METEOROLOGICAL APPLICATIONS
Meteorol. Appl. **21**: 431–443 (2014)
Published online 15 October 2013 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/met.1429

Principal components of sea surface temperatures as predictors of seasonal rainfall in rainfed wheat growing areas of Pakistan

Floris van Ogtrop,[a] Mukhtar Ahmad[b] and Carina Moeller[c]*

[a] *Faculty of Agriculture and Environment, University of Sydney, Eveleigh, Australia*
[b] *Department of Agronomy, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan*
[c] *Tasmanian Institute of Agriculture (TIA), University of Tasmania, Hobart, Australia*

ABSTRACT: Time-lagged relationships were explored between VARIMAX rotated principal components (RCs) of sea surface temperatures (SSTs) and rainfall periods that are important for rainfed wheat production in Pakistan. Seasonal forecasts were developed using Generalized Additive Models. The first 10 RCs explained 54% of the variance in the SST data. Individual RCs were strongly ($r^2 \geq |0.5|$) to moderately ($r^2 \geq |0.3|$) correlated with climatic indices of SST anomalies associated with the El-Niño Southern Oscillation, Pacific Decadal Oscillation, Indian Ocean Dipole, and the tropical Atlantic Ocean. Forecasts of monsoon (July to September), total growing season (November to April), early (November to January) and late season (February to April) rainfall (1961–2010) were developed for Chakwal, Talagang and Islamabad. Important, linear or non-linear, time-lagged relationships were found between the RCs of SSTs and rainfall. Cross-validated forecasts were compared with real-time forecasts to evaluate the 'true' forecasting ability of the models. Continuous and categorical probabilistic forecasts were tested with an array of skill scores. Skilful forecasts of pre-season, monsoon and late-season rainfall were produced for the drier sites Chakwal and Talagang and to a lesser extent for the wetter site Islamabad. These simple, statistical forecasts can be developed with minimal financial investment. However, consideration of the potential uses of such forecasts will require a reflective decision framework that engages stakeholders and addresses socio-economic and agro-ecological constraints not included here.

KEY WORDS seasonal climate forecasting; probabilistic forecasts; categorical forecasts; forecast skill; generalized additive model; monsoon; wheat

Received 2 June 2012; Revised 17 June 2013; Accepted 19 June 2013

1. Introduction

Crop production in the *barani* (rainfed) areas of the Pothwar plateau in the northern Punjab, Pakistan, is risky because of high rainfall variability. Pothwar is the largest rainfed cropping area in Pakistan. Most of the arable land is allocated to wheat (*Triticum aestivum*), Pakistan's main stable crop and contributor to food security (Punjab Barani Commission, 1976; Khan and Shah, 2010). Rainfall in the region is closely associated with the South-Asian summer monsoon: winds from the Indian Ocean bring heavy rains from June to September, which peak during July and August, contributing over 50% to the annual rainfall (Suleman *et al.*, 1995). Wheat is grown in the following *rabi* (winter) season, which begins in October to December and ends in April to May. Variability of both pre-sowing and in-crop rainfalls causes large fluctuations in yield from season to season. This hinders the development of environmentally and economically sustainable cropping systems in Pothwar (Byerlee and Husain, 1993).

Farmers in dryland regions such as Pothwar are responsive to climatic fluctuations (Stewart, 1991; Stone and Meinke, 2005). However, the experiential knowledge gained from past climatic vagaries, such as erratic and deficient rainfall, does not necessarily prevent future failures: farmers in Pothwar did not foresee the drought of the 2009/2010 season, which caused widespread crop failures and left about 60% of farms without return on investment (M. Ahmad, personal communications). A means to increase the preparedness of farm managers above a level that is solely based on experience is the application of skilful seasonal forecasts ahead of critical periods in the farming and crop calendar. The importance of each month's rainfall for yield depends on crop physiological factors related to growth and development, and farmers must devise management practices that optimize the productive use of scarce rainfall and its conversion into yield to be successful (Passioura and Angus, 2010; Fischer, 2011). The summer monsoon is important for moisture storage in the soil profile prior to the start of the cropping season, especially in the low rainfall zone of Pothwar (Arif and Malik, 2009). Among agricultural decisions that could benefit from forecasts of monsoon rainfall are those related to the area sown and the consequent amount of seed and fertilizer to be purchased, choice of soil type (higher *vs* lower water-holding capacity) as well as the crop/cultivar type (long- *vs* short-season maturity types) (Meinke *et al.*, 2005; Stone and Meinke, 2005; Moeller *et al.*, 2009). The rainfall during October to December is critical for seed germination and crop establishment, and farmers may adjust the rates of fertilizer applied at the start of the season depending on the outlook (Moeller *et al.*, 2008). Late winter and early spring rainfalls are vital for biomass accumulation prior to flowering, which is positively correlated with yield (a vigorous crop stand at this stage is a prerequisite for potentially achieving

* Correspondence: C. Moeller, Tasmanian Institute of Agriculture (TIA), University of Tasmania, Hobart, Australia. E-mail: carina.moeller@utas.edu.au

high yield), while March to May rainfall is important for grain formation and filling (Passioura and Angus, 2010). Decisions that can benefit from forecasting mid and late season rainfall are those related to tactical crop management (fertilizer and pesticide use), harvest preparations and marketing (Everingham *et al.*, 2002; Stone and Meinke, 2005). With all these decisions, the aim is to minimize the chance of economic and/or environmental losses and profit from the upsides of rainfall variability (Hansen, 2005; Moeller *et al.*, 2008). At the regional/national level, skilful seasonal climate forecasts issued 2–3 months ahead may assist planners to estimate the size of the harvest to make timely decisions about importing agricultural commodities to meet future demand (Barlow *et al.*, 2002; Hansen and Indeje, 2004; Hansen, 2005).

To realize opportunities related to the application of seasonal climate forecasts in agriculture, it requires that the observed pattern of rainfall variability in an agricultural region is forced by changes in climatic predictors such as sea surface temperatures (SSTs). Moreover, it is important that there is a time-lagged relationship between the climatic predictors and rainfall (Drosdowsky and Chambers, 2001; Goddard *et al.*, 2001; Stone and Meinke, 2005). Due to their large heat capacity, the oceans are a key source of predictable climate variability on several time scales (e.g. monthly, seasonal, and inter-annual). The tropical oceans are the most important regions for coupled ocean–atmosphere interactions (Chang *et al.*, 2006); better understanding of these interactions has improved long-range climate forecasting capabilities (Cane, 2000; Goddard *et al.*, 2001). However, the climate system has also many chaotic features, which limit predictability beyond a theoretical threshold (Westra and Sharma, 2010).

