasonal precipitation forecast skill over Iran, *Int. J. Climatol.*, n/a–n/a,
10 doi:10.1002/joc.4467, 2015.
- 11 Sigdel, M. and Ikeda, M.: Summer Monsoon Rainfall over Nepal Related with Large-Scale Atmospheric Circulations,
12 *Journal of Earth Science & Climatic Change*, 2012, doi:10.4172/2157-7617.1000112, 2013.
- 13 Smith, D. M., Scaife, A. A. and Kirtman, B. P.: What is the current state of scientific knowledge with regard to seasonal
14 and decadal forecasting?, *Environmental Research Letters*, 7(1), 15602–15612, doi:10.1088/1748-9326/7/1/015602,
15 2012.
- 16 Smith, T. M. and Reynolds, R. W.: Extended Reconstruction of Global Sea Surface Temperatures Based on COADS
17 Data (1854–1997), *J. Climate*, 16(10), 1495–1510, doi:10.1175/1520-0442-16.10.1495, 2003.
- 18 Smith, T. M., Reynolds, R. W., Peterson, T. C. and Lawrimore, J.: Improvements to NOAA’s Historical Merged Land–
19 Ocean Surface Temperature Analysis (1880–2006), *J. Climate*, 21(10), 2283–2296, doi:10.1175/2007JCLI2100.1, 2008.
- 20 Stone, R. C., Hammer, G. L. and Marcussen, T.: Prediction of global rainfall probabilities using phases of the Southern
21 Oscillation Index, *Nature*, 384(6606), 252–255, doi:10.1038/384252a0, 1996.
- 22 Suárez-Moreno, R. and Rodríguez-Fonseca, B.: S4CAST v2.0: sea surface temperature based statistical seasonal
23 forecast model, *Geosci. Model Dev.*, 8(11), 3639–3658, doi:10.5194/gmd-8-3639-2015, 2015.
- 24 Surendran, S., Gadgil, S., Francis, P. A. and Rajeevan, M.: Prediction of Indian rainfall during the summer monsoon
25 season on the basis of links with equatorial Pacific and Indian Ocean climate indices, *Environ. Res. Lett.*, 10(9),
26 094004, doi:10.1088/1748-9326/10/9/094004, 2015.
- 27 Syed, F. S., Giorgi, F., Pal, J. S. and King, M. P.: Effect of remote forcings on the winter precipitation of central
28 southwest Asia part I: observations, *Theor. Appl. Climatol.*, 86(1-4), 147–160, doi:10.1007/s00704-005-0217-1, 2006.
- 29 Syed, F. S., Giorgi, F., Pal, J. S. and Keay, K.: Regional climate model simulation of winter climate over Central–
30 Southwest Asia, with emphasis on NAO and ENSO effects, *Int. J. Climatol.*, 30(2), 220–235, doi:10.1002/joc.1887,
31 2010.
- 32 Tian, B. and Fan, K.: A Skillful Prediction Model for Winter NAO Based on Atlantic Sea Surface Temperature and
33 Eurasian Snow Cover, *Wea. Forecasting*, 30(1), 197–205, doi:10.1175/WAF-D-14-00100.1, 2015.
- 34 Ummenhofer, C. C., England, M. H., McIntosh, P. C., Meyers, G. A., Pook, M. J., Risbey, J. S., Gupta, A. S. and
35 Taschetto, A. S.: What causes southeast Australia’s worst droughts?, *Geophys. Res. Lett.*, 36(4), L04706,
36 doi:10.1029/2008GL036801, 2009.
- 37 Wang, C.: Atmospheric Circulation Cells Associated with the El Niño–Southern Oscillation, *J. Climate*, 15(4), 399–419,
38 doi:10.1175/1520-0442(2002)015<0399:ACCAWT>2.0.CO;2, 2002.
- 39 Wang, H. and He, S.: Weakening relationship between East Asian winter monsoon and ENSO after mid-1970s, *Chin.*
40 *Sci. Bull.*, 57(27), 3535–3540, doi:10.1007/s11434-012-5285-x, 2012.
- 41 Wulf, H., Bookhagen, B. and Scherler, D.: Seasonal precipitation gradients and their impact on fluvial sediment flux in
42 the Northwest Himalaya, *Geomorphology*, 118(1–2), 13–21, doi:10.1016/j.geomorph.2009.12.003, 2010.

