A statistically based seasonal precipitation forecast model with automatic predictor selection and its application to Central and South Asia

Lars Gerlitz, Sergiy Vorogushyn, Heiko Apel, Abror Gafurov, Katy Unger-Shayesteh, & Bruno Merz
GFZ German Research Centre for Geosciences, Section 5.4: Hydrology, Telegrafenberg, 14473 Potsdam, Germany

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

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

The model is applied and evaluated for selected target regions in Central and South Asia. Particularly for winter and spring in westerlies-dominated Central Asia, correlation coefficients between forecasted and observed precipitation reach values up to 0.48, although the variability of precipitation rates is strongly underestimated. Likewise, for the monsoonal precipitation amounts in the South Asian target area correlations of up to 0.5 were detected. The skill of the model for the dry winter season over South Asia is found to be low.

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

1 Introduction

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

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

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

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

Particularly the state of the El Nino Southern oscillation (ENSO) is known to influence the large-scale precipitation patterns almost everywhere on the globe (Dai and Wigley, 2000; Mason and Goddard, 2001; Stone et al., 1996). The precipitation variability in the tropical regions is directly determined by ENSO due to its impact on the tropical Walker Circulation. During El Nino events, positive SST anomalies occur over the eastern tropical Pacific as a result of weakened easterly trade winds. A common consequence is the occurrence of drought periods in South East Asia, especially over

Indonesia, and the simultaneous presence of long lasting precipitation events over the arid regions of the western slopes of the South American Andes (Julian and Chervin, 1978; Wang, 2002). However, several studies demonstrated a statistically significant correlation of El Nino-Indices (usually derived from SST-observations in the El Nino core regions or from associated pressure gradients between Darwin and Tahiti) with seasonal precipitation time series in other parts of the tropics and also in temperate climate zones. For example. various studies detected a robust statistical relationship between Australian monsoonal precipitation and the ENSO state during previous months (Cai et al., 2011; Ummenhofer et al., 2009). A significant influence of El Nino events was also found for monsoonal precipitation amounts in Eastern and Southern Africa (Liebmann et al., 2014; Ratnam et al., 2014) and the Sahel region (Parhi et al., 2015). For the South Asian monsoon a negative response to El Nino events has been frequently perceived (Krishnaswamy et al., 2014; Lau and Wu, 2001; Surendran et al., 2015). For the semi-arid regions of Central Asia and for the Mediterranean region a positive relationship of winter and spring precipitation to El Nino events during previous autumn was found e.g. by Barlow et al. (2002); Hoell et al. (2013); Roghani et al. (2015) and Syed et al. (2006). Moreover, Fraedrich (1994) and Wu and Lin (2012) detected a statistically significant influence on extra tropical circulation anomalies such as the position of large-scale Rossby waves and the associated North Atlantic Oscillation. This subsequently leads to a certain impact of El Nino events on the European winter climate, although correlations are in general less robust compared with tropical regions. Other tropical SST modes frequently used in seasonal forecasts include the Indian Ocean Dipole (IOD), the Atlantic Multi-decadal Oscillation (AMO) and the Pacific Decadal Oscillation (PDA), which have a significant predictive skill for their adjacent coastal regions (Eden et al., 2015).

Numerous studies additionally used customized SST indices as predictor variables for seasonal precipitation forecasts. For example Schepen e. al. (2011) give a comprehensive overview of oceanic and atmospheric climate indices with predictive potential for seasonal rainfall amounts in Australia. They illustrate that oceanic indices from the pacific region comprise a high forecast skill, particularly for autumn and winter precipitation totals. Hartmann et al. (2016) tested the predictive skill of mean SSTs from various ocean basins surrounding the Asian continent for the precipitation variability in the arid Tarim basin in North Western China. Hertig and Jacobeit (2010) investigated the predictive skill of EOF-derived SST patterns of the Northern Atlantic, in order to forecast winter precipitation amounts in the Mediterranean. Seibert et al., (2016) recently demonstrated that customized SST indices from the Indian and Southern Atlantic Ocean improve the quality of statistical seasonal forecasts for the Limpopo-basin in Southern Africa. Suárez-Moreno and Rodríguez-Fonseca (2015) showed that particularly for coastal regions, adjacent Sea surface temperatures can significantly improve the seasonal forecast of precipitation.

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

Local land cover characteristics are also frequently applied in statistical seasonal forecast models. For example, Cohen and Entekhabi (1999) and Cohen and Barlow (2005) showed that the snow cover over Eurasia during autumn and spring alters the large-scale atmospheric circulation over the Northern hemisphere with wide implications on precipitation patterns during subsequent months. Brands et al. (2012) reported a statistical significant relationship between late autumn snow cover over Eurasia and winter precipitation over Europe. Tian and Fan (2015) argued that the state of the NAO and

the associated precipitation patterns over Europe are influenced by both Atlantic SSTs and snow cover rates over Eurasia. Some studies indicate a negative response of the South Asian monsoon to higher snow cover rates over Eurasia, most likely due to a delayed surface heating of the Asian continent (Wu and Qian, 2003; Zhang et al., 2004). Recently some studies also included local soil moisture or previous rainfall into statistical forecasting models in order to capture water recycling due to autochthonous weather conditions and persistent circulation characteristics (Eden et al., 2015; van den Hurk et al., 2010).

As shown, most statistical forecast applications utilize either well-known climate indices or expert knowledge based customized indices from SSTs or land cover characteristics. Customized indices are frequently included, since typical climate indices do not cover regional scale anomalies of SST or pressure patterns which might be important predictor variables for seasonal climate forecasts in certain regions. However, these customized indices are usually calibrated with regard to specific target areas and thus are not transferable to any other regions. Hence state of the art seasonal climate forecast models are either based on a fixed number of climate indices (and thus might not consider important predictor variables) or are highly site specific and barely transferable to other regions. Recently some advances towards an automatic predictor selection were made by Suárez-Moreno and Rodríguez-Fonseca (2015), who used gridded SST fields as potential predictors in order to automatically identify SST patterns, which are relevant for the seasonal precipitation forecast in selected target areas.

With the aim of developing an operational seasonal forecast model, which is easily transferable to any region in the world, we present a generic data mining approach which automatically selects potential predictors from gridded SST observations and large-scale atmospheric circulation patterns derived from reanalysis data. Subsequently the approach generates robust statistical relationships with posterior precipitation anomalies for user selected target regions. The statistical package R (R Development Core Team, 2008) as well as the scripting environment of the free and open source GIS system SAGA (Conrad et al., 2015) are utilized. The precipitation forecast model is based on a cell-wise correlation analysis of various gridded variables with regional precipitation estimates, which identifies grid cells with potential predictive skill for a specific target area with different time lags. Grid cells, which significantly correlate with precipitation anomalies during subsequent months, are aggregated to predictor regions by means of an automatic cluster analysis for every variable and time lag. Thus, for every target area, specific predictor variables are automatically derived. The cluster regions are afterwards utilized as potential predictors in a non-parametric and non-linear random forest based modelling approach. Based on 4-fold split sample test, the model performance for the selected target area is evaluated before an operational forecast is generated based on real time predictor fields.

In the following section, we provide a detailed overview about the utilized data sets and the main model components, including predictor selection, model calibration and evaluation. Subsequently we provide some applications of the model for selected target areas in Central and South Asia. In order to make the individual modelling steps more comprehensible, we already provide some major interim results for one target area in Northern Central Asia (Figure 5) when explaining the methods in the next section.

2 Methods and Data

2.1 Modelling Structure

The major objective of the presented model is to derive suitable predictor variables from global oceanic and atmospheric fields and to develop robust statistical relationships which enable a seasonal precipitation forecast for user selectable target regions. The underlying data sets as well as the major model components are summarized in Fig. 1. In order to

analyze the precipitation variability in selected target areas the model is based on the CRU TS 2.0 precipitation data set, which provides monthly precipitation estimates for the 20th century on a global grid with a resolution of 0.5° lat./long. (Harris et al., 2014; New et al., 1999). The data set is based on a dense network of observations for the period from 1961 to 1990, which were used for the regionalization of monthly mean precipitation amounts, and a compilation of station records with longer time series available, which were used for the calculation of anomalies and were subsequently spatially interpolated based on inverse distances. New et al. (1999) showed that this approach is suitable for the resolution of 0.5° since it combines a climatic baseline, which is highly influenced by the underlying topography with simple interpolated anomalies, which are mainly driven by large-scale weather conditions. Areal mean monthly precipitation sums for the selected target region are extracted from the CRU data set. Due to a temporally varying number of stations used for the interpolation of gridded precipitation estimates, the data may incorporate inhomogenities in some regions. Thus a Standardized Normal Homogeneity Test (SNHT) for absolute annual values (Wijngaard et al., 2003) is conducted, which identifies abrupt changes of the annual precipitation sums. The results serve as a background information for the interpretation of the model results.

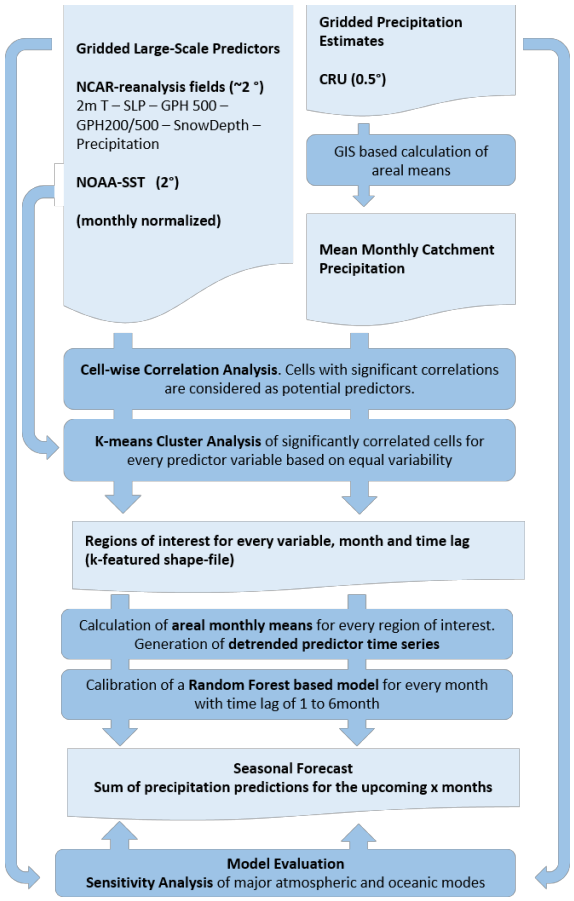


Fig 1: Flow chart representing the major components of the seasonal forecast model.

Since monthly time series of precipitation are usually positively skewed, which might not compromise the assumptions of the subsequent correlation analysis, the actual values are converted into the Standardized Precipitation Index (SPI) (Guttman, 1998; McKee et al., 1993) for every single month of the year. Therefore, the precipitation distribution for each month is fitted to a gamma distribution with suitable shape and scale parameters. The exceedance probability of observed precipitation amounts is then converted into z-values of the normal distribution. The SPI values, which are normally

distributed by definition, are subsequently cell-wise correlated with gridded global SST and climate data with lead times ranging from 1 to 6 months. For every variable and lead time grid cells are identified, which significantly correlate with the mean monthly SPI time series. These grid cells are subsequently aggregated to predictor regions with similar variability by means of a Hill-Climbing based k-means cluster analysis. For every large-scale variable and time lag, the areal mean anomalies of those cluster regions are considered as potential predictor variables for a random forest based precipitation forecast model. All data sets (predictants and predictor variables) are automatically processed for the period from 1948 to 2014. In order to find robust predictor variables for monthly precipitation amounts and to exclude incidental correlations, the data set is randomly partitioned into two subsets. One is utilized for the cell-wise correlation analysis, the second one is employed for the subsequent calibration of a random forest based forecast model. Since precipitation usually shows a rather random temporal variability at a monthly time scale, results of the monthly precipitation forecast are in general unreliable. Thus, modelling results are aggregated to running three-month precipitation totals.