Large-scale SST patterns in the Pacific, Atlantic and Indian Ocean are related to rainfall anomalies across the Indian Ocean basin (Clark *et al.*, 2000; Li *et al.*, 2001; Chang *et al.*, 2001, 2006; Krishnan and Sugi, 2003; Goswami *et al.*, 2006; Zhang and Delworth, 2006; Ashok and Saji, 2007; Schott *et al.*, 2009). The El-Niño Southern Oscillation (ENSO) is the leading mode of tropical climate variability, and a climatic phenomenon of global significance characterized by warming/cooling cycles of sea surface waters in the eastern equatorial Pacific Ocean and the consequent changes in zonal air pressure gradients and convection shifts in the equatorial Pacific Ocean (Allan, 2000; Cane, 2000; Schott *et al.*, 2009). During the 'warm' ENSO event (El-Niño), warm SST anomalies occur in the eastern Pacific; cells of intense convection shift eastward and rainfall intensifies over the central to eastern equatorial Pacific. The climatic pattern is the opposite during the 'cold' ENSO event (La-Niña). The ENSO is locked into the seasonal cycle: eastern Pacific SST anomalies typically develop during the boreal summer, peak in winter and break during the boreal spring. However, extreme phases of the ENSO have a tendency to last for about 18–24 months once they have become established (Allan, 2000). The large-scale shifts in convection over the Pacific affect the atmospheric circulation over the Indian Ocean. The tropical Indian Ocean gradually warms during El-Niño episodes. Following the height of El-Niño (boreal winter), the Indian Ocean warming peaks during the boreal spring (February to April), and persists until the summer following the event. This warming increases rainfall over the Indian Ocean and the surrounding regions after the El-Niño decays (Schott *et al.*, 2009). Severe drought years on the Indian subcontinent have always been accompanied by El-Niño (Kumar *et al.*, 2006).

The Indian Ocean Dipole (IOD) has been described as a coupled ocean-atmosphere phenomenon defined by the difference in SSTs between two areas in the tropical Indian Ocean and accompanying anomalous low-level winds, and a significant contributor to rainfall variability in countries that surround the Indian Ocean (Saji and Yamagata, 2003; Terray and Dominiak, 2005; Schott *et al.*, 2009). During positive IOD events, warm SST anomalies occur in the western Indian Ocean and cold SST anomalies in the east off Sumatra. The IOD has a strong seasonality and peaks during the boreal fall (September to November); an east–west dipole of anomalous rainfall establishes with increased rainfall in the west. Easterly wind anomalies blow from the cold/dry to the warm/wet regions and amplify the SST cooling in the east. This SST pattern is of the opposite sign during negative IOD events (Saji and Yamagata, 2003; Ashok *et al.*, 2004; Schott *et al.*, 2009).

However, while it has been originally proposed that the IOD is largely distinct from the ENSO (e.g. Saji *et al.*, 1999) others have argued that the IOD is a representation of another phase of the ENSO and therefore not an independent mode of Indian Ocean climate variability (Allan *et al.*, 2001). It is thought that Pacific and Indian Ocean variability are linked through an extension of the Walker Circulation to the west and associated Indonesian throughflow (Bracco *et al.*, 2005; Schott *et al.*, 2009; Ummenhofer *et al.*, 2011). Analyses of IOD and ENSO events show that positive (negative) IOD events often, but not always, co-occur with El Niño (La Niña), and that a significant number of IOD events occur during neutral ENSO years (Meyers *et al.*, 2007). In the Asian summer monsoon regions, a strong positive IOD generally reduces the rainfall deficit induced by El-Niño over western India and parts of Pakistan (Ashok *et al.*, 2004; Ashok and Saji, 2007).

Monsoonal rainfall is thought to be modulated by variability in northern Pacific Ocean SSTs occurring on 10–15 year timescales. This modulation has been described as the Pacific Decadal Oscillation (PDO), and has been associated with a decreased (increased) strength of the monsoon over the subcontinent during its warm (cold) phase (Krishnan and Sugi, 2003; D'Arrigo and Wilson, 2006). However, others have shown that the PDO is strongly correlated with the ENSO at all timescales (Newman *et al.*, 2003). Power and Colman (2006) argue that the PDO may be simply the ENSO signal plus an unknown (random) modulation which appears to be periodic on an approximately decadal scale. While PDO signals have been linked to climate fluctuations in the northern Pacific and Asian region (Barlow *et al.*, 2001; D'Arrigo *et al.*, 2001; D'Arrigo and Wilson, 2006), questions remain about whether the PDO is a robust, predictable feature of Pacific Ocean climate variability (Meinke *et al.*, 2005).

North Atlantic SSTs have been linked to the Indian monsoon through the modulation of tropospheric temperatures over Eurasia and by coupled ocean-atmosphere feedbacks in the Indian and western Pacific Oceans (Chang *et al.*, 2001; Goswami *et al.*, 2006; Lu *et al.*, 2006). Furthermore, tropical Atlantic SSTs have been shown to moderate the influence of ENSO on the strength of the Indian monsoon (Kucharski *et al.*, 2007, 2008). However, the mechanisms underlying the connections between the ENSO and tropical Atlantic SSTs have yet to be further explored (Kucharski *et al.*, 2007).

This study explores climatic predictors (indices derived from global SSTs) associated with the climate phenomena reviewed above that influence rainfall variability over the Indian subcontinent for their usefulness in forecasting seasonal rainfall (monsoon rainfall and three rainfall intervals during the wheat growing season) at three agriculturally important sites on the Pothwar plateau. Firstly, VARIMAX rotated principal components of SSTs (the climatic predictors) were calculated and the strength of the correlation with the well-described climate phenomena was assessed. Secondly, time-lagged relationships between the climatic predictors and rainfall in upcoming months were established using Generalized Additive Models (GAM) (van Ogtrop *et al.*, 2011). The GAM models were used to forecast seasonal rainfall at each of the study sites. Thirdly, the forecast performance was assessed and the forecast skill was evaluated using a number of cross-validated skill scores (Jolliffe and Stephenson, 2003; Wilks, 2011). To our best knowledge, this is the first study to develop and conduct statistical forecasts of seasonal rainfall using SSTs for rainfed wheat growing locations in Pakistan.

2. Methods

2.1. Site description

The Pothwar (or Pothohar) plateau (350–580 m a.s.l.) in the northern Punjab extends from 32° 10–34° 9′ N to 71° 10–73° 10′ E and covers about 1.82 million ha of which one third is arable land. Pothwar has a subtropical thermal climate with summer-dominant rainfall (Fischer *et al.*, 2002). Rainfall environments range from semi-arid (about 250 mm average annual rainfall) in the southern and western parts of the plateau to sub-humid ($>$ 1000 mm average annual rainfall) in the northeastern regions around the capital Islamabad (Byerlee and Husain, 1993; Arif and Malik, 2009).