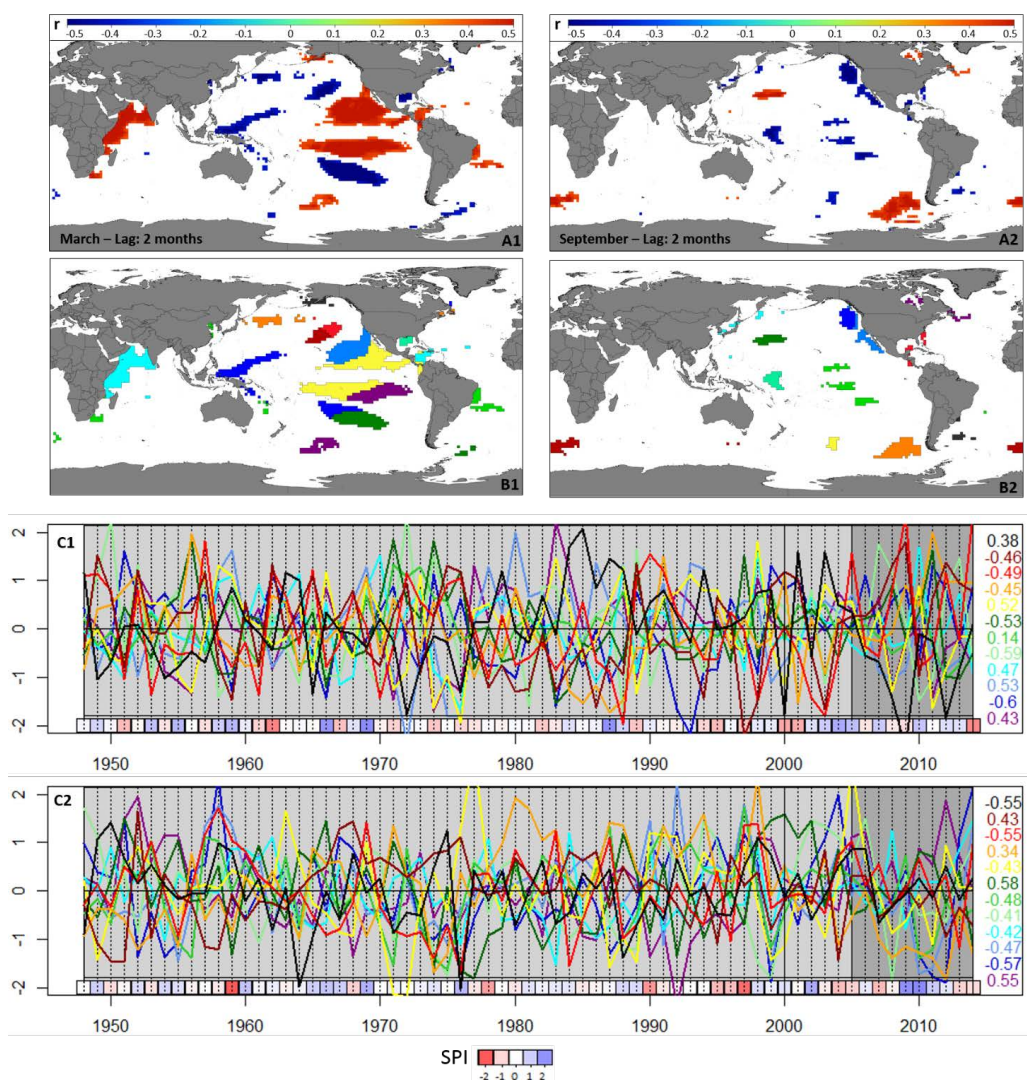


- 1 Wu, T.-W. and Qian, Z.-A.: The Relation between the Tibetan Winter Snow and the Asian Summer Monsoon and
2 Rainfall: An Observational Investigation, *J. Climate*, 16(12), 2038–2051, doi:10.1175/1520-
3 0442(2003)016<2038:TRBTTW>2.0.CO;2, 2003.
- 4 Wu, Z. and Lin, H.: Interdecadal variability of the ENSO–North Atlantic Oscillation connection in boreal summer,
5 *Q.J.R. Meteorol. Soc.*, 138(667), 1668–1675, doi:10.1002/qj.1889, 2012.
- 6 Wu, Z., Wang, B., Li, J. and Jin, F.-F.: An empirical seasonal prediction model of the East Asian summer monsoon using
7 ENSO and NAO, *Journal of Geophysical Research: Atmospheres* (1984–2012), 114(D18) [online] Available from:
8 <http://onlinelibrary.wiley.com/doi/10.1029/2009JD011733/full> (Accessed 17 December 2015), 2009.
- 9 Yadav, R. K., Yoo, J. H., Kucharski, F. and Abid, M. A.: Why Is ENSO Influencing Northwest India Winter
10 Precipitation in Recent Decades?, *J. Climate*, 23(8), 1979–1993, doi:10.1175/2009JCLI3202.1, 2010.