2.2 Predictor selection

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

In order to reveal the influence of nearby or remote SST anomalies on precipitation characteristics, we make use of the NOAA Extended Global Sea Surface Temperature ERSST V3b (Smith et al., 2008; Smith and Reynolds, 2003), which is available at a resolution of 2°x2° for the period from 1854 onwards. The data set is based on in situ sea surface temperature observations only, which are regionalized by means of statistical methods, considering both, low and high frequency oceanic modes. With the aim of avoiding statistical artefacts, resulting from the variability of the sea ice extent in polar oceans, we restricted the analysis of SST patterns to the geographical region between 65° N and 65° S. Further we utilize variables representing the state of the large-scale atmospheric circulation from the NCAR-NCEP reanalysis (Kalnay et 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 observations with climate modelling results by means of a 4d-assimilation system for the period from 1948 onwards. As potential atmospheric predictors we used monthly aggregated values of sea level pressure (SLP), the 500 hPa geopotential height (GPH500) as well as the geopotential thickness between the 500 hPa and 200 hPa pressure level (GPH500-200). In order to investigate the land surface characteristics and their subsequent effects, we additionally utilize monthly aggregated global grids of near surface temperature (TEMP), antecedent precipitation amounts (PREC) and snow water equivalent (SWE) from the NCAR-NCEP reanalysis as potential predictor variables. While the intrinsic pressure related variables are provided at a spatial resolution of 2.5°x2.5° lat./long., the diagnostic land surface variables are available at a resolution of approximately 1.875°x1.904°. All predictor fields are cell-wise normalized for every month respectively. Since the utilized large-scale predictor variables are updated regularly and freely downloadable they are suitable for the development of an operational seasonal precipitation forecast system.

Acronyme	Variable Name	Unit	Source	Spatial Resolution
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

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

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

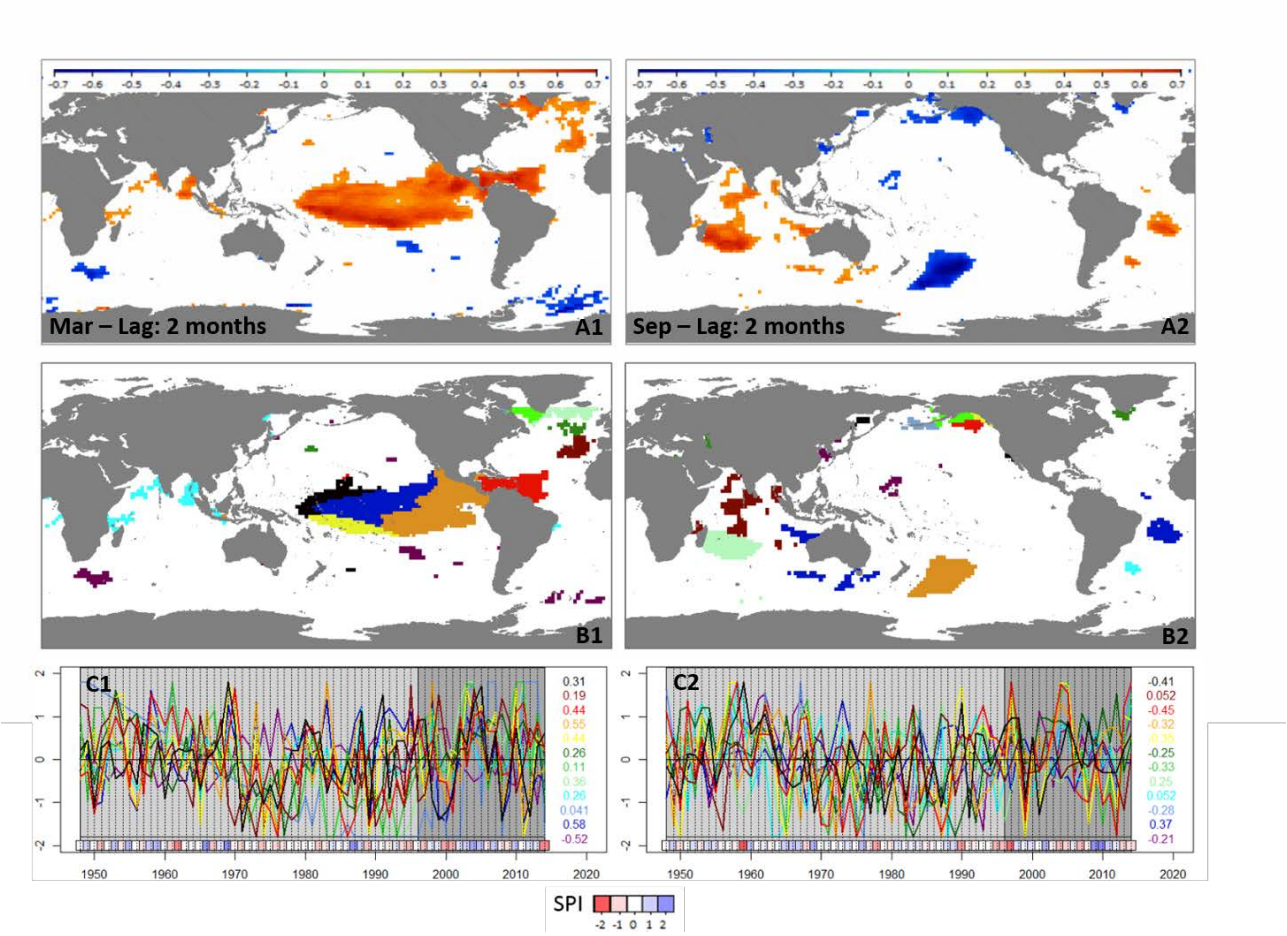


Fig.2: Correlation analysis results for precipitation anomalies of Northern Central Asia for March and September with a lead time of two months A1 and A2 show SST grid cells which are significantly correlated. B1 and B2 show the aggregation to predictor regions based on the Hill Climbing k-means cluster analysis. 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 Northern Central Asia (expressed as red to blue rectangles, indicting 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.

During March (representing the wet season in Northern Central Asia) monthly precipitation shows a clear positive response to January SST variations in the El Nino core region – a result which has been frequently reported for Central Asia – and to SST anomalies in the Arabian Sea and the Bay of Bengal. Further, a positive correlation with SST anomalies in the North Atlantic has been detected. During September (representing rather dry climate conditions) a positive response to SST variations in the Indian Ocean is evident, however the spatial distribution of potential predictive SST-regions is rather scattered, indicating less robust statistical relationships.

In a subsequent step each of the correlation grids (which usually contain a large number of potentially predictive grid cells) is aggregated to a distinct number of correlation regions, by means of a SAGA-GIS based Hill-Climbing k-means Cluster Analysis. (Hartigan & Wong, 1979). For every specific month, time lag and predictor variable, the complete normalized time series of all potentially predictive grid cells is considered. The iterative and unsupervised classification technique firstly randomly allocates every grid cell to one of *k* clusters. The error sum of squares is calculated as the sum of Euclidian distances of all associated grid cells from the cluster centroid and displays the quality of the cluster estimation. Every grid cell is subsequently reallocated to the nearest cluster and cluster centroids and error terms are recalculated. This procedure is iteratively conducted, until the error sum of squares converges to its minimum value. Basically the

clustering algorithm minimizes the error sum of squares within the cluster groups and maximizes the error sum of squares among them. This leads to definition of regions with similar temporal variability during the calibration period and thus identifies important large-scale patterns of the considered predictor variable with high predictive potential for the seasonal precipitation forecast.

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

As shown in Fig. 2, the El Nino core regions in January (orange, blue, yellow and black clusters in B1) are identified as important regions for the forecast of monthly precipitation amounts in March for Northern Central Asia, for instance the prolonged 1999-2001 winter and spring drought in Central Asia is associated with negative anomalies of the ENSO-related predictor variables. In general dry periods usually coincide with La Nina events, characterized by negative SST anomalies. The January SST of the Arabian Sea and the Bay of Bengal is identified as an independent predictor variable for the precipitation amounts in March. For the precipitation variability in September the majority of predictive SST clusters is located in the Indian Ocean.

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

2.3 Forecast Model Calibration

For every month of the year and every lead time, one separate statistical forecast model is established based on the potential predictor variables derived from the correlation and cluster analysis. In order to avoid overfitting and to develop a robust regression relationship, the model calibration is based on the second random sample and thus is independent from the predictor selection procedure. Some of the potential predictor variables are highly correlated due to their association to the same phenomenon, e.g. the El Nino Southern Oscillation is manifested in various SST regions and significantly influences the large-scale pressure and precipitation patterns in many regions of the world. Additionally the distribution of potential predictor variables is unknown, e.g. precipitation or snow water equivalent are most likely extremely skewed and not normally distributed. Thus a reliable forecasting approach requires a non-parametric statistical technique, without any assumption concerning the distribution and statistical independence of predictor variables. We make use of a random forest based approach (Breiman, 2001), a widely utilized data mining technique, which stands out due to its flexibility concerning the characteristics of predictant and predictor variables and due to its ability to detect non-linear and conditional statistical relationships. Basically random forest models represent an advancement of regression tree algorithms (Breiman et al., 1984) which automatically classify large data sets by means of adequate predictor variables in order to identify statistical structures in the predictor space, which are highly associated with a response variable (Gerlitz, 2014; Zorita et al., 1995).

Classification is conducted by means of an iterative procedure. In every processing step one predictor variable and one split value are identified, which classify the learning sample into two sub groups, characterized by a maximal homogeneity (i.e. a minimum variance) of the predictant variable. However, since the recursive regression tree approach tends to considerably overfit the predictor-predictant relationships and does not only classify important structures within the feature space but also the inherent noise of the predictant variable, the predictive skill of single regression trees is

frequently insufficient. Therefore, random forest applications consider an ensemble of various trees, which are based on a subset of the complete data set respectively. By means of this bagging approach a large number of trees is constructed. Prediction values are eventually calculated as the mean of predictions from all single trees. The bagging approach and the ensemble composition of the final random forest model avoid overfitting and additionally provide an internal error and confidence estimation (Chen et al., 2012).

The specific forecast models for every month and lead time are constructed based on random forests with 500 realizations. Regression trees are recursively constructed until the final leaves include 3 observations or less. For the determination of each splitting criterion, a randomly selected bagging-sample 2/3 of the entire learning sample is utilized.

As the predictant variable the absolute amount of monthly precipitation is used. This allows the subsequent additive aggregation of the monthly forecast values to seasonal precipitation amounts and the evaluation of the model at different temporal scales. Fig. 3 shows as an example the results of the monthly precipitation forecast with varying lead times for the Northern Central Asian target area for an independent period 1996 to 2010. The remaining time series has been utilized for the predictor selection and the model calibration.

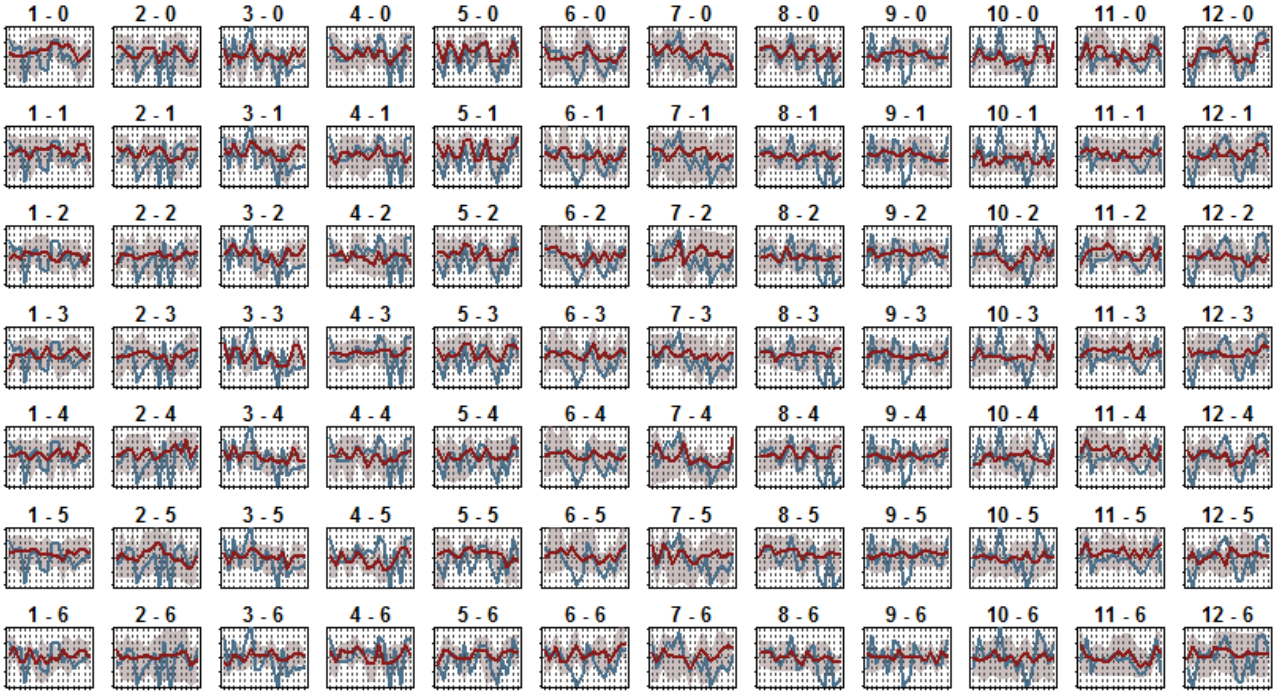


Fig. 3: Results of the random forest based monthly precipitation forecast models (red) and observations (blue) for the Northern Central Asia for the period from 1996 to 2010 (x-axis). Values are displayed as monthly SPI values between -2 and 2. The numbers indicate the month for which the forecast is conducted and the particular lead time (e.g. 1-2 shows the results for January precipitation based on predictor variables from November). The shaded area indicates the range of prediction values of all single tree models belonging to the random forest forecast model.