The three study sites were Islamabad (33° 40′ N, 73° 10′ E, 508 m a.s.l.), Chakwal (32° 56′ N, 72° 52′ E, 513 m a.s.l.), and Talagang (32° 55′ N, 72° 25′ E, 458 m a.s.l.). Chakwal and Talagang are located about 130 km southwest of Islamabad. The climate at Islamabad is dry sub-humid with $>$ 1000 mm average annual rainfall (high rainfall zone) and an average annual temperature of 21.3°C. The semi-arid site Chakwal is situated in the medium rainfall zone, and has an average annual temperature of 22.4°C. Talagang (about 45 km west of Chakwal) has an arid climate and is located in the low rainfall zone. The average annual temperature is 23.7°C at Talagang. The annual potential evapotranspiration (FAO Penman–Monteith) is about 1600 mm at all sites, and the rainfall variability during the wheat growing season (November to April) is greater than the annual variability (Table 1).

2.2. Derivation of climatic predictors

Global SST anomalies (1961–2010) were extracted from version 2 of the extended Kaplan SST dataset (Reynolds and Smith, 1994; Kaplan *et al.*, 1998). The data can be downloaded from http://iridl.ldeo.columbia.edu/SOURCES/.KAPLAN/.EXTENDED/.v2/.ssta (International Research Institute for Climate and Society, 2012a). Principal Component Analysis (also referred to as Empirical Orthogonal Function Analysis) of the SSTs (1207 grid cells) was conducted to reduce the dataset into a manageable number of covariates, and understand the SST data in terms of a much smaller number of prominent modes (Westra *et al.*, 2010). Principal components are essentially linearly uncorrelated, which is important to avoid collinearity in the data used for statistical rainfall forecasting (Graham, 2003). The principal components (PCs) of the SSTs were calculated using the principal function in the 'psych' package (Revelle, 2011) in the open source program R (R Development Core Team, 2011). For the specific analyses presented here, the PCs of the SSTs were rotated using VARIMAX rotation (Westra *et al.*, 2010; Revelle, 2011). The rotated PCs were chosen following initial tests, which showed that the forecasting models tended to perform better using rotated PCs compared to unrotated PCs (results not shown). An advantage of rotated PCs is that they allow more localized features, which are often better aligned with the physical mechanisms underlying climatic phenomena (Appendix) than unrotated PCs, to be obtained (Hannachi *et al.*, 2007; International Research Institute for Climate and Society, 2012b). Advantages/disadvantages such as robustness of the results and physical justification have been discussed in detail by Jolliffe (1987, 2002), Hannachi *et al.* (2007) and Richman (1986, 1987).

The first 10 VARIMAX rotated principal components (referred to as RC1, RC2 … RC10), which explained 54% of the variance in the SST data, were used as predictors in the forecasting model described in Section 2.3. The strength of the relationships between these RCs and the well-described climate indices (Appendix) were assessed using Pearson's correlation analysis. Because the RCs are not the same as the climate indices (Appendix), correlation analysis is typically used to assist with the interpretation of the RCs (Dommenget and Latif, 2002; Westra *et al.*, 2010). These climate indices can be downloaded from http://www.esrl.noaa.gov/psd/data/climateindices/list/ (National Oceanic and Atmospheric Administration, 2012).

2.3. Rainfall forecasting

The forecast periods chosen are relevant for decision-making and productivity in wheat-based farming systems of the study region. These were: the (1) peak monsoon period (three forecast periods: July to September, July to August, and August to September), (2) the wheat growing season (November to April), (3) crop establishment and early growth (November to January), and (4) the period of rapid biomass accumulation, anthesis, and maturation (February to April (Table 1)). The SST predictors were obtained for month/s preceding the forecast periods, and lag-periods (forecast lead-times) between 1 and 6 months were considered. For the monsoon period, SST predictors were obtained for the April to May period. For all other forecast periods, SST predictors were derived for the month/s: July, August and September, July to August, August to September, and July to September. If not otherwise indicated, names of months are denoted by their first letter in subsequent sections.

Time-lagged statistical relationships between the climatic predictors (i.e. the 10 VARIMAX RCs) and rainfall were established using the Generalized Additive Model (GAM) in R. A GAM is an extension of a Generalized Linear Model and allows for non-normal response distributions (e.g. rainfall), non-linearity in the model structure (e.g. relationship between rainfall and the sum of the VARIMAX RCs), and non-linearity in the predictor variables (e.g. relationship between rainfall and VARIMAX RCs) (Wood, 2006).

Accounting for non-linearity in the covariates is achieved by adding smoothing functions (splines) of the covariates to the models (Wood, 2008). Fewer predictors can be used in a GAM as including a spline in a regression equation is equivalent to fitting, depending on the complexity of the chosen spline, two or more covariates.

To implement the forecasting models, the 'gam function' in the 'mgcv package' of R was applied (Wood, 2006). The default 'thin plate splines' was used (Wood, 2003). For each site, model parameter selection was achieved by applying a ridge penalty term allowing each term to shrink to zero (Copas, 1983; Wood,

Table 1. Rainfall statistics for five periods at the study sites (individual months are denoted by their first letter).

Site		Annual	NDJFMA[a]	JAS[b]	NDJ[c]	FMA[d]
Islamabad	Mean	1070	302	590	97	206
	Median	1060	270	593	89	176
	CV%[e]	28	54	39	47	66
Chakwal	Mean	600	240	282	61	180
	Median	576	257	283	55	166
	CV%	30	42	51	74	49
Talagang	Mean	290	114	124	30	85
	Median	292	115	136	27	77
	CV%	29	43	56	75	46

[a]Total growing season rainfall for wheat. [b]Pre-sowing, monsoon rainfall. [c]Crop establishment and early growth. [d]Rapid biomass accumulation, anthesis and maturation. [e]The co-efficient of variation was calculated as CV% = mean/standard deviation × 100. Monthly rainfall data (1961–2010) were sourced from the Meteorology Department of Pakistan (http://www.pmd.com.pk/)

2006). In other words, this shrinkage approach was used to select only those RCs for the forecasting models that showed an important time-lagged relationship with rainfall at the study sites.