- 11 Yim, S.-Y., Wang, B., Liu, J. and Wu, Z.: A comparison of regional monsoon variability using monsoon indices, *Clim
12 Dyn*, 43(5-6), 1423–1437, doi:10.1007/s00382-013-1956-9, 2013.
- 13 Yin, Z.-Y., Wang, H. and Liu, X.: A Comparative Study on Precipitation Climatology and Interannual Variability in the
14 Lower Midlatitude East Asia and Central Asia, *J. Climate*, 27(20), 7830–7848, doi:10.1175/JCLI-D-14-00052.1, 2014.
- 15 Zhang, Y., Li, T. and Wang, B.: Decadal Change of the Spring Snow Depth over the Tibetan Plateau: The Associated
16 Circulation and Influence on the East Asian Summer Monsoon*, *J. Climate*, 17(14), 2780–2793, doi:10.1175/1520-
17 0442(2004)017<2780:DCOTSS>2.0.CO;2, 2004.
- 18 Zorita, E., Hughes, J. P., Lettenmaier, D. P. and von Storch, H.: Stochastic Characterization of Regional Circulation
19 Patterns for Climate Model Diagnosis and Estimation of Local Precipitation, *J. Climate*, 8(5), 1023–1042,
20 doi:10.1175/1520-0442(1995)008<1023:SCORCP>2.0.CO;2, 1995.
- 21



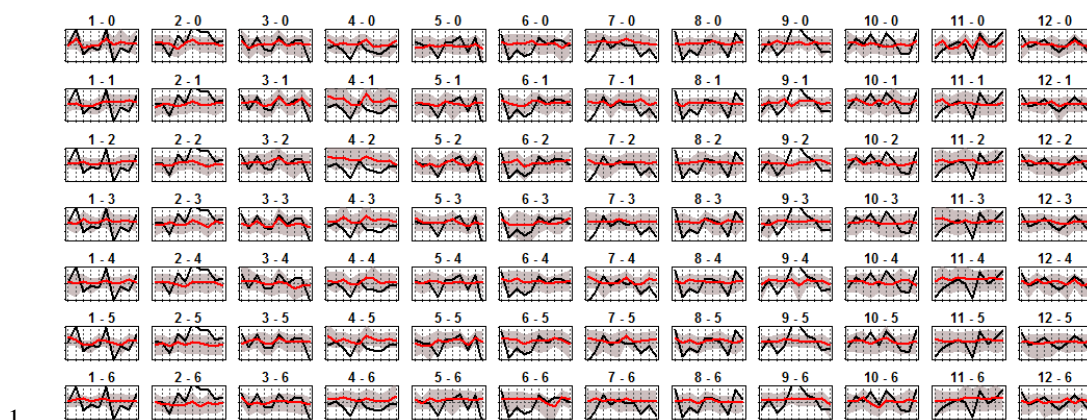
1

Fig 1: Flow Chart representing the major components of the seasonal forecast model.