Values are converted to the monthly standardized precipitation index based on observations from the entire model calibration period. Obviously the variability of precipitation amounts is highly underestimated by the random forest based precipitation forecast models, which is a typical feature of regression based statistical models, particularly if the predictant variable is characterized by a large non-predictable noise. Furthermore, the correlation of forecasted and observed precipitation is low with values distinctly below 0.2 for most months and lead times. The rather poor results at the monthly scale certainly reflect the non-predictable noise of monthly precipitation amounts and thus lead to the assumption that

modelling results should not be evaluated based on discrete monthly values due to the high frequency variability of precipitation events. This is confirmed by the aggregation of observations and modelling results to three-month running totals, which leads to a significant increase of correlation and variance.. Fig. 4 shows the entire SPI time series for running three-month total precipitation amounts and the corresponding model results. (In order to generate a statistically independent forecast for the entire period, a four-fold split sample test has been conducted, see sec. 2.4 for details). Although the variability of precipitation amounts remains underestimated, the smoothed model results better capture the explicit features in terms of dry and moist periods of the observations. Taking into consideration the entire time series of three monthly precipitation amounts, the correlation between observed and forecasted values increases to r>0.5 for a lag time of 1 month. Correlations rapidly decrease with higher lead times, however, even for a lead time of 6 months a certain skill is detected (r=0.13).

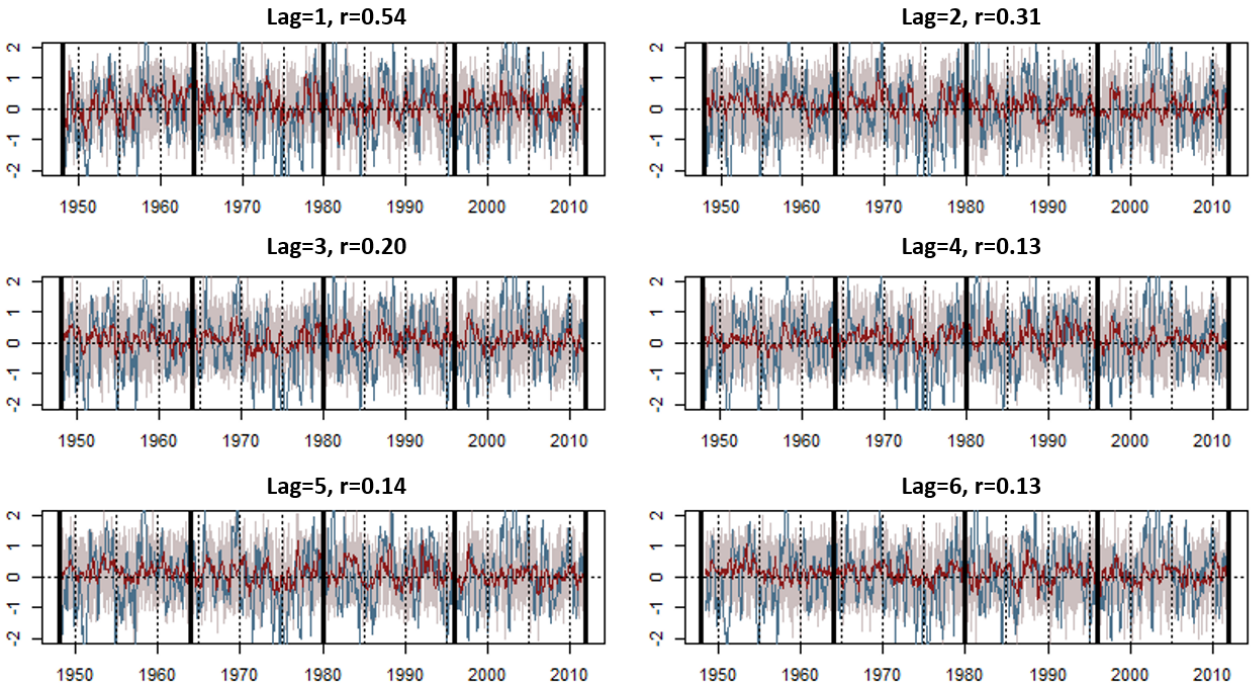


Fig. 4: Time series of observed (blue line) and forecasted (red line) running three-month SPI values for the Northern Central Asia. The shaded areas indicates the three month total of maximum and minimum forecasts of single trees of the random forest model. Black verticals indicate the division of the time series into four independent evaluation samples.

With this in mind we define two composite forecast periods with a length of three months respectively. The $F[1:3]^m$-forecast model is defined as the sum of random forest model results based on predictor variables from the month m with lead times of 1, 2 and 3 months. The $F[4:6]^m$-forecast is equally based on predictor variables from month m, but involves the random forest models with lead times of 4, 5 and 6 months.

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

where *RF(m,l)* is the specific Random Forest forecast model based on predictor variables of the month *m* and precipitation anomalies occurring after a lead time of *l* months. As an example, the $F[1:3]^{12}$-composite forecast including January, February and March is defined as the sum of three RF model results, which are all based on predictor variables from

previous December. RF(m=12,l=1) utilizes predictor variables from December for the January forecast, RF(m=12, l=2) indicates the December based forecast for February, RF(m=12,l=3) is the forecast march.

2.4 Model Evaluation

Since the skill of the automatic forecast model is likely to vary depending on the target area and the associated precipitation regimes during different seasons, an evaluation of the automatic seasonal forecast model performance is necessary in order to assess the reliability of the forecast and to interpret the results. Based on a 4-fold split sample test the deterministic forecasts of three-month running totals are automatically evaluated. Therefore, the entire time series from 1948 to 2014 is split into four sub-periods of equal length. The statistical forecast model is then applied four times, always taking one sub-period as an independent sample for the evaluation. The remaining three sub-periods are combined and split into two parts of equal length, which are utilized for the predictor selection and the model calibration, respectively. Eventually the independent predictions are compounded to one time series, comprising forecast values for the entire period. We abstained from the implementation of a full cross-validation procedure due to the high computational demands of the predictor selection routine. For each of the running three-month periods, traditional performance indicators such as correlation, bias and root mean square error (RMSE) are computed, which enables the assessment of the model performance for various seasons. In order to achieve a maximal comparability of different target areas, bias and RMSE are specified as the percentage of the long term precipitation totals for each three-month period respectively.
Moreover, since stakeholders often require robust predictions of anomalous periods, the ability of the forecast model to forecast drought and moist conditions is evaluated by means of receiver operating characteristics (ROC) for each three-month period and areas under the curve (AUC) are provided. Therefore, the running three-month precipitation totals are converted to the associated standardized precipitation indices, based on observations of the entire model calibration period. The deterministic SPI forecast is then converted into a probabilistic prediction by means of a simple residual based approach. Assuming that SPI residuals are normally distributed for each three-month period respectively, we estimate the standard deviation of residuals for each of the three-month periods, which is subsequently utilized to transform the deterministic forecast into a normalized probability distribution. ROC curves are then constructed for SPI threshold values of -0.5, representing moderate drought, and +0.5, indicating wet conditions. For various probability thresholds, positive hit rates (defined as the number of correctly identified droughts divided by the overall number of drought events) are plotted against the false negative rate (defined as the coefficient of the number of false alarms and the number on non-drought conditions). ROC curves for moist conditions are equivalently constructed. Eventually the area under the curve is interpreted as a performance measure of the seasonal forecast model. AUC-values near 1 indicate a perfect predictive skill considering the forecast of droughts or moist periods, values of 0.5 or less indicate no predictive skill at all.

3. Model Application to Central and South Asia

With regard to an increasing demand of climatological and hydrological forecasts in this vulnerable region, we applied the presented model to three target regions covering different climatic settings in Central and South Asia (see Fig. 4). The Northern Central Asian target area covers Uzbekistan, Kyrgyzstan and parts of Kazakhstan and comprises the majority of

the Syr Darya catchment. The Southern Central Asian target region covers wide parts of Iran, Afghanistan, Turkmenistan, Tajikistan and Pakistan and encompasses the Amu Darya river system. As presented by the mean 850 hPa wind field of the NCAR reanalysis (Fig. 5), both regions are mainly controlled by extratropical westerly circulation patterns (with contributions from South during winter and from high latitudes during summer) and receive a precipitation maximum during winter and spring season. Due to the location in continental Central Asia, both regions are characterized by a high precipitation variability with monthly coefficients of variation up to 0.5. In the high elevations of the Central Asian mountain ranges precipitation during moist season mainly falls as snow and is released during warm and dry summer season (Barlow and Tippett, 2008; Dixon and Wilby, 2015; Schär et al., 2004). Thus winter and spring precipitation amounts in the mountainous areas provide a vast share of the Central Asian river flow during the vegetation period and form the basis for the irrigation dependent agriculture of the riaparian countries, which are characterized by semi-arid to arid climate conditions throughout the year.

The Northern Indian domain covers the entire Himalayan range and the catchment of the Ganges river. During winter, the region is under influence of westerly winds and receives a certain amount of precipitation due to the passage of westerly disturbances, however the maximum of precipitation is associated with the Indian Summer Monsoon, which transports moist air masses from the Arabian Sea and the Bay of Bengal into the target area. Although, it is well documented, that particularly for Central Asia the number of stations utilized for the generation of gridded precipitation data is highly variable in time (Unger-Shayesteh et al., 2013), the Standard Normal Homogeneity Test does not detect any statistically significant shifts (a=0.05, see red line in Fig. 5) of the areal mean annual precipitation sums during the considered period in any of the target areas.

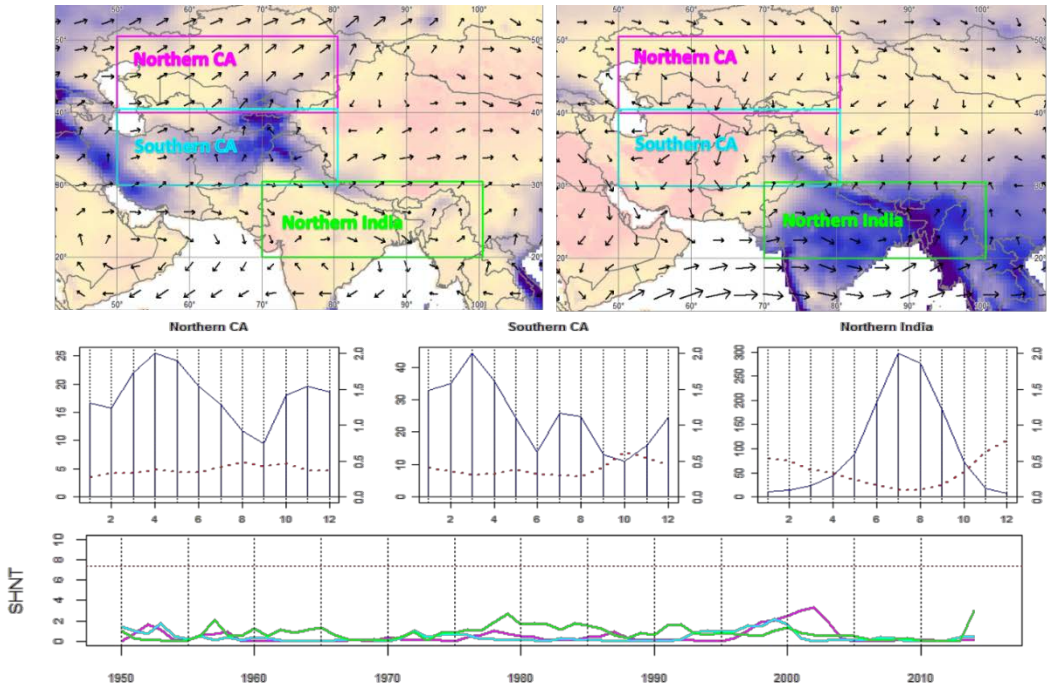


Fig. 5: Location of selected target areas as well as the mean precipitation total [mm] (CRU-TS) and the mean 850 hpa wind field (NCEP-NCAR) during DJF (left) and JJA (right).
Diagrams show the mean monthly precipitation amount for every catchment in mm (blue bars) as well as the Coefficient of Variation (red line) (middle panels) and the result of the Standardized Normal Homogeneity test. The red line indicate the 0.95-singnificance level (lower panel)