The following model was considered:

$$Seasonal\ rainfall = \beta_o + \sum_{i=1}^{n} g_i\,(\text{lag } RC_i)$$
$$Seasonal\ rainfall \sim N\,(\mu, \sigma) \quad (1)$$

where g is a thin plate spline, i is the i^{th} selected RC covariate and n is the total number of RC covariates used. The coefficient β_o is the intercept and lag RC_i is the i^{th} lagged RC covariate where the lags correspond to those mentioned above. Due to the limited number of data points (49 seasons; 1961–2010), the maximum number of knots was restricted to five in the model. The exact number of knots was chosen using the default generalized cross-validation (Wood, 2006). The combinations of site, forecast and lag period, and month/s for which VARIMAX RCs were calculated, resulted in 63 forecasting models.

2.4. Model evaluation

Leave-one-out cross-validation was used for model calibration and evaluation (van Ogtrop *et al.*, 2011). The aim was to minimize the risk of artificial forecast skill, which is an optimization bias generated by small sample sizes (here: 49 seasons). The cross-validation process involved iterative cycles in which the models were calibrated on the full data set minus one season and validated on the left-out seasons.

In addition to cross-validating the models' ability to forecast rainfall, a split- and an expanding-window evaluation was performed for independent evaluation. Large deviations from the cross-validated forecasts would indicate that those results are not indicative of the true predictive skill of the forecasts. In the split-window approach, the calibration/training period was 1961–1995 and the verification period was 1996–2008. The expanding-windows approach involved forecasting one season ahead, i.e. data from 1961 to 1995 was used to forecast 1996 rainfall, and then data from 1961 to 1996 was used to forecast 1997 rainfall. This process was repeated for seasons up to 2008. This process is equivalent to a real-time forecast as it assumes no knowledge of future climate. The evaluation approaches were compared using the measures r^2 and percent bias (pBIAS; an indicator of whether a model is consistently over- or under-predicting) for the 1996–2008 period.

2.5. Goodness of fit

Seven skill scores were used to test the goodness of fit for both the continuous and probabilistic forecasts produced by the models. The R package 'hydroGOF' (Zambrano-Bigiarini, 2011) was used to calculate the root mean square error (RMSE), a measure of the magnitude of the forecast error, and pBIAS (explained above). A RMSE score of zero indicates a perfect forecast and a pBIAS value of 0% indicates that the forecast is unbiased. The R package 'verification' (NCAR – Research Application Program, 2010) was used to calculate the linear error in probability space (LEPS), a measure of the forecast error, and the Brier score (BS) and response operating characteristic curve, both measures of a forecasts ability to predict two alternative outcomes correctly. A LEPS score of one indicates a perfect forecast and any value equal to or below zero indicates no skill. A Brier score of one indicates no skill and zero indicates a perfect forecast. The ROC value is calculated as a p-value with $p < 0.05$ indicating a skilful forecast. The S-score (S%) is the percentage of correct forecasts above the percentage expected due to chance alone (Moeller *et al.*, 2008). A value of S% greater than 0% is considered a skilful forecast. Further detail on goodness of fit tests can be found in Wilks (2011) and Jolliffe and Stephenson (2003). The skill scores were cross-validated as described above in Section 2.4. The probabilistic rainfall forecasts were obtained from the assumed rainfall distribution (van Ogtrop *et al.*, 2011).

3. Results

3.1. Interpreting the VARIMAX rotated principal components

The first 10 VARIMAX RCs explained 54% of the variance in the SST data. Figure 1 shows contour plots of RC2 and RC9, which together explained 20% of the variance. The contours represent the loadings of each component. It can be seen that higher loadings tend to centre in certain oceanic regions: RC2 shows high positive loadings in the Pacific Ocean, while RC9 shows high loadings around the Atlantic Ocean south of the equator.

The RC2 was highly correlated with SSTs in the central and western Pacific Ocean indicating its relation to the ENSO phenomenon (Table 2, Figure 1). Furthermore, RC9 was strongly correlated with the Tropical Southern Atlantic index (TSA) (Table 2, Figure 1) and RC3 with the Tropical Northern Atlantic index (TNA), and the Nino 1 + 2, and Nino 3 regions. The IOD