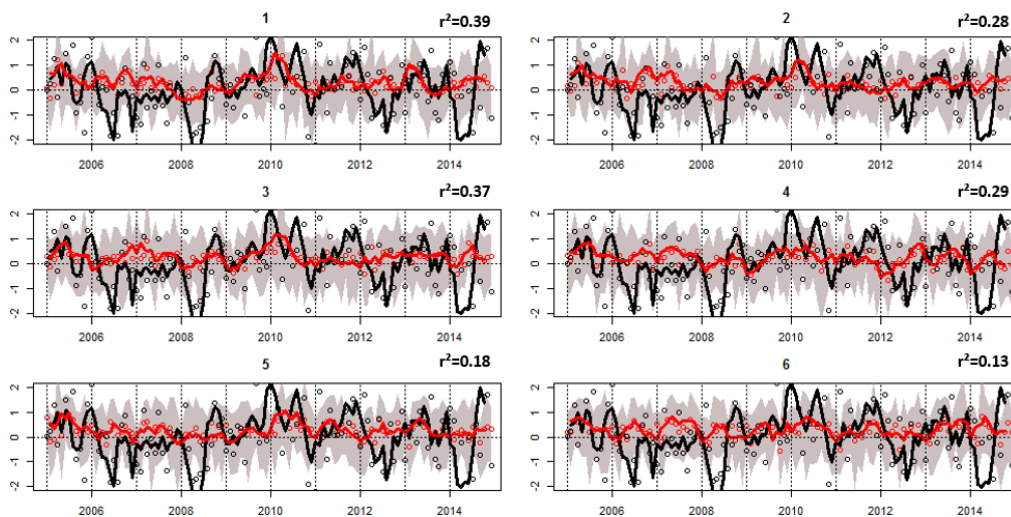
1

Fig.2: Exemplary representation of SST grid cells which are significantly correlated with precipitation anomalies of the Naryn basin for March and September with a lead time of two months (A1 and A2) and the aggregation to predictor regions based on the Hill Climbing k-means cluster analysis (B1 and B2). The diagrams (C1 and C2) show the time series of z-normalized mean SSTs during the selected months for each of the cluster regions (same color) and the subsequent hydroclimatic variations in the Naryn catchment (expressed as red to blue rectangles, indicating SPI values between -2 to 2). The colored values on the right indicate the correlation of mean cluster SST anomalies and the corresponding SPI values.



1

Fig. 3: Results of the random forest based monthly precipitation forecast models (red) and observations (black) for the Naryn catchment for the period from 2005 to 2014 (x-axis). Values are displayed as monthly SPI values between -2 and 2. The numbers indicate the month for which the forecast is conducted and the particular lead time (e.g. 1-1 shows the results for January precipitation based on predictor variables from December). The shaded area indicates the range of prediction values of all single tree models belonging to the random forest forecast model.



2

Fig. 4: Time series of observed (black line) and forecasted (red line) running three-months SPI values for the Naryn catchment. The shaded areas indicates the three month total of maximum and minimum forecasts of single trees of the random forest model. Points show the observations and forecast at a one-monthly scale.