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

3.1 Pluviometric regimes and precipitation variability over Central and South Asia

In general the climate of Central and South Asia is influenced by two major pluviometric regimes, which are related to westerly and monsoonal circulation systems. During boreal cold season the entire region is influenced by westerly circulation patterns and precipitation is mainly associated with mid-latitude disturbances originating over the Atlantic Ocean and the Mediterranean (Bohner, 2006; Bothe et al., 2011; Gerlitz et al., 2015; Maussion et al., 2014). Since the track of westerly disturbances is mainly determined by the position of the 200 hPa westerly Jetstream at the polar frontal zone, a seasonal cycle of precipitation is distinctly defined. Particularly the Western parts of the Himalayas receive a considerable amount of winter precipitation associated with the uplift of westerly air masses, which reaches up to 60 % of the annual precipitation total (Bohner, 2006; Gerlitz et al., 2015; Wulf et al., 2010). During spring, the zone of westerly precipitation migrates towards north, reaches the Hindu Kush region and the Pamir in March and continues to the Tien Shan region in May. Mariotti (2007) showed that during winter season, a northward current over the Arabian countries transports tropical air masses into Central Asia, which represents an important moisture source for the westerly air masses. While the continental Central Asian countries remain under influence of extratropical westerly air masses throughout the year, the tropical monsoon circulation is established over South Asia during summer season (Bohner, 2006; Bookhagen and Burbank, 2006; Gerlitz et al., 2015). Due to a declining strength of the monsoonal moisture fluxes towards west, a clear gradient of precipitation totals from East to West has been detected (Bohner, 2006; Wulf et al., 2010).

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

In contrast, for the variability of winter and spring precipitation (associated with westerly weather patterns over Central and South Asia) a positive relationship with the El Nino Southern Oscillation has been observed. Severe droughts have been linked to the El Nino cold phase (La Nina) (Barlow et al., 2002, 2015; Hoell et al., 2013). Roghani et al. (2015) and Shirvani and Landman (2015) found statistically significant correlations of the Southern Oscillation index during summer

and autumn with precipitation amounts over Iran in subsequent winter. Likewise, a significant positive correlation of the ENSO state with winter precipitation amounts over the Southern Himalayan slopes has been detected (Dimri, 2013; Yadav et al., 2010). Mariotti (2007) showed that the moisture fluxes originating over the Arabian Sea are enhanced during ENSO warm phase due to the strengthening of the southwesterly current over the Arabian countries. Beside tropical SST modes, the impact of Northern Atlantic climatic conditions on the winter climate of Central Asia have been frequently investigated. Bothe et al. (2011) demonstrated that drought and moist winter seasons over Central Asia are dominated by different wave patterns over the Eurasian sector and particularly mention the North Atlantic Oscillation (NAO) and the East Atlantic pattern (EA) (which represent the first two modes of SLP variability in the North Atlantic domain) as important covariates. Schiemann et al. (2008) reported that an anomalous location or a decreasing strength of the westerly Jetstream result in drought conditions over parts of Central Asia due to modified tracks and intensities of westerly disturbances. Dimri (2013) found that a distinct southward shift of the westerly Jetstream is associated with wet winter conditions over the Himalayas. Syed et al. (2006, 2010) indicated that positive winter precipitation anomalies over Afghanistan, Pakistan and Tajikistan are usually associated with El Nino events combined with a positive state of the North Atlantic Oscillation. Negative correlations between the NAO index and observed precipitation anomalies were found for Kyrgyzstan and Northern Uzbekistan. Investigations by Bastos et al. (2016) indicated that both NAO and EA simultaneously control the winter moisture fluxes into Northern Central Asia. Maximum fluxes were found during negative NAO conditions, coupled with a positive EA index. Yin et al., (2014) further showed that the positive phase of the Eastern-Atlantic/Western-Russia and the Polar/Eurasian patterns lead to enhanced moisture fluxes into Central Asia. Most recently, Hartmann et al. (2016) suggested that beside of well-known atmospheric modes, the sea surface temperatures of the main moisture sources might influence the precipitation climate of the Tarim basin.

3.2 Modelling results

The seasonal precipitation model, including the automatic predictor selection routine, has been applied to each of the selected target regions and the results have been evaluated with regard to different seasons and the accompanying precipitation regimes. Fig. 6 shows the time series of observed three-month running totals (blue bars) and the composite results (red lines) for the F[1:3] forecast model. The evaluation results of the F[4:6] composite forecast model are presented in Fig. A1 (supplementary material). In order to keep the annual cycle, values are displayed at the center of each three-month period. The date of forecast generation is 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 three-month periods are presented and the 90% confidence interval of the residual based probabilistic forecast is illustrated. Fig. 7 summarizes the modelling results in terms of correlation, bias, RMSE and AUC for moderate drought and moist conditions. The performance measures are provided for each of the running three-month periods, respectively.

For the North Central Asian domain, drought and moist conditions during winter and spring, which are characterized by maximum moisture fluxes into the target region, are well captured by the statistical model.For example the recent moist spring seasons in 2005 and 2010 are adequately predicted by the F[1:3] forecast model. Also the spring drought of 2008 and particularly the prolonged drought of 1999-2001 are accurately predicted by the forecast model, although the severity of the extreme 1999-2001 drought is highly underestimated. Correlations between observed and modelled precipitation totals are high ($r>0.4$) for winter and spring. AUC values >0.7 indicate that the model is capable to forecast moderate drought and moist conditions in Northern Central Asia during winter and spring. RMSE is in the order of 20% of the precipitation mean. For the dry summer season the skill of the forecast model is distinctly lower with correlation around 0.2 and AUC values in the order of 0.5 for both, moderate drought and moist seasons. RMSE values in summer reach up

to 40% of the mean precipitation amounts. The SPI time series for Southern Central Asia shows a similar variability and is significantly correlated with the North Central Asian record, which indicates a common large-scale climatic forcing of the Central Asian target areas. For example. the recent drought conditions during boreal cold seasons of 2007/2008 and 1999-2001 are evident in both the observational and the modelled time series. However the variability of precipitation rates in Southern Central Asia is highly underestimated by the statistical model. Correlations reach highest values in late autumn ($r>0.4$), but some three-month composite periods with correlation below 0.2 were detected throughout the year. AUC values exceed 0.7 in autumn, winter and spring, during dry summer season the evaluation results are highly heterogeneous, with some AUC values in the order of 0.5, indicating a limited skill of the precipitation forecast model.

For the monsoonal influenced target domain, maximum correlations were achieved during summer season. Particularly for the late monsoon season, high correlations ($r>0.4$) and AUC values above 0.7 were detected for the F[1:3] forecast, which indicates the ability of the model to predict monsoonal drought periods several months in advance. For example the negative precipitation anomaly during summer monsoon of 2009 (which was the second worst drought of the entire period) is well captured by the forecast model, however the magnitude of extreme events is mostly underestimated. For the winter and transition seasons, negative correlations, high RMSE values of up to 60% of the long-term mean and AUC values below 0.5 indicate a poor performance of the statistical model. In overall, the statistical model adequately captures the variability of westerly precipitation amounts for the Central Asian target domains, particularly during moist winter and spring seasons. For the Northern Indian region, the evaluation measures reach highest values during summer season, when precipitation is associated with monsoonal circulation modes. During winter and the transition seasons associated with westerly weather patterns over Northern India, the model fails to reproduce the inter-annual precipitation variability.

The F[4:6] composite forecast model in general shows a distinctly lower skill compared with F[1:3] (see supplement Figures A1 and A2). Correlations remain positive for the Northern Central Asian domain for the moist cold season, however, values seldom exceed $r=0.2$ in the F[4:6] model. AUC values during moist seasons are in the order of 0.6 for both Central Asian domains, which indicates a low but still positive skill of the F[4:6] forecast model. The skill of the F[4:6] model for the monsoonal Northern Indian target area is in general low with negative correlations and AUC values <0.5 in most of the months.

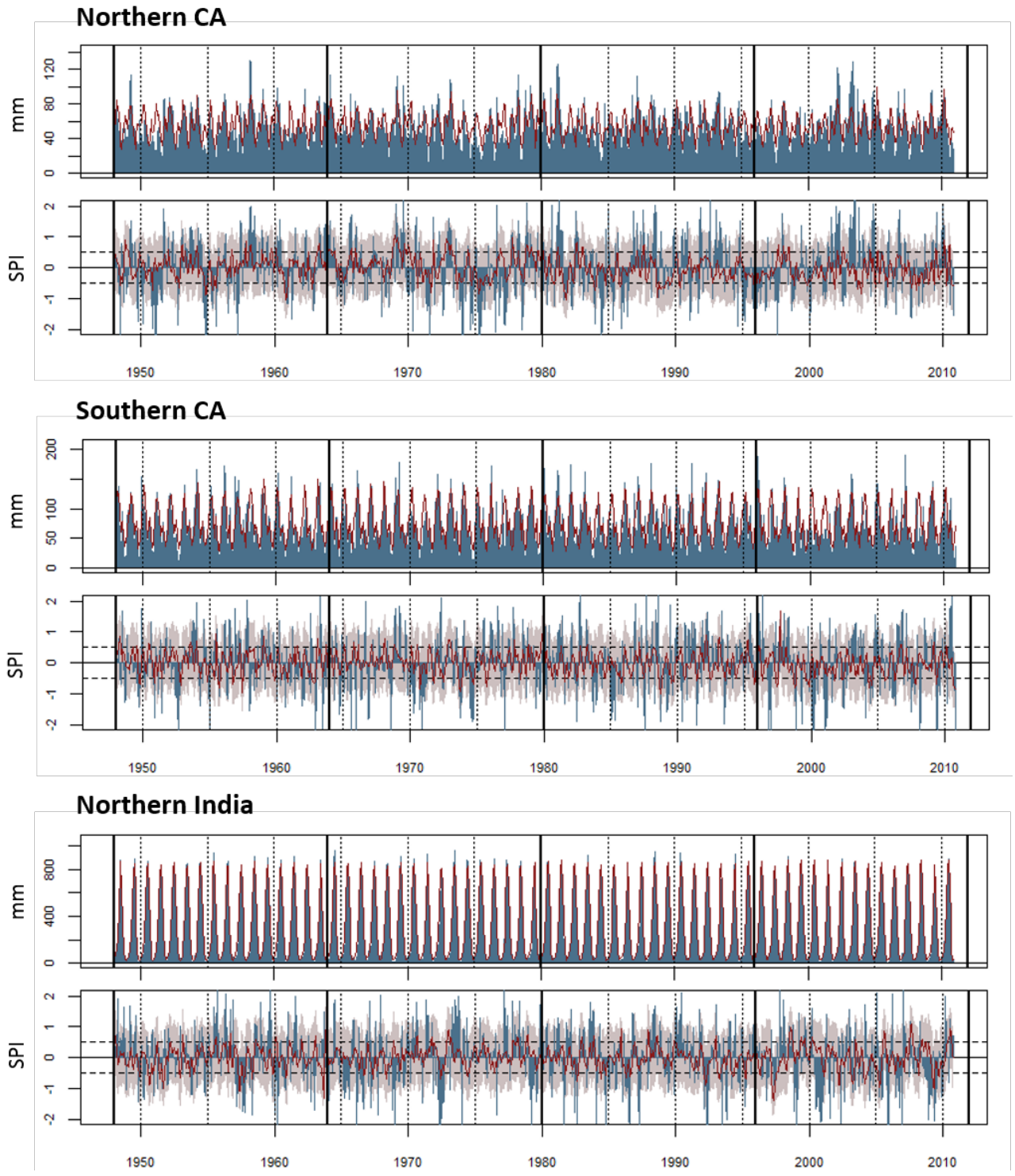


Fig. 6: Observed running three-month precipitation totals (blue bars) and modelling results (red line) of the F[1:3] model for selected target regions. The upper panels show absolute precipitation totals for running three-month periods, the lower panels show the corresponding SPI index for each three-month period respectively. Shaded areas indicate the 90% interval of the residual based probabilistic forecast. Black verticals indicate the division of the time series into four independent evaluation samples.