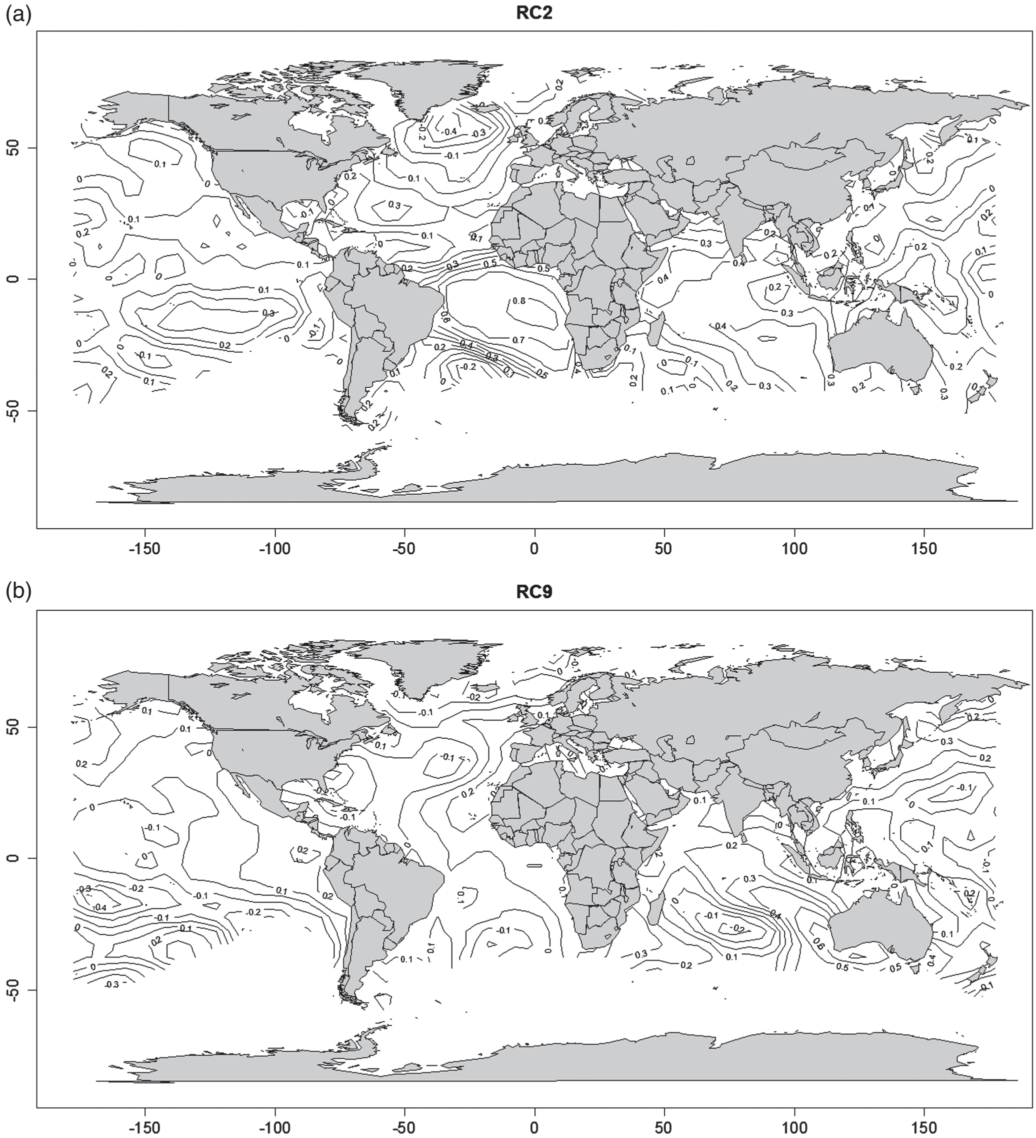

Figure 1. Contour plots of the first two VARIMAX rotated principal components, which explained 20% of the variance in the SST data, showing high loadings in the central Pacific Ocean (RC2) and the southern Atlantic Ocean (RC9).

was most strongly related to RC1, RC2 and RC8. The PDO was reasonably well correlated with the RC2, RC4 and RC10.

3.2. The forecasting models

The forecasts produced using the three different approaches for forecast evaluation (cross-validation, split- and moving-window methods) all followed a similar pattern (Figure 2). The positive trend in the rainfall data between 1970 and 1995 (Figure 3) was captured in all evaluation approaches as exemplified here for monsoon forecasts (JAS) at Chakwal (Figure 2). This suggests that the relationship between rainfall and SSTs is approximately stationary in this region. Forecast results obtained using the cross-validation and expanding-window approach were in close agreement as revealed by an r^2 of 0.71 and pBIAS of −0.3% (Figure 2). The cross-validation and split-window evaluation approaches were less similar as indicated by an r^2 of 0.42 and pBIAS of 12.9% (Figure 2). Because the cross-validation approach makes maximum use of the limited, available data (49 seasons only), and the evaluation methods produced overall reasonably similar forecast results, only the cross-validated forecasts are presented in subsequent sections.

The model that best described time-lagged relationships between a specific rainfall period and the climatic predictors depended on the site. Significant RCs are shown in Table 3. Generally, the relationships between total seasonal rainfall and the RCs related to Indian, Pacific and Atlantic Ocean SSTs were significant at all sites. This section presents only the skilful subset of cross-validated forecast models (see Section 3.3). Results for all forecasts and lag periods considered in this study are given in an Appendix.

The forecasting models captured the increasing trend in total seasonal rainfall in the monsoon period (Figure 3(a) and (b)). Similarly, wet and dry periods during the mid 1970s, late 1970s, and early 1980s have also been captured in the models for the FMA period (Figure 3(c)). Forecasts for the FMA period performed better up until the 1980s and then after 2007. In contrast, forecasts of monsoon period performed better in the 1980s, 1990s and early 2000s. The wet period in the late 1990s was captured by most models.

Table 2. Pearson correlation co-efficients for 10 VARIMAX rotated principal components (RC1, RC2 ... RC10) and climate indices describing known climate phenomena.

	Cumulative variance explained (%)	Nino1 + 2	Nino 3.4	Nino3	Nino4	IOD	TNA	TSA	PDO
RC2	12	0.30	0.79	0.58	0.76	0.39	0.14	−0.20	0.38
RC9	20	−0.23	−0.07	−0.12	−0.09	0.17	−0.07	0.80	−0.10
RC1	26	0.35	−0.26	0.04	−0.41	−0.32	0.26	0.23	−0.12
RC3	31	−0.51	−0.08	−0.30	0.19	0.06	0.69	0.12	−0.13
RC6	36	0.19	0.00	0.12	−0.08	0.04	−0.09	−0.41	0.25
RC4	40	−0.04	−0.04	−0.05	0.03	−0.26	−0.09	−0.07	0.32
RC10	44	0.02	0.02	−0.01	0.16	0.01	−0.02	−0.01	−0.59
RC5	48	−0.04	0.00	−0.01	0.06	0.23	0.34	0.11	−0.19
RC7	51	0.13	0.17	0.18	0.05	0.11	0.14	−0.05	−0.02
RC8	54	0.37	0.04	0.21	−0.11	−0.31	−0.03	−0.10	0.13

A value of one indicates perfect correlation and zero indicates no correlation. The RCs are noted in order of the variance explained. See appendix for a list of climate indices.

Table 3. Time-lagged relationship between rainfall during different periods and 10 VARIMAX rotated principal components of SSTs (RC1, RC2 ... RC10). The RC was either significant at $p <= 0.05$ and selected in the forecasting model (S), was selected but was not highly significant (+), or was non-significant (−).

Region	Chakwal	Talagang			Islamabad
SST month/s	AM	AM	S	AS	S
Rainfall period	JAS	JAS	FMA	NDJFMA	FMA
RC2	−	−	S	S	S
RC9	S	S	S	S	S
RC1	S	S	S	S	−
RC3	−	−	+	−	S
RC6	S	+	+	S	+
RC4	−	−	−	−	−
RC10	+	S	S	S	−
RC5	−	−	−	−	−
RC7	−	−	S	S	S
RC8	−	−	+	+	S

The RCs are noted in order of the variance explained. Individual months are denoted by their first letter.

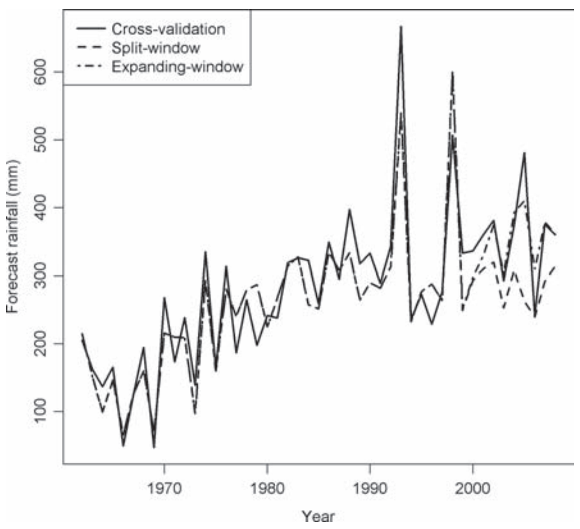

Figure 2. Comparison of three forecast evaluation methods: cross validation, split- and expanding-window method. Shown are forecasts of monsoon (JAS) rainfall using VARIMAX rotated principal components of April to May SSTs. The location is Chakwal.