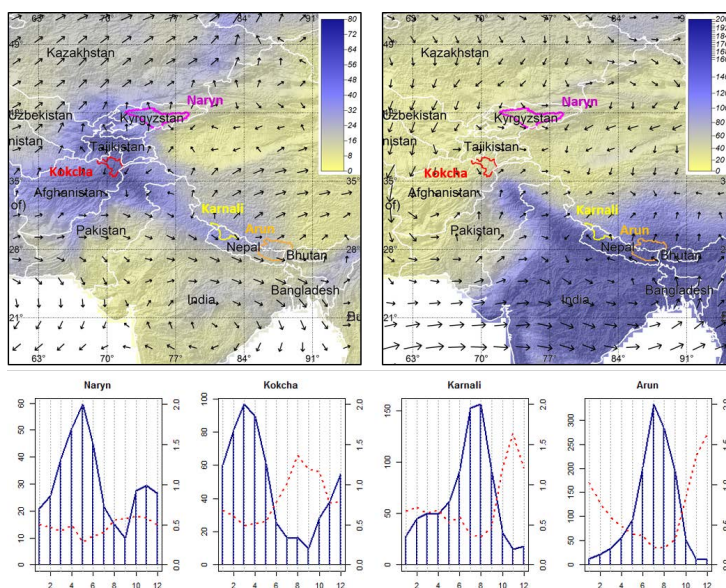
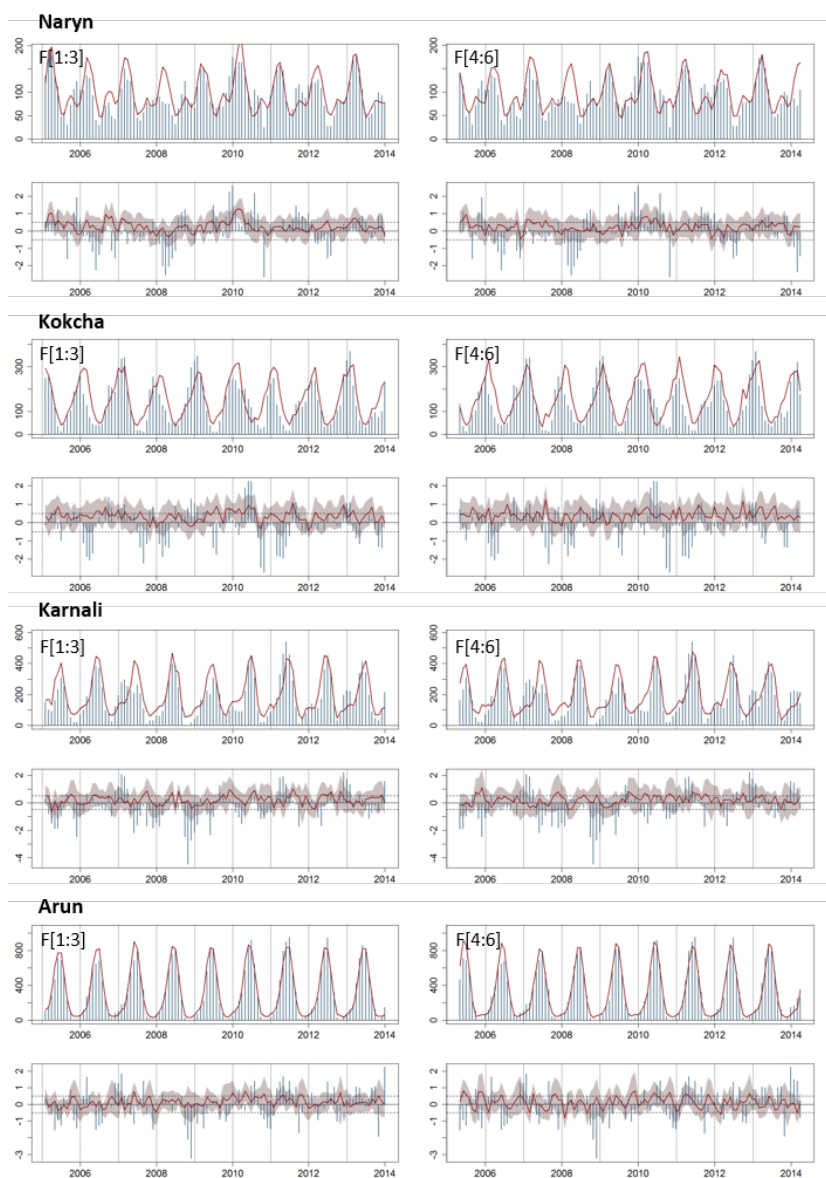