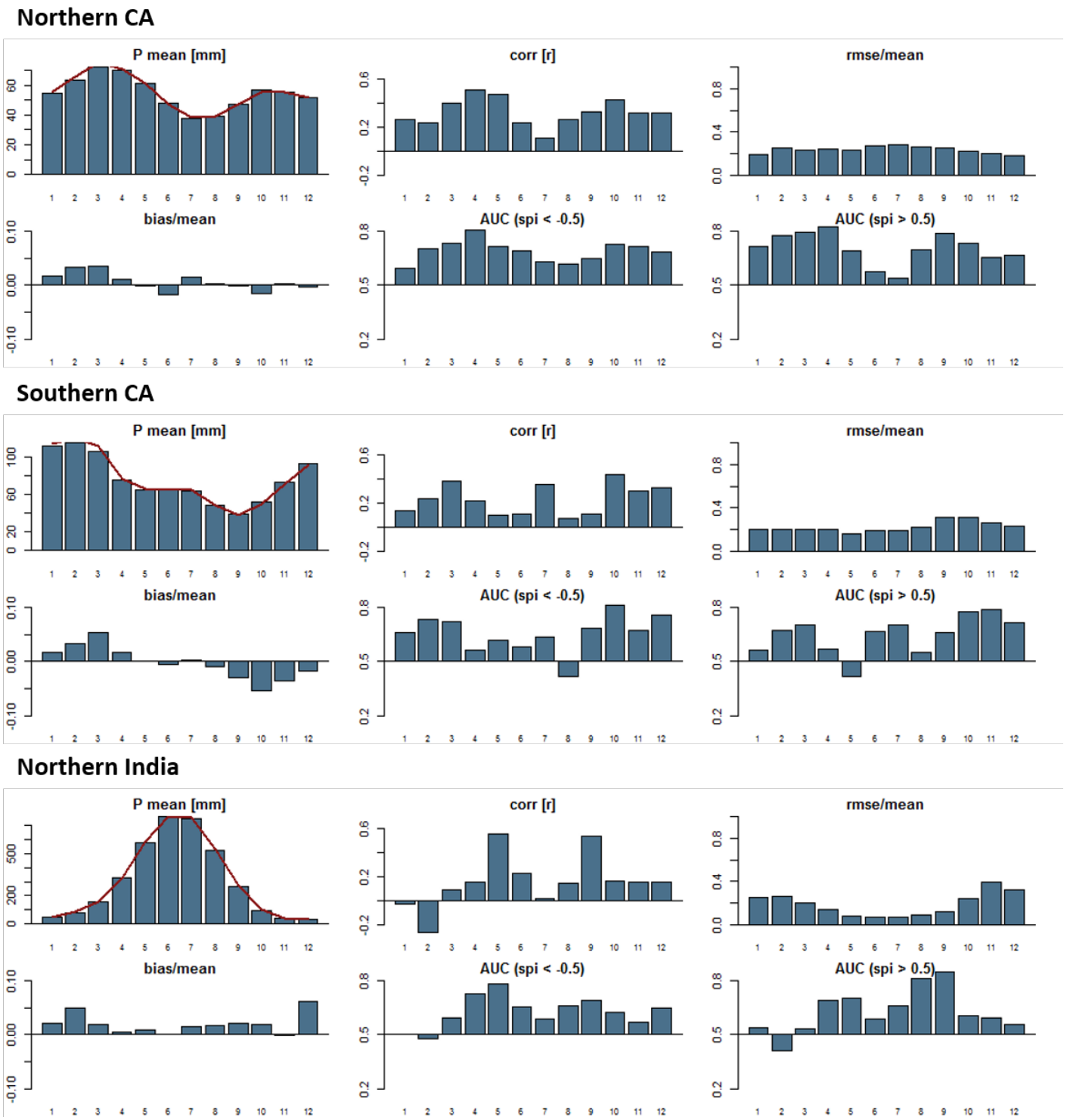


Fig. 7: Summary of evaluation measures of the F[1:3] forecast for selected target areas. In order to keep the annual cycle of precipitation amounts, the specified month at the x-axis indicate the middle of the forecast period.

3.3 Sensitivity Analysis

In comparison to linear models with a small set of independent predictor variables, the complex structure of the presented random forest based model does not directly reveal physically interpretable input–output relationships. Particularly the fact that the predictor selection procedure generates a large sample of partially highly correlated predictor variables, which basically comprise the same information concerning the large scale climatic variability, impedes a direct interpretation of the predictor importance and variable response. Frequently utilized random forest variable importance measures are based on the increase of the model error, in case of a random modification of one particular variable (permutation importance). If the predictor space is not statistically independent, i.e. it includes highly correlated predictor variables, every variable can easily be substituted, which results in unrealistically low values of the random forest importance measure (Gregorutti et al., 2016).

In order to overcome the blackbox character of the statistical model, we conducted a sensitivity analysis for the selected target areas under consideration of well-known atmospheric indices. Therefore individual random forest models were forced with modified input data, containing only those predictor variables, which are highly correlated with the considered indices. This facilitates the estimation of the fractional response of the model to a considered predictor and reveals the underlying influence of major atmospheric modes on hydro-climatic variability of the target regions. The results of the sensitivity analysis enable a comparison of the model results with previous studies, which utilized traditional climate indices (see section 3.1 for a brief summary) and thus serve as a plausibility test of the presented approach.

With the aim of investigating the model response to a selected climate index, the time series of potential predictor variables, which are significantly correlated with the index (alpha=0.01) are maintained, while the others were set to zero. All maintained predictor variables (which are associated with the considered large scale atmospheric mode) are modified to an equal distance record of values ranging from -2 to 2 standard deviations, if the predictor is positively correlated with the considered climate index (if the correlation is negative, modified values range from 2 to -2). The statistical forecast model is then applied to the modified predictor-data. The results are converted to SPI values and indicate the response of the model to increasing values of the considered large scale climate index. Fig. 8 shows the results of the sensitivity analysis for December, March, June and September, as representative for winter, spring, summer and autumn season, respectively. Since the sensitivity procedure is only valid for individual random forest models, we analyzed the monthly forecast models with different lead times, in order to estimate the influence of selected climate indices. A direct sensitivity study for the F[1:3] composite forecast model is not feasible, due to its complex aggregation of various random forest models, however the results can be regarded as generally valid, if the sensitivity is constant for varying lead times.

However, due to the non-linear nature of the statistical model, the response fractions should not be perceived as independent or additive and should rather be interpreted as a general sensitivity of the model.

As potentially important large scale climate indices we make use of the El Nino-3 index (ENSO) as well as the North Atlantic Oscillation (NOA) and the East Atlantic pattern (EA), which are frequently mentioned as important influencing factors on the Central and South Asian precipitation climate (Barlow et al., 2002; Hoell et al., 2013; Khidher and Pilesjö, 2014; Syed et al., 2006).

The plotted predictor responses (Fig. 8) clearly indicate that the state of the El Nino Southern Oscillation determines the precipitation variability in all target areas. For winter and spring season (represented by the forecast models for December and March) a positive response to predictors related to the ENSO-3 index is evident for all target areas, indicating an intensification of moisture fluxes and associated westerly disturbances over the entire domain during ENSO warm phase and a reversed effect during cold phase of El Nino, which is consistent with previous studies on the variability of cold

season precipitation totals in the vast target domain (Barlow et al., 2002, 2015; Dimri, 2013; Mariotti, 2007; Syed et al., 2006). The model response is strongest for the moist seasons, which is late winter for the Southern Central Asia and spring for the Northern Central Asian domain. This coincides with a moderate to high model performance for the Central Asian target regions during winter and spring season and emphasizes the relevance of ENSO for the winter and spring moisture fluxes into Central Asia. As indicated in Fig. 6 winter and spring drought conditions frequently occur simultaneously in Northern and Southern Central Asia, which indicates a common large scale forcing, most likely associated with an ENSO cold phase. The positive response to increasing values of the ENSO index remains constant for lead times ranging from 0 to 3 months, indicating a high forecast potential for the F[1:3] model in Central Asia. Winter drought conditions in the Northern Indian domain are likewise associated with La Nina events.

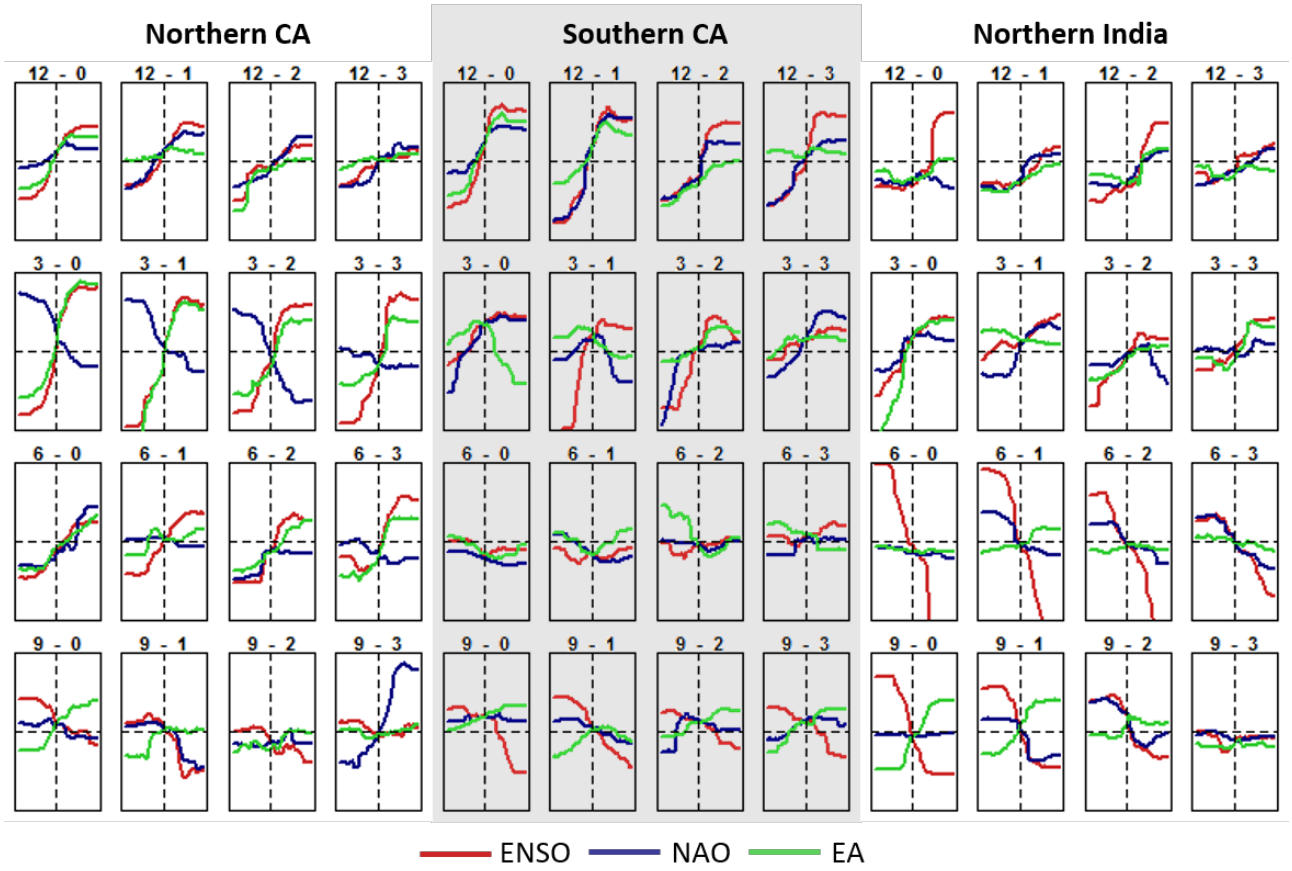


Fig. 8: Results of sensitivity analysis for selected target regions, months and lead times (from 0 to 3 months). The x-axis represents the range of the considered large scale index (from -2 to 2 standard deviations). The y-axis indicates the model response to associated predictor variables (ranging from SPI=-1 to SPI=1)

During summer, the response of the model results to variations of ENSO remains positive for Northern Central Asia, however the response magnitude is distinctly lower. Although the model performance for the Central Asian target regions is poor during summer season, the results of the sensitivity study are mostly consistent with findings by Mariotti (2007), who proposed seasonal independent enhanced South Easterly moisture fluxes into Central Asia during the ENSO warm phase.

For the monsoonal influenced North Indian target area, a distinct negative relationship between ENSO variations and summer and autumn precipitation is evident in the sensitivity results, which confirms a number of previous studies

(Rajeevan and Pai, 2007; Sigdel and Ikeda, 2013; Wu et al., 2009). In autumn, a slight negative response has also been detected for Southern Central Asia, indicating a monsoonal influence, which is certainly prevalent in Pakistan.