The model parameters RC1, RC6, RC9 and RC10 explained rainfall during the month JAS at Chakwal and Talagang (see Appendix for the climate indices of well-described climate phenomena). The model parameters RC2, RC6, RC7, RC8 and RC9 were included in the FMA forecasts for both Talagang and Islamabad (Table 3, Figure 3). The RC1 and RC10 were only significant for the FMA forecast at Talagang. The significant RCs were similar for the FMA and NDJFMA forecasts for Talagang. The RC 4 and RC5 were not significant in any of the models given in Table 3.

In addition to assessing which RCs affect rainfall, linear/non-linear properties of the relationship between each of the significant RCs and rainfall were explored. Figure 4 shows the significant relationships between the monsoon (JAS) at Talagang and the significant mean April to May RCs. The RC1 and RC9 both showed non-linear behaviour. The SST anomaly values above 0.5 had a marginal effect on rainfall as indicated by the fact that the value zero was included in the confidence intervals. Below this value, RC9 was approximately linear and decreasing. The inverse was true for RC1. In contrast, RC10 was positive linear, which also demonstrated the flexibility of using a GAM whereby a relationship can be either linear (the number of knots is approximately 1) or non-linear (the number of knots is greater than 1). In summary, Figure 4 suggested an inverse relationship between RC1 and RC9 and a similar relationship between RC1 and RC10. That is, a lower (higher) value of RC9 and higher (lower) values of RC1 and RC10 indicated higher (lower) monsoon rainfall at Talagang. As the values for the RCs moved towards the positive and negative extremes, the uncertainty in the relationship with rainfall increased.

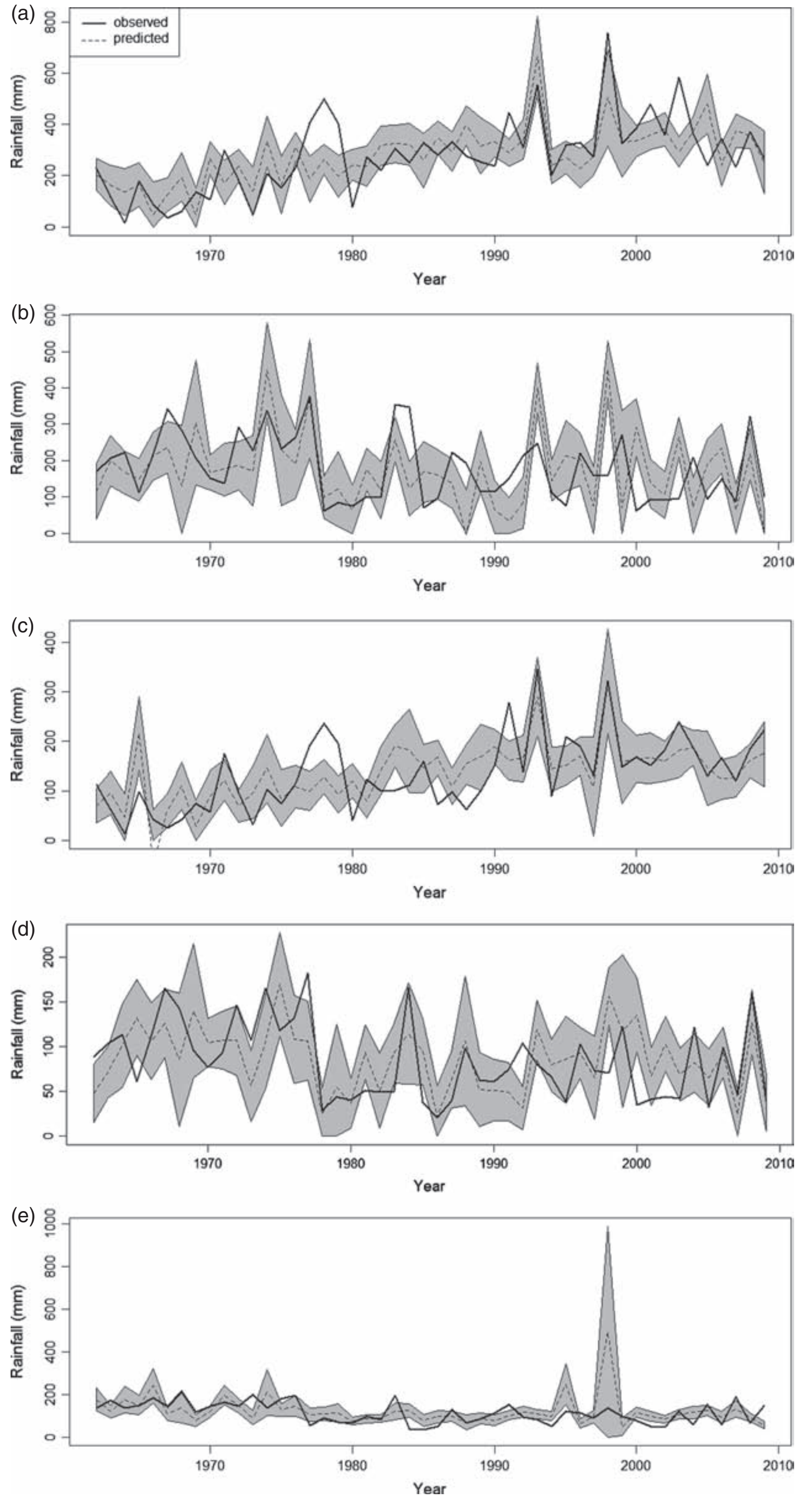

Figure 3. Continuous forecast of seasonal rainfall (dotted line), 95% confidence intervals for forecasts (shaded area), and measured seasonal rainfall (solid line) for Pothwar, Pakistan. The locations and rainfall periods were: (a) Chakwal JAS, (b) Talagang JAS, (c) Talagang FMA, (d) Talagang NDJFMA and (e) Islamabad FMA.