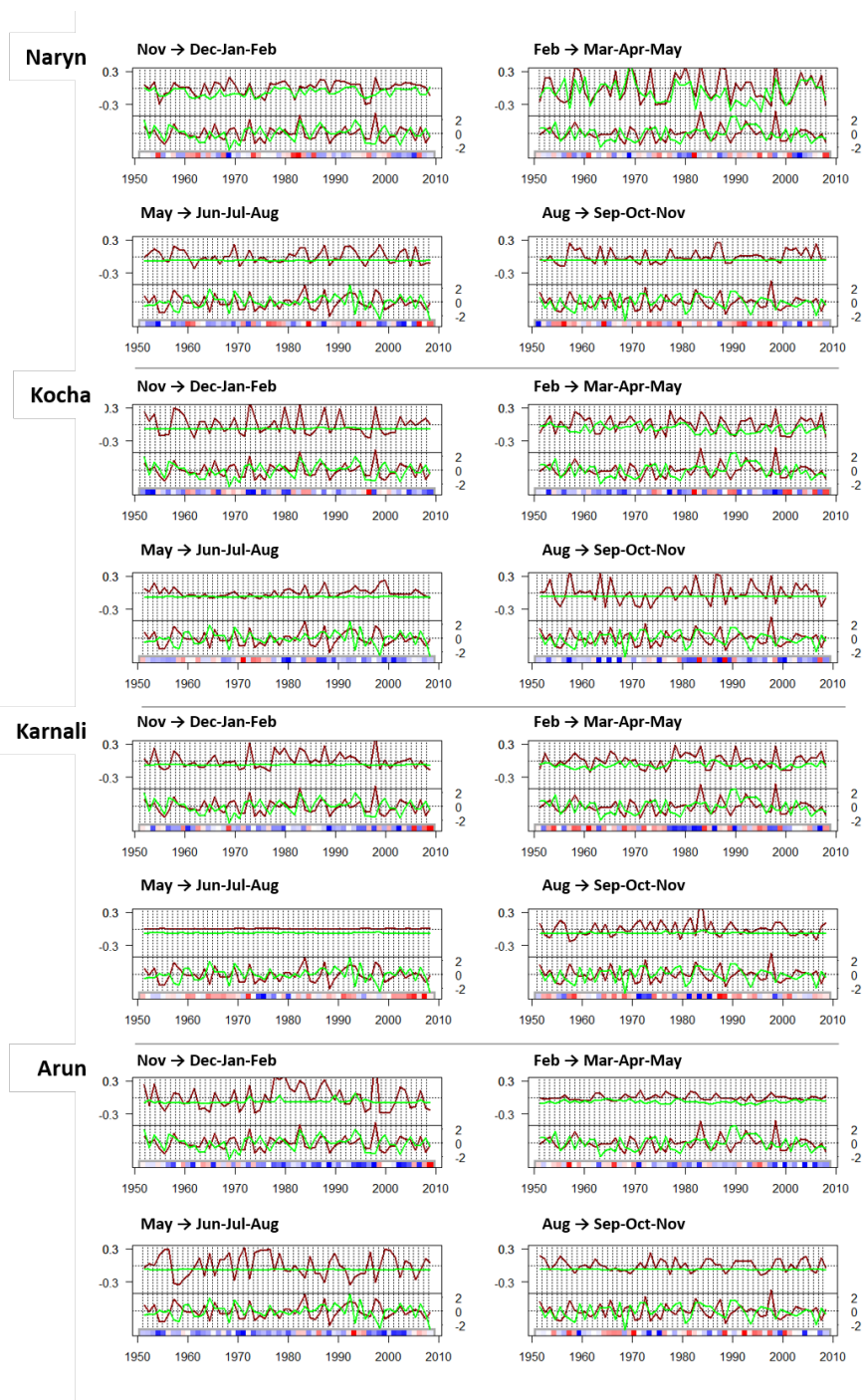
Fig. 5: Location of selected target areas as well as the mean precipitation total [mm] (CRU-TS) and the mean 850 hpa wind field (NCEP-NCAR) during DJF (left) and JJA (right).

Diagrams show the mean monthly precipitation amount for every catchment in mm (blue bars) as well as the Coefficient of Variation (red line)



1

Fig. 6: Observed running three-months precipitation totals (blue bars) and modelling results (red line) of the F[1:3] and the F[4:6] model for Naryn, Kokcha, Karnali and Arun. The upper panels show absolute precipitation totals for running three-months periods, the lower panel show the corresponding SPI index for each three-month period respectively. Shaded areas indicate the 90% interval of the residual based probabilistic forecast.



1

Fig. 7: Response of the F[1:3] model to predictors associated to ENSO (red) and NAO (green) for different seasons and target areas (upper panels of each diagram). Time series of the considered climate indices (lower panel) for the month of forecast generation.