In addition, the winter and spring precipitation forecast models for Northern and Southern Central Asia distinctly respond to variations of predictor variables related to the North Atlantic Oscillation and the East Atlantic pattern, which reveals the influence of pressure anomalies in the temperate climate zones on the Central Asian precipitation variability. In December the model positively responds to increasing NAO and EA indices. Particularly for the South Central Asian target area, the magnitude is in the order of the response to ENSO related predictor variables for lead times of 0 and 1 month (the zero forecast is based on mean predictor variables from the same month and has not been considered in the forecast procedure). For larger lead times the response magnitude for NAO and EA decreases, which indicates a lower forecast potential. The positive response to the EA pattern is likewise evident in March for the Northern Central Asian target area. However, for the NAO, a strong negative response has been detected for lead times up to 2 months. The response to NAO in Southern Central Asia remains positive in March. Again this is confirmed by previous studies of Syed et al. (2006) who found a negative correlation between the state of the NAO and cold season precipitation over the Northern Central Asian countries and a reverse relationship for a band covering Iran, southern Afghanistan and Pakistan. Thus, the combination of a negative NAO and EA phase with a warm phase of the El Nino Southern Oscillation is likely to trigger drought conditions over Southern Central Asia during winter season. During spring and particularly for the Northern Central Asian region, drought conditions are associated with an El Nino cold phase in combination with a negative state of the EA pattern and positive state of the NAO. While the response to ENSO and EA is similar in both Central Asian target regions, the differentiated response to NAO leads to a diverging precipitation signal in Northern and Southern Central Asia.

4. Summary and Outlook

We presented a statistically based modelling framework, which automatically identifies suitable predictors from globally gridded climate variables by means of an extensive data mining procedure and explicitly avoids the utilization of typical large-scale climate indices. This leads to an enhanced flexibility of the model and enables its automatic calibration for any target area without any prior assumption concerning adequate predictor variables. Potential predictor variables are derived by means of a cell-wise correlation analysis of precipitation anomalies within a user selectable target area with global climate variables. The correlation analysis is conducted for monthly values with lead times ranging from one to six months. For each potential predictor variable, month and lead time, significantly correlated grid cells are aggregated to predictor regions by means of a variability based cluster analysis. Finally, for every month and lead time, an individual random forest based forecast model is constructed, by means of the preliminary generated predictor variables. In order to reduce the risk of overfitting, predictor selection and model calibration are based on independent samples. Due to the large noise of observed precipitation amounts at a monthly time scale, the random forest based forecasts based on predictor variables of one specific month withlead times of one to three months and four to six months are aggregated to running three-month composite predictions. These are automatically evaluated based on a 4-fold split sample test and modelling performance measures are provided for each of the running three-month predictions, which enables the assessment of the model performance for different seasons of the year.

The model has been applied to selected target regions in Central and South Asia. While the Central Asian catchments are primarily under influence of westerly air masses throughout the year, the target area in Southern Asia receives moisture fluxes from westerly winds during winter and is under influence of the South Asian Monsoon during summer season.

Particularly for the Central Asian target domains correlations between observations and forecast results reach values r>0.4, especially for the moist winter and spring seasons. The capability of the model to predict moderate drought events or anomalous moist conditions is reflected by AUC values >0.7. Due to the fact that precipitation in the high elevations mainly falls as snow and is released during dry summer season, the irrigated agriculture of the downstream countries is highly vulnerable to drought events during winter and spring. Some studies indicate, that the natural summer discharge of the tributaries of the major Central Asian rivers can be accurately forecasted by means of winter precipitation amounts or snow cover rates, which are usually available in spring (Barlow and Tippett, 2008; Dixon and Wilby, 2015). A modelling chain including statistical precipitation forecasting and runoff prediction could extend the forecast range and foster adequate adaption strategies.

For the Northern Indian target area, the model performance was found to be slightly lower, but particularly for the economically important monsoonal precipitation amounts correlation values reach 0.4 and higher, and AUC values exceed 0.7.

A sensitivity analysis of the complex statistical model using well-known climate indices shows, that the model automatically identifies relevant predictor variables, among others, those that are associated with typical climatic modes, such as the El Nino Southern Oscillation, North Atlantic Oscillation and the East Atlantic pattern. Further, the sensitivity analysis enables the estimation of the model response to specified climatic modes and thus reveals the major influencing factors for the observed precipitation variability. The winter and spring precipitation amounts in the entire target area were found to be highly influenced by the state of the El Nino Southern Oscillation with positive precipitation anomalies during El Nino events. Additionally for the Central Asian catchments, the states of the North Atlantic Oscillation and the East Atlantic pattern were identified as important controlling factors. The sensitivity analysis of the model suggests that drought events in Northern Central Asia are frequently triggered by an ENSO cold phase in combination with a positive NAO and a negative EA state. Drought in the Southern Central Asian domain is associated with an El Nino cold phase in combination with negative NAO and EA indices. Concerning the forecast of summer precipitation amounts in the monsoonal Northern Indian domain the model shows a distinct negative response to El Nino events.

In general, the statistical model is characterized by a large underestimation of variance, but the forecast of a drought risk appears feasible to a certain extent. The accurate prediction of severe drought periods, however, remains difficult by means of statistical techniques. Therefore, the atmospheric and oceanic patterns, which trigger extreme drought or moist conditions and the interaction of potential influencing factors, such as the state of the North Atlantic Oscillation, the East Atlantic pattern or the El Nino Southern, need to be further investigated. Additionally, since climatic conditions in the selected target areas show a large noise, which is not predictable by means large scale atmospheric and oceanic predictor variables, the implementation of a real probabilistic forecast model should be considered for further model development. The generation of a model ensemble based on randomly selected predictor variables and a subsequent model averaging approach, for example based on Beyesian techniques as proposed by Wang et al. (2012), appears promising in this regard.

Acknowledgements

This work was carried out within the framework of the CAWa (Water in Central Asia) project (http://www.cawa-project.net, contract no. AA7090002), funded by the German Federal Foreign Office as part of the “Berlin Process”. We further thank M. Barlow and another anonymous reviewer as well as the HESS editor Q.J. Wang, who helped to considerably improve the presented manuscript.

References

Ashok, K., Guan, Z. and Yamagata, T.: Impact of the Indian Ocean dipole on the relationship between the Indian monsoon rainfall and ENSO, Geophys. Res. Lett., 28(23), 4499–4502, doi:10.1029/2001GL013294, 2001.

Barlow, M., Cullen, H. and Lyon, B.: Drought in Central and Southwest Asia: La Niña, the Warm Pool, and Indian Ocean Precipitation, J. Climate, 15(7), 697–700, doi:10.1175/1520-0442(2002)015<0697:DICASA>2.0.CO;2, 2002.

Barlow, M., Zaitchik, B., Paz, S., Black, E., Evans, J. and Hoell, A.: A Review of Drought in the Middle East and Southwest Asia, J. Climate, doi:10.1175/JCLI-D-13-00692.1, 2015.

Barlow, M. A. and Tippett, M. K.: Variability and Predictability of Central Asia River Flows: Antecedent Winter Precipitation and Large-Scale Teleconnections, J. Hydrometeor, 9(6), 1334–1349, doi:10.1175/2008JHM976.1, 2008.

Bastos, A., Janssens, I. A., Gouveia, C. M., Trigo, R. M., Ciais, P., Chevallier, F., Peñuelas, J., Rödenbeck, C., Piao, S., Friedlingstein, P. and Running, S. W.: European land CO2 sink influenced by NAO and East-Atlantic Pattern coupling, Nat Commun, 7, 10315, doi:10.1038/ncomms10315, 2016.

Bohner, J.: General climatic controls and topoclimatic variations in Central and High Asia, Boreas, 35(2), 279–295, doi:10.1111/j.1502-3885.2006.tb01158.x, 2006.

Bookhagen, B. and Burbank, D. W.: Topography, relief, and TRMM-derived rainfall variations along the Himalaya, Geophys. Res. Lett., 33(8), L08405, doi:10.1029/2006GL026037, 2006.

Bothe, O., Fraedrich, K. and Zhu, X.: Precipitation climate of Central Asia and the large-scale atmospheric circulation, Theor Appl Climatol, 108(3–4), 345–354, doi:10.1007/s00704-011-0537-2, 2011.

Brands, S., Manzanas, R., Gutiérrez, J. M. and Cohen, J.: Seasonal Predictability of Wintertime Precipitation in Europe Using the Snow Advance Index, J. Climate, 25(12), 4023–4028, doi:10.1175/JCLI-D-12-00083.1, 2012.

Breiman, L.: Random Forests, Machine Learning, 45(1), 5–32, doi:10.1023/A:1010933404324, 2001.

Breiman, L., Friedman, J., Stone, C. J. and Olshen, R. A.: Classification and Regression Trees, Taylor & Francis., 1984.

Cai, W., van Rensch, P., Cowan, T. and Hendon, H. H.: Teleconnection Pathways of ENSO and the IOD and the Mechanisms for Impacts on Australian Rainfall, J. Climate, 24(15), 3910–3923, doi:10.1175/2011JCLI4129.1, 2011.

Chang, C.-P., Harr, P. and Ju, J.: Possible Roles of Atlantic Circulations on the Weakening Indian Monsoon Rainfall–ENSO Relationship, J. Climate, 14(11), 2376–2380, doi:10.1175/1520-0442(2001)014<2376:PROACO>2.0.CO;2, 2001.

Chen, J., Li, M., Wang, W., Chen, J., Li, M. and Wang, W.: Statistical Uncertainty Estimation Using Random Forests and Its Application to Drought Forecast, Statistical Uncertainty Estimation Using Random Forests and Its Application to Drought Forecast, Mathematical Problems in Engineering, Mathematical Problems in Engineering, 2012, 2012, e915053, doi:10.1155/2012/915053, 10.1155/2012/915053, 2012.

Chiew, F. H. S., Zhou, S. L. and McMahon, T. A.: Use of seasonal streamflow forecasts in water resources management, Journal of Hydrology, 270(1–2), 135–144, doi:10.1016/S0022-1694(02)00292-5, 2003.

Cohen, J. and Barlow, M.: The NAO, the AO, and Global Warming: How Closely Related?, J. Climate, 18(21), 4498–4513, doi:10.1175/JCLI3530.1, 2005.

Cohen, J. and Entekhabi, D.: Eurasian snow cover variability and northern hemisphere climate predictability, Geophys. Res. Lett., 26(3), 345–348, doi:10.1029/1998GL900321, 1999.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V. and Böhner, J.: System for Automated Geoscientific Analyses (SAGA) v. 2.1.4, Geosci. Model Dev., 8(7), 1991–2007, doi:10.5194/gmd-8-1991-2015, 2015.

Dai, A. and Wigley, T. M. L.: Global patterns of ENSO-induced precipitation, Geophys. Res. Lett., 27(9), 1283–1286, doi:10.1029/1999GL011140, 2000.

Dimri, A. P.: Interannual variability of Indian winter monsoon over the Western Himalayas, Global and Planetary Change, 106, 39–50, doi:10.1016/j.gloplacha.2013.03.002, 2013.

Dixon, S. G. and Wilby, R. L.: Forecasting reservoir inflows using remotely sensed precipitation estimates: a pilot study for the River Naryn, Kyrgyzstan, Hydrological Sciences Journal, 0(0), 1–16, doi:10.1080/02626667.2015.1006227, 2015.

Douville, H. and Chauvin, F.: Relevance of soil moisture for seasonal climate predictions: a preliminary study, Climate Dynamics, 16(10–11), 719–736, doi:10.1007/s003820000080, 2000.

Eden, J. M., van Oldenborgh, G. J., Hawkins, E. and Suckling, E. B.: A global empirical system for probabilistic seasonal climate prediction, Geosci. Model Dev., 8(12), 3947–3973, doi:10.5194/gmd-8-3947-2015, 2015.

Fraedrich, K.: An ENSO impact on Europe?, Tellus A, 46(4), 541–552, doi:10.1034/j.1600-0870.1994.00015.x, 1994.

Gerlitz, L.: Using fuzzified regression trees for statistical downscaling and regionalization of near surface temperatures in complex terrain, Theor Appl Climatol, 122(1–2), 337–352, doi:10.1007/s00704-014-1285-x, 2014.

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

Gregorutti, B., Michel, B. and Saint-Pierre, P.: Correlation and variable importance in random forests, Statistics and Computing, doi:10.1007/s11222-016-9646-1, 2016.