The normal quantile plot (qqplot), which was derived for each of the 63 cross-validated models, indicated that the assumption that seasonal rainfall follows a Gaussian distribution is not entirely valid because not all of the values, particularly values lower than −50, lie on the 1:1 line (Figure 5(a)). This is also reflected by the histogram being slightly skewed (Figure 5(b)). However, without the two outliers the residual scatter plot showed a reasonable scatter (Figure 5(c)). There appeared to be more values concentrated between −50 and +50, and a thin scatter above +50 illustrated some skewness

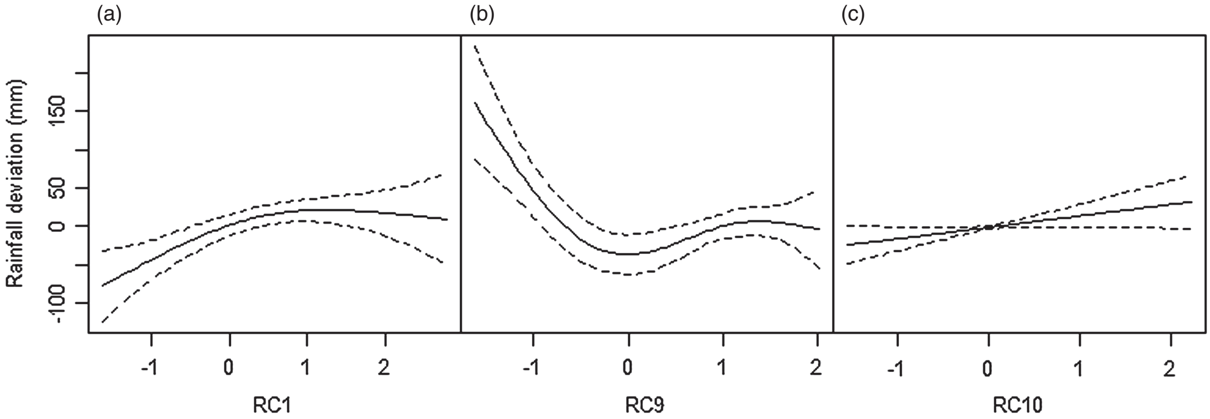

Figure 4. Relationships between individual VARIMAX rotated principal components (covariates RC1, RC9, and RC10) and monsoon rainfall (JAS) at Talagang. For each covariate, the modelled rainfall is given as deviation from the intercept (132.6 mm). The dotted lines indicate the 95% confidence intervals for each covariate. The number of knots in the smoothing functions was (a) 2.11 for RC1, (b) 3.18 for RC9, and (c) 0.84 for RC10. In the forecasting model, the predicted rainfall is the sum of the three smoothing functions (a–c) plus the intercept.

in the residuals. The scatter-plot of the observed *versus* the fitted model showed a reasonable fit, although the model slightly under-predicted seasonal rainfall. The two extreme rainfall seasons (>300 mm) were modelled with reasonable accuracy.

3.3. Forecast skill

Subsequently, the skill of the models for forecasting rainfall at the study sites was evaluated. There was generally greater skill in forecasting rainfall at the drier sites Chakwal and Talagang than at the wetter site Islamabad. Table 4 shows the subset forecasts for which all cross-validated skill scores were significant. Values of the skill scores for all forecasts and lag periods considered in this study are given in an Appendix.

Forecasts for the monsoon period showed significant skill at both Chakwal and Talagang and for the entire growing season at Talagang (Table 4). The model had generally no skill for forecasting rainfall during NDJ (crop establishment and early crop growth). Forecasts for Talagang and Islamabad showed significant skill at the end of the growing season (FMA).

The results of the skilful continuous forecasts for Chakwal, Talagang and Islamabad (Table 4) showed that the VARIMAX RCs of the SSTs explained between 16 and 38% of total seasonal rainfall variability ($r^2 \times 100$) with a RMSE between 42.5 and 122.1 mm rainfall. The pBIAS ranged between −1.90 and 4.90%, and the LEPS scores also showed skill with values ranging between 0.23 and 0.31.

The skill score S% for categorical forecasts (above and below median seasonal rainfall) ranged from 22 to 71% indicating a marked improvement over relying on chance alone (i.e. $S\% <= 0$). Similarly, the probabilistic forecasts, which are presented as the probability of exceeding median seasonal rainfall, gave Brier score values ranging from 0.14 to 0.28 and the ROC values were significant at $p < 0.05$.

4. Discussion

The analyses showed that there is scope for skilful seasonal rainfall forecasting in Pothwar, especially in the arid to semi-arid areas of Talagang and Chakwal, and, to a lesser extent, at the dry sub-humid site Islamabad. Parameter selection was generally consistent for Talagang and Chakwal which would be expected considering their close proximity (Talagang is only 45 km west of Chakwal but rainfall drops considerably over this distance). However, subtle differences occur, such as the level of significance with RC6 and RC10 for the two JAS forecasts (Table 3). Given the complexity of the interactions being modelled and potential input uncertainties, it is likely that these differences may occur between different rainfall stations particularly when parameters are only marginally influential in the model. Generally, the parameters selected and dropped in the cross validation process were consistent for each model.

The modelled relationships between the RCs and rainfall were linear and non-linear (Figure 4). It has been shown that being able to capture non-linearity is essential for describing relationships between SST anomalies and rainfall (Boulanger *et al.*, 2005). The fact that RC1 and RC9 are significant for both Talagang and Chakwal suggests significant influence of the Indian and Atlantic Oceans on rainfall (Ashok and Saji, 2007; Kucharski *et al.*, 2008). It would also be expected that monsoonal rainfall is well correlated with the ENSO (Kumar *et al.*, 2006). This is represented in the model by RC1, which shows a positive correlation with eastern Pacific Ocean SSTs and a negative correlation with western Pacific Ocean SSTs. Similarly, as RC1 increased so did rainfall, which is consistent with a La-Niña event. Additionally, RC10, which is negatively correlated with the PDO (−0.59, Table 2), was selected in both models. This may support the finding that the PDO moderates the effect of ENSO on the Indian monsoon (Krishnan and Sugi, 2003).

The skilful seasonal climate forecasts presented here are locally relevant, and could potentially inform risk management strategies in dryland farming systems if embedded in a reflective decision framework (Weaver *et al.*, 2013). Generally, skillful forecasts have economic and/or environmental benefits only if they can change the course of decision-making therefore reducing vulnerability (Hansen, 2002). For wheat systems in Australian drylands, Moeller *et al.* (2008) showed that a skill level of at least $S\% = 26$ was required for seasonal forecasts of above/below median rainfall to be useful in managing applications of fertilizer nitrogen. They defined the deterministic value of the forecast as the additional gross margin achieved over fertilizer strategies that ignore any forecasts. The threshold level of skill required for the forecast