<u>Acronym</u>	<u>Variable Name</u>	<u>Unit</u>	<u>Source</u>	<u>Spatial Resolution</u>
SST	Sea Surface Temperature	°C	ERSST V3b	2°x2°
SLP	Sea Level Pressure	hPa	NCAR-reanalysis	2.5°x2.5°
GPH500	Geopotential Height at 500 hPa	m	NCAR-reanalysis	2.5°x2.5°
GPH500-200	Geopotential Thickness between 500 and 200 hPa	m	NCAR-reanalysis	2.5°x2.5°
TEMP	Near Surface Temperature	°C	NCAR-reanalysis	1.875°x1.904°
PREC	Previous Precipitation	mm	NCAR-reanalysis	1.875°x1.904°
SWE	Snow Water Equivalent	mm	NCAR-reanalysis	1.875°x1.904°

Tab. 1: Globally gridded variables utilized as potential predictor variables by the statistical forecast model

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25



Naryn													
Month	Forecast Range	COR	BIAS[%]	RMSE[%]	AUC(-0.5)	AUC(0.5)	Month	Forecast Range	COR	BIAS[%]	RMSE[%]	AUC(-0.5)	AUC(0.5)
1	2-4	0.71	0.07	0.13	1.00	0.57	1	5-7	0.29	0.09	0.20	0.56	0.55
2	3-5	0.68	0.16	0.20	1.00	0.86	2	6-8	0.50	0.12	0.26	0.64	0.83
3	4-6	0.70	0.16	0.20	0.80	0.86	3	7-9	0.50	0.10	0.28	0.94	0.78
4	5-7	0.58	0.07	0.18	0.83	0.45	4	8-10	0.15	0.12	0.44	0.43	0.25
5	6-8	0.53	0.10	0.25	0.86	0.83	5	9-11	0.35	0.04	0.31	0.86	0.56
6	7-9	0.08	0.07	0.31	0.56	0.39	6	10-12	0.26	0.02	0.24	--	0.36
7	8-10	0.31	0.13	0.44	0.21	0.30	7	11-1	0.10	0.07	0.32	0.14	0.45
8	9-11	0.22	0.02	0.38	0.50	0.22	8	12-2	0.66	0.26	0.42	0.88	0.55
9	10-12	0.28	0.05	0.25	--	0.36	9	1-3	0.53	0.20	0.40	0.38	0.50
10	11-1	0.19	0.03	0.35	0.36	0.65	10	2-4	0.39	0.08	0.18	0.64	1.00
11	12-2	0.23	0.22	0.41	0.25	0.70	11	3-5	0.46	0.17	0.25	0.78	0.63
12	1-3	0.86	0.16	0.31	0.75	0.72	12	4-6	0.50	0.18	0.24	0.90	0.88

Kokcha													
Month	Forecast Range	COR	BIAS[%]	RMSE[%]	AUC(-0.5)	AUC(0.5)	Month	Forecast Range	COR	BIAS[%]	RMSE[%]	AUC(-0.5)	AUC(0.5)
1	2-4	0.27	0.07	0.25	0.64	0.44	1	5-7	0.39	0.22	0.44	0.95	0.50
2	3-5	0.30	0.26	0.35	0.55	0.75	2	6-8	0.19	0.12	0.67	1.00	0.13
3	4-6	0.45	0.21	0.33	0.80	0.43	3	7-9	0.06	0.10	0.80	0.65	0.64
4	5-7	0.54	0.16	0.43	0.75	0.79	4	8-10	0.18	0.32	0.48	0.50	0.57
5	6-8	0.46	0.15	0.58	0.38	1.00	5	9-11	0.53	0.19	0.37	0.79	0.64
6	7-9	0.41	0.10	0.73	0.60	0.50	6	10-12	0.12	0.22	0.43	0.86	0.75
7	8-10	0.28	0.31	0.45	0.21	0.93	7	11-1	0.25	0.06	0.40	0.50	0.15
8	9-11	0.45	0.16	0.37	0.93	0.79	8	12-2	0.04	0.03	0.29	0.57	0.50
9	10-12	0.75	0.07	0.37	0.71	1.00	9	1-3	0.21	0.01	0.19	--	0.36
10	11-1	0.12	0.10	0.42	0.57	0.25	10	2-4	0.22	0.14	0.27	0.36	0.56
11	12-2	0.12	0.02	0.30	0.57	0.70	11	3-5	0.18	0.22	0.35	0.55	0.57
12	1-3	0.14	0.03	0.19	--	0.36	12	4-6	0.03	0.24	0.37	0.40	0.00