Guttman, N. B.: Comparing the Palmer Drought Index and the Standardized Precipitation Index1, JAWRA Journal of the American Water Resources Association, 34(1), 113–121, doi:10.1111/j.1752-1688.1998.tb05964.x, 1998.

Harris, I., Jones, P. d., Osborn, T. j. and Lister, D. h.: Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, Int. J. Climatol., 34(3), 623–642, doi:10.1002/joc.3711, 2014.

Hartmann, H., Snow, J. A., Stein, S., Su, B., Zhai, J., Jiang, T., Krysanova, V. and Kundzewicz, Z. W.: Predictors of precipitation for improved water resources management in the Tarim River basin: Creating a seasonal forecast model, Journal of Arid Environments, 125, 31–42, doi:10.1016/j.jaridenv.2015.09.010, 2016.

Hasson, S., Lucarini, V., Khan, M. R., Petitta, M., Bolch, T. and Gioli, G.: Early 21st century snow cover state over the western river basins of the Indus River system, Hydrol. Earth Syst. Sci., 18(10), 4077–4100, doi:10.5194/hess-18-4077-2014, 2014.

Hertig, E. and Jacobeit, J.: Predictability of Mediterranean climate variables from oceanic variability. Part II: Statistical models for monthly precipitation and temperature in the Mediterranean area, Clim Dyn, 36(5–6), 825–843, doi:10.1007/s00382-010-0821-3, 2010.

Hoell, A., Funk, C. and Barlow, M.: The regional forcing of Northern hemisphere drought during recent warm tropical west Pacific Ocean La Niña events, Clim Dyn, 42(11–12), 3289–3311, doi:10.1007/s00382-013-1799-4, 2013.

Hoerling, M., Eischeid, J. and Perlwitz, J.: Regional Precipitation Trends: Distinguishing Natural Variability from Anthropogenic Forcing, J. Climate, 23(8), 2131–2145, doi:10.1175/2009JCLI3420.1, 2010.

Hurk, B. van den, Doblas-Reyes, F., Balsamo, G., Koster, R. D., Seneviratne, S. I. and Jr, H. C.: Soil moisture effects on seasonal temperature and precipitation forecast scores in Europe, Clim Dyn, 38(1–2), 349–362, doi:10.1007/s00382-010-0956-2, 2010.

Julian, P. R. and Chervin, R. M.: A Study of the Southern Oscillation and Walker Circulation Phenomenon, Mon. Wea. Rev., 106(10), 1433–1451, doi:10.1175/1520-0493(1978)106<1433:ASOTSO>2.0.CO;2, 1978.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R. and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, Bull. Amer. Meteor. Soc., 77(3), 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

Khidher, S. A. and Pilesjö, P.: The effect of the North Atlantic Oscillation on the Iraqi climate 1982–2000, Theor Appl Climatol, 122(3–4), 771–782, doi:10.1007/s00704-014-1327-4, 2014.

Krishnaswamy, J., Vaidyanathan, S., Rajagopalan, B., Bonell, M., Sankaran, M., Bhalla, R. S. and Badiger, S.: Non-stationary and non-linear influence of ENSO and Indian Ocean Dipole on the variability of Indian monsoon rainfall and extreme rain events, Clim Dyn, 45(1–2), 175–184, doi:10.1007/s00382-014-2288-0, 2014.

Kumar, A., Chen, M. and Wang, W.: Understanding Prediction Skill of Seasonal Mean Precipitation over the Tropics, J. Climate, 26(15), 5674–5681, doi:10.1175/JCLI-D-12-00731.1, 2013.

Kumar, K. K., Rajagopalan, B. and Cane, M. A.: On the Weakening Relationship Between the Indian Monsoon and ENSO, Science, 284(5423), 2156–2159, doi:10.1126/science.284.5423.2156, 1999.

Lau, K.-M. and Wu, H. T.: Principal Modes of Rainfall–SST Variability of the Asian Summer Monsoon: A Reassessment of the Monsoon–ENSO Relationship, J. Climate, 14(13), 2880–2895, doi:10.1175/1520-0442(2001)014<2880:PMORSV>2.0.CO;2, 2001.

Li, C. and Yanai, M.: The Onset and Interannual Variability of the Asian Summer Monsoon in Relation to Land–Sea Thermal Contrast, J. Climate, 9(2), 358–375, doi:10.1175/1520-0442(1996)009<0358:TOAIVO>2.0.CO;2, 1996.

Liebmann, B., Hoerling, M. P., Funk, C., Bladé, I., Dole, R. M., Allured, D., Quan, X., Pegion, P. and Eischeid, J. K.: Understanding Recent Eastern Horn of Africa Rainfall Variability and Change, J. Climate, 27(23), 8630–8645, doi:10.1175/JCLI-D-13-00714.1, 2014.

Mariotti, A.: How ENSO impacts precipitation in southwest central Asia, Geophys. Res. Lett., 34(16), L16706, doi:10.1029/2007GL030078, 2007.

Mason, S. J. and Goddard, L.: Probabilistic Precipitation Anomalies Associated with ENSO, Bull. Amer. Meteor. Soc., 82(4), 619–638, doi:10.1175/1520-0477(2001)082<0619:PPAAWE>2.3.CO;2, 2001.

Maussion, F., Scherer, D., Mölg, T., Collier, E., Curio, J. and Finkelnburg, R.: Precipitation Seasonality and Variability over the Tibetan Plateau as Resolved by the High Asia Reanalysis*, J. Climate, 27(5), 1910–1927, doi:10.1175/JCLI-D-13-00282.1, 2014.

McKee, T. B., Doesken, N. J. and Kleist, J.: The Relationship of Drought Frequency and Duration to Time Scales, [online] Available from: http://ccc.atmos.colostate.edu/relationshipofdroughtfrequency.pdf, 1993.

New, M., Hulme, M. and Jones, P.: Representing Twentieth-Century Space–Time Climate Variability. Part I: Development of a 1961–90 Mean Monthly Terrestrial Climatology, J. Climate, 12(3), 829–856, doi:10.1175/1520-0442(1999)012<0829:RTCSTC>2.0.CO;2, 1999.

Orsolini, Y. J., Senan, R., Balsamo, G., Doblas-Reyes, F. J., Vitart, F., Weisheimer, A., Carrasco, A. and Benestad, R. E.: Impact of snow initialization on sub-seasonal forecasts, Clim Dyn, 41(7–8), 1969–1982, doi:10.1007/s00382-013-1782-0, 2013.

Palmer, T. N. and Anderson, D. L. T.: The prospects for seasonal forecasting—A review paper, Q.J.R. Meteorol. Soc., 120(518), 755–793, doi:10.1002/qj.49712051802, 1994.

Parhi, P., Giannini, A., Gentine, P. and Lall, U.: Resolving contrasting regional rainfall responses to El Niño over tropical Africa, J. Climate, doi:10.1175/JCLI-D-15-0071.1, 2015.

Peings, Y. and Douville, H.: Influence of the Eurasian snow cover on the Indian summer monsoon variability in observed climatologies and CMIP3 simulations, Clim Dyn, 34(5), 643–660, doi:10.1007/s00382-009-0565-0, 2009.

Pokhrel, S., Chaudhari, H. S., Saha, S. K., Dhakate, A., Yadav, R. K., Salunke, K., Mahapatra, S. and Rao, S. A.: ENSO, IOD and Indian Summer Monsoon in NCEP climate forecast system, Clim Dyn, 39(9–10), 2143–2165, doi:10.1007/s00382-012-1349-5, 2012.

Prodhomme, C., Terray, P., Masson, S., Boschat, G. and Izumo, T.: Oceanic factors controlling the Indian summer monsoon onset in a coupled model, Clim Dyn, 44(3–4), 977–1002, doi:10.1007/s00382-014-2200-y, 2014.

R Development Core Team: R: The R Project for Statistical Computing, R Foundation for Statistical Computing, [online] Available from: https://www.r-project.org/ (Accessed 17 December 2015), 2008.

Rajeevan, M. and Pai, D. S.: On the El Niño-Indian monsoon predictive relationships, Geophys. Res. Lett., 34(4), L04704, doi:10.1029/2006GL028916, 2007.

Rajeevan, M., Pai, D. S., Kumar, R. A. and Lal, B.: New statistical models for long-range forecasting of southwest monsoon rainfall over India, Clim Dyn, 28(7–8), 813–828, doi:10.1007/s00382-006-0197-6, 2006.

Ratnam, J. V., Behera, S. K., Masumoto, Y. and Yamagata, T.: Remote Effects of El Niño and Modoki Events on the Austral Summer Precipitation of Southern Africa, J. Climate, 27(10), 3802–3815, doi:10.1175/JCLI-D-13-00431.1, 2014.

Roghani, R., Soltani, S. and Bashari, H.: Influence of southern oscillation on autumn rainfall in Iran (1951–2011), Theor Appl Climatol, 1–13, doi:10.1007/s00704-015-1423-0, 2015.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M. P., van den Dool, H., Zhang, Q., Wang, W., Chen, M. and Becker, E.: The NCEP Climate Forecast System Version 2, J. Climate, 27(6), 2185–2208, doi:10.1175/JCLI-D-12-00823.1, 2014.

Schär, C., Vasilina, L., Pertziger, F. and Dirren, S.: Seasonal Runoff Forecasting Using Precipitation from Meteorological Data Assimilation Systems, J. Hydrometeor, 5(5), 959–973, doi:10.1175/1525-7541(2004)005<0959:SRFUPF>2.0.CO;2, 2004.

Schepen, A., Wang Q.J. and Robertson, D.: Evidence for Using Lagged Climate Indices to Forecast Australian Seasonal Rainfall, J. Climate, 35, 1230-1246, doi:10.1175/JCLI-D-11-00156.1, 2011.

Schiemann, R., Lüthi, D., Vidale, P. L. and Schär, C.: The precipitation climate of Central Asia—intercomparison of observational and numerical data sources in a remote semiarid region, Int. J. Climatol., 28(3), 295–314, doi:10.1002/joc.1532, 2008.

Seibert, M., Apel, H. and Merz, B.: Seasonal forecasting of hydrological drought in the Limpopo basis: A comparison of statistical methods, 2016.

Shirvani, A. and Landman, W. A.: Seasonal precipitation forecast skill over Iran, Int. J. Climatol., n/a-n/a, doi:10.1002/joc.4467, 2015.

Sigdel, M. and Ikeda, M.: Summer Monsoon Rainfall over Nepal Related with Large-Scale Atmospheric Circulations, Journal of Earth Science & Climatic Change, 2012, doi:10.4172/2157-7617.1000112, 2013.

Smith, D. M., Scaife, A. A. and Kirtman, B. P.: What is the current state of scientific knowledge with regard to seasonal and decadal forecasting?, Environmental Research Letters, 7(1), 15602–15612, doi:10.1088/1748-9326/7/1/015602, 2012.

Smith, T. M. and Reynolds, R. W.: Extended Reconstruction of Global Sea Surface Temperatures Based on COADS Data (1854–1997), J. Climate, 16(10), 1495–1510, doi:10.1175/1520-0442-16.10.1495, 2003.

Smith, T. M., Reynolds, R. W., Peterson, T. C. and Lawrimore, J.: Improvements to NOAA's Historical Merged Land–Ocean Surface Temperature Analysis (1880–2006), J. Climate, 21(10), 2283–2296, doi:10.1175/2007JCLI2100.1, 2008.

Stone, R. C., Hammer, G. L. and Marcussen, T.: Prediction of global rainfall probabilities using phases of the Southern Oscillation Index, Nature, 384(6606), 252–255, doi:10.1038/384252a0, 1996.

Suárez-Moreno, R. and Rodríguez-Fonseca, B.: S4CAST v2.0: sea surface temperature based statistical seasonal forecast model, Geosci. Model Dev., 8(11), 3639–3658, doi:10.5194/gmd-8-3639-2015, 2015.

Surendran, S., Gadgil, S., Francis, P. A. and Rajeevan, M.: Prediction of Indian rainfall during the summer monsoon season on the basis of links with equatorial Pacific and Indian Ocean climate indices, Environ. Res. Lett., 10(9), 94004, doi:10.1088/1748-9326/10/9/094004, 2015.

Syed, F. S., Giorgi, F., Pal, J. S. and King, M. P.: Effect of remote forcings on the winter precipitation of central southwest Asia part 1: observations, Theor. Appl. Climatol., 86(1–4), 147–160, doi:10.1007/s00704-005-0217-1, 2006.