Karnal 1													
Month	Forecast Range	COR	BIAS[%]	RMSE[%]	AUC(-0.5)	AUC(0.5)	Month	Forecast Range	COR	BIAS[%]	RMSE[%]	AUC(-0.5)	AUC(0.5)
1	2-4	0.42	0.02	0.48	0.50	0.71	1	5-7	0.24	0.06	0.31	0.70	0.57
2	3-5	0.05	0.06	0.47	0.22	0.67	2	6-8	0.66	0.17	0.27	0.78	1.00
3	4-6	0.49	0.00	0.38	0.50	0.86	3	7-9	0.50	0.10	0.17	0.80	0.61
4	5-7	0.30	0.09	0.35	0.45	0.14	4	8-10	0.32	0.12	0.19	0.64	0.50
5	6-8	0.45	0.16	0.28	0.83	0.63	5	9-11	0.16	0.00	0.31	0.56	0.21
6	7-9	0.58	0.10	0.17	0.80	0.94	6	10-12	0.11	0.48	0.66	0.50	0.57
7	8-10	0.53	0.17	0.21	0.71	0.63	7	11-1	0.01	0.51	0.60	0.39	--
8	9-11	0.13	0.12	0.30	0.50	0.43	8	12-2	0.73	0.19	0.71	0.00	0.17
9	10-12	0.23	0.49	0.63	0.50	0.79	9	1-3	0.35	0.01	0.55	0.40	0.11
10	11-1	0.33	0.46	0.56	0.50	--	10	2-4	0.00	0.03	0.54	0.70	0.44
11	12-2	0.40	0.13	0.56	0.80	0.50	11	3-5	0.36	0.17	0.56	0.50	0.50
12	1-3	0.45	0.02	0.52	0.80	0.61	12	4-6	0.15	0.03	0.41	0.29	0.43

Arun													
Month	Forecast Range	COR	BIAS[%]	RMSE[%]	AUC(-0.5)	AUC(0.5)	Month	Forecast Range	COR	BIAS[%]	RMSE[%]	AUC(-0.5)	AUC(0.5)
1	2-4	0.38	0.09	0.30	0.71	0.56	1	5-7	0.24	0.01	0.14	0.61	0.57
2	3-5	0.13	0.20	0.30	--	0.65	2	6-8	0.03	0.11	0.14	0.55	0.25
3	4-6	0.65	0.00	0.19	1.00	0.78	3	7-9	0.33	0.00	0.13	0.29	0.33
4	5-7	0.19	0.05	0.15	0.56	0.57	4	8-10	0.37	0.06	0.10	0.50	0.71
5	6-8	0.23	0.07	0.12	0.55	0.38	5	9-11	0.04	0.23	0.32	0.25	0.25
6	7-9	0.50	0.01	0.11	0.57	0.44	6	10-12	0.29	0.16	0.64	0.45	0.36
7	8-10	0.70	0.02	0.07	1.00	0.86	7	11-1	0.36	0.74	0.77	0.44	--
8	9-11	0.29	0.17	0.26	0.75	0.75	8	12-2	0.09	0.44	0.68	0.35	0.63
9	10-12	0.05	0.24	0.57	0.50	0.36	9	1-3	0.40	0.10	0.51	1.00	0.79
10	11-1	0.21	0.73	0.76	0.50	--	10	2-4	0.53	0.28	0.44	0.64	0.90
11	12-2	0.33	0.27	0.67	0.55	0.75	11	3-5	0.40	0.30	0.37	--	0.78
12	1-3	0.19	0.12	0.60	0.21	0.43	12	4-6	0.03	0.05	0.23	0.39	0.33

1
 2 *Tab. 2: Performance Measures for the F[1:3] and the F[4:6] model for all selected target areas. In order to better*
 3 *visualize the results, correlation and AUC values are colored using a color gradient from dark green (correlations and*
 4 *AUC near +1) and darkred (correlations near -1, AUC near 0). If no AUC values is provided, no drought/moist event*
has been observed during the evaluation period.