Syed, F. S., Giorgi, F., Pal, J. S. and Keay, K.: Regional climate model simulation of winter climate over Central–Southwest Asia, with emphasis on NAO and ENSO effects, Int. J. Climatol., 30(2), 220–235, doi:10.1002/joc.1887, 2010.

Tian, B. and Fan, K.: A Skillful Prediction Model for Winter NAO Based on Atlantic Sea Surface Temperature and Eurasian Snow Cover, Wea. Forecasting, 30(1), 197–205, doi:10.1175/WAF-D-14-00100.1, 2015.

Ummenhofer, C. C., England, M. H., McIntosh, P. C., Meyers, G. A., Pook, M. J., Risbey, J. S., Gupta, A. S. and Taschetto, A. S.: What causes southeast Australia's worst droughts?, Geophys. Res. Lett., 36(4), L04706, doi:10.1029/2008GL036801, 2009.

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

Wang, C.: Atmospheric Circulation Cells Associated with the El Niño–Southern Oscillation, J. Climate, 15(4), 399–419, doi:10.1175/1520-0442(2002)015<0399:ACCAWT>2.0.CO;2, 2002.

Wang, H. and He, S.: Weakening relationship between East Asian winter monsoon and ENSO after mid-1970s, Chin. Sci. Bull., 57(27), 3535–3540, doi:10.1007/s11434-012-5285-x, 2012.

Wang, Q.J., Schepen, A., Robertson, D.: Merging Seasonal Rainfall Forecasts from Multiple Statistical Models through Bayesian Model Averaging, J. Climate, 25, 5524-5536, dio:10.1175/JCLI-D-11-00386.1, 2012.

Wijngaard, J. B., Klein Tank, A. M. G. and Können, G. P.: Homogeneity of 20th century European daily temperature and precipitation series: HOMOGENEITY OF EUROPEAN CLIMATE SERIES, International Journal of Climatology, 23(6), 679–692, doi:10.1002/joc.906, 2003.

Wu, T.-W. and Qian, Z.-A.: The Relation between the Tibetan Winter Snow and the Asian Summer Monsoon and Rainfall: An Observational Investigation, J. Climate, 16(12), 2038–2051, doi:10.1175/1520-0442(2003)016<2038:TRBTTW>2.0.CO;2, 2003.

Wu, Z. and Lin, H.: Interdecadal variability of the ENSO–North Atlantic Oscillation connection in boreal summer, Q.J.R. Meteorol. Soc., 138(667), 1668–1675, doi:10.1002/qj.1889, 2012.

Wu, Z., Wang, B., Li, J. and Jin, F.-F.: An empirical seasonal prediction model of the East Asian summer monsoon using ENSO and NAO, Journal of Geophysical Research: Atmospheres (1984–2012), 114(D18) [online] Available from: http://onlinelibrary.wiley.com/doi/10.1029/2009JD011733/full (Accessed 17 December 2015), 2009.

Wulf, H., Bookhagen, B. and Scherler, D.: Seasonal precipitation gradients and their impact on fluvial sediment flux in the Northwest Himalaya, Geomorphology, 118(1–2), 13–21, doi:10.1016/j.geomorph.2009.12.003, 2010.

Yadav, R. K., Yoo, J. H., Kucharski, F. and Abid, M. A.: Why Is ENSO Influencing Northwest India Winter Precipitation in Recent Decades?, J. Climate, 23(8), 1979–1993, doi:10.1175/2009JCLI3202.1, 2010.

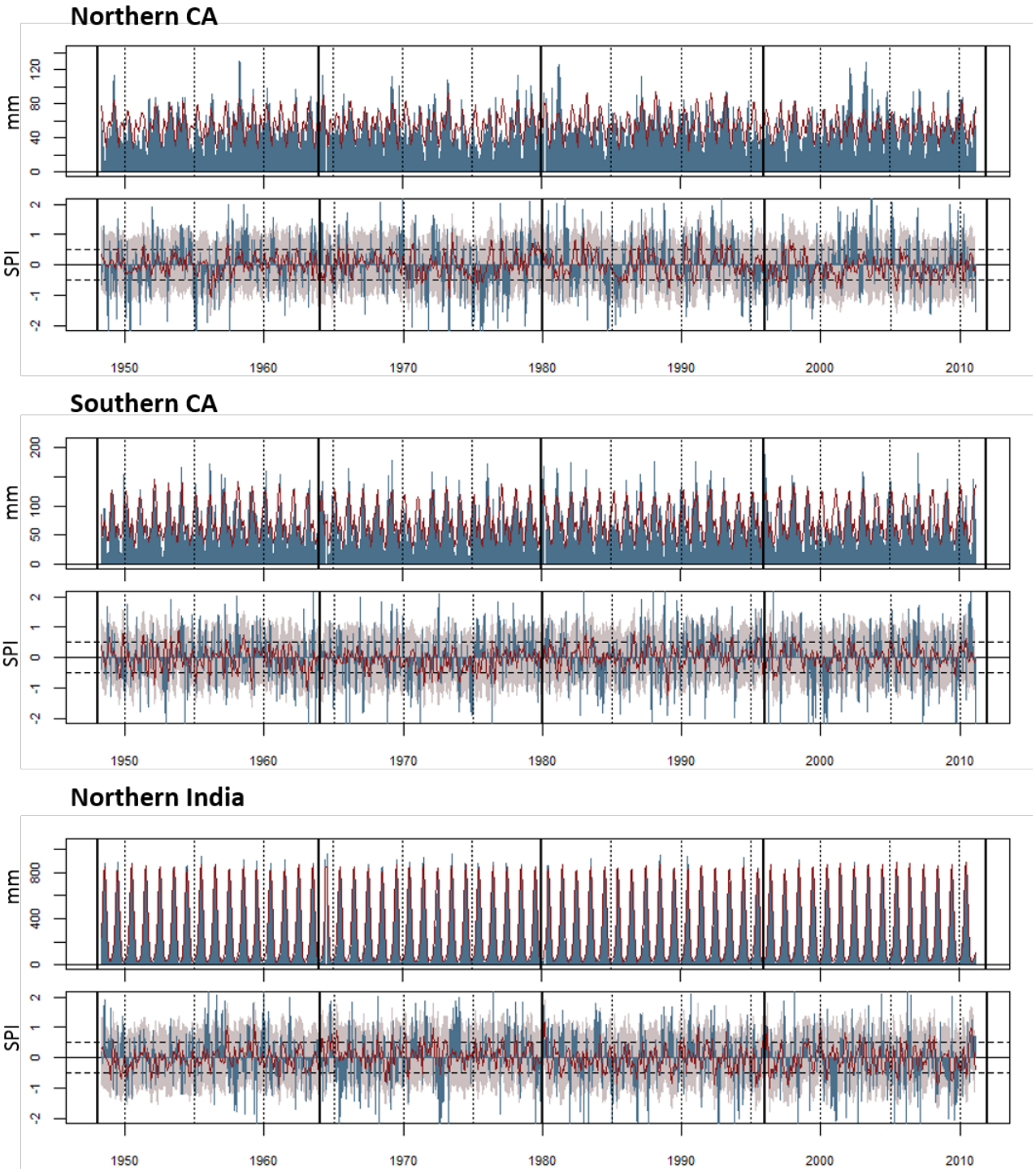
Yim, S.-Y., Wang, B., Liu, J. and Wu, Z.: A comparison of regional monsoon variability using monsoon indices, Clim Dyn, 43(5–6), 1423–1437, doi:10.1007/s00382-013-1956-9, 2013.

Yin, Z.-Y., Wang, H. and Liu, X.: A Comparative Study on Precipitation Climatology and Interannual Variability in the Lower Midlatitude East Asia and Central Asia, J. Climate, 27(20), 7830–7848, doi:10.1175/JCLI-D-14-00052.1, 2014.

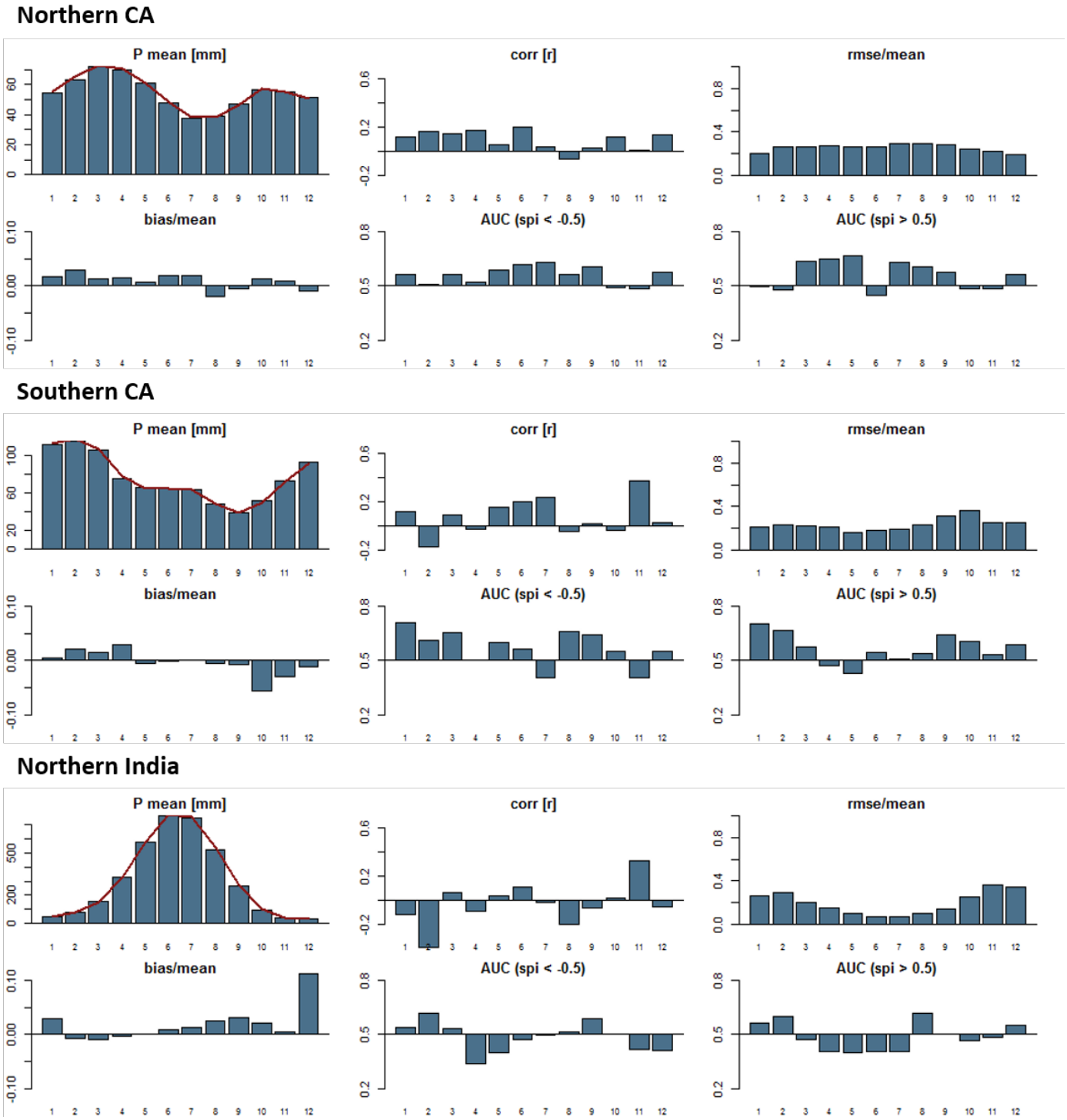
Zhang, Y., Li, T. and Wang, B.: Decadal Change of the Spring Snow Depth over the Tibetan Plateau: The Associated Circulation and Influence on the East Asian Summer Monsoon*, J. Climate, 17(14), 2780–2793, doi:10.1175/1520-0442(2004)017<2780:DCOTSS>2.0.CO;2, 2004.

Zorita, E., Hughes, J. P., Lettemaier, D. P. and von Storch, H.: Stochastic Characterization of Regional Circulation Patterns for Climate Model Diagnosis and Estimation of Local Precipitation, J. Climate, 8(5), 1023–1042, doi:10.1175/1520-0442(1995)008<1023:SCORCP>2.0.CO;2, 1995.

Suppelements:



A1: Observed running three-month precipitation totals (blue bars) and modelling results (red line) of the F[4:6] model for selected target regions. The upper panels show absolute precipitation totals for running three-month 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. Black verticals indicate the division of the time series into four independent evaluation samples.



A2: Summary of evaluation measures of the F[4:6] forecast for selected target areas. In order to keep the annual cycle of precipitation amounts, the specified month at the x-axis indicate the middle of the forecast period.