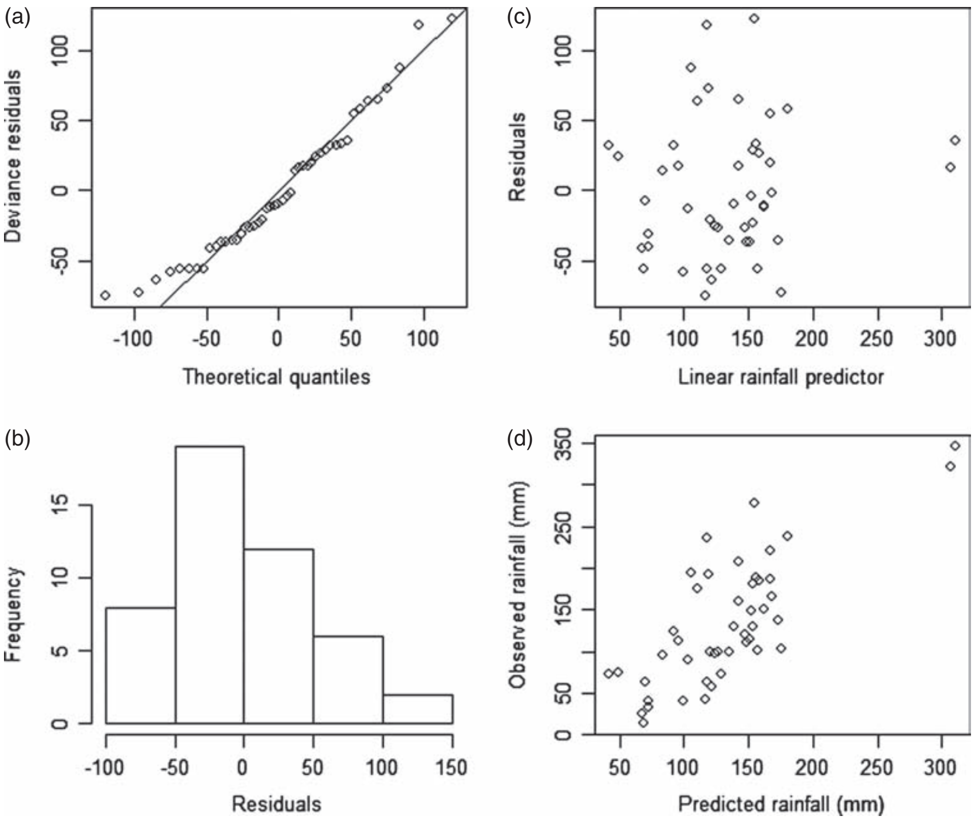

Figure 5. Quality checks for one of the 63 cross-validated models forecasting monsoon rainfall (JAS) at Talagang: (a) normal quantile plot, (b) histogram of residuals, (c) residual plot, and (d) xy scatter plot of observed *versus* predicted rainfall. Details are given in text.

Table 4. Values of skill scores for skilful forecasts of seasonal rainfall at three locations in Pothwar, Pakistan (individual months are denoted by their first letter).

	Chakwal	Talagang	Talagang	Talagang	Islamabad
SST month/s	AM	AM	S	AS	S
Forecast period	JAS	JAS	FMA	NDJFMA	FMA
Skill score					
r^2	0.37	0.38	0.21	0.19	0.16
RMSE	118.2	57.7	42.5	51.4	122.1
pBIAS	0.6	−0.2	−2.6	−3.0	−4.9
LEPS	0.31	0.23	0.24	0.23	0.23
S%	22	27	37.5	71	33.3
BS	0.28	0.28	0.24	0.14	0.25
ROC	0.01	$<$ 0.001	$<$ 0.001	$<$ 0.001	0.014

to be valuable was higher (lower) the lesser (more) the growth and yield of the wheat crop was water-limited. In other words, the forecast skill needed to be higher at wetter sites compared to drier sites. Thus, high skill levels of S% =71% for seasonal and S% = 37.5% for FMA rainfall at Talagang could be useful for decision-making. However, the 'true' value of skilful but imperfect forecasts depends on a range of socio-economic and agro-ecological constraints that were not considered here (Hansen, 2002; Weaver *et al.*, 2013).

Potential benefits from applications of seasonal climate forecasting necessitate an understanding of the forecast information, including the forecast probabilities, format, accuracy, and skill (Nicholls, 2000). Thus, forecast users need to be 'climate-literate'. For example, a shift in the odds towards a greater chance for above median rainfall also means that there remains a chance for below median rainfall; forecasts can be skilful but 'inaccurate' (Figure 3). Furthermore, the analyses showed that there is spatial and temporal variation in forecast skill (Figure 3, Table 4), which forecast users would need to consider when using forecasts in decision-making. Here, the selected seasonal forecasts were out by between 42.5 and 122.1 mm as approximately indicted by the RMSE. This is about half to one-third of the average seasonal rainfall in Pothwar, where wheat is produced in regions delineated by the 100–350 mm isohyets of November to April rainfall (Table 1). This may compromise the usefulness of forecasts, which appeared to perform better during certain periods in the climatic records that were available from the study locations (Figure 3).

Temporal variations in forecast performance (Figure 3, Table 4) have been shown to be related to underlying climatic modulations that are not captured in statistical forecasting models (Power *et al.*, 1999). Another source for the temporal variation may be that the model cannot fully explain non-linear relationships between the RCs and rainfall. Comparison of models for the more skilful and less skilful periods may assist in determining whether a particular combination of RCs is responsible for the variability in model performance. Results of this further analysis may assist in refining forecast output to account for increased uncertainty in periods where forecast performance is reduced. In contrast to the variability in forecast skill at annual timescales, the relationship between SSTs and rainfall appeared to be stationary at intra-decadal or longer timescales (Figure 3(a) and (c)). This suggests that the forecasts may be able to capture general trends in the rainfall data as a result of climate variability at longer periodicities than annual.

The forecasts presented here are skilful but imperfect. As a consequence of such forecast uncertainty, climate sensitive management decisions do not always result in better outcomes (Stone and Meinke, 2005; Meinke *et al.*, 2007; Moeller *et al.*, 2008). Apart from discrepancies between forecast and observed rainfall amounts, other reasons for the possibility of poor outcomes include the variability in the timing and intensity of rainfall events relative to critical crop growth stages such as flowering (Fischer, 2011), temporal changes in productions costs and commodity prices, and how forecast users perceive the climate information provided (Weaver *et al.*, 2013). In fact, vulnerability can increase if agricultural decision makers perceive forecast information as an explicit and unambiguous signal (Meinke *et al.*, 2005). Weaver *et al.* (2013) argue that 'predict-then-act' frameworks have generally failed to inform decision-making meaningfully. Thus, for applications of the seasonal forecasts discussed here to be successful would require a reflective decision framework that engages stakeholders, and which considers socio-economic and agro-ecological constraints and their interactions with decision-making under climatic uncertainty in Pothwar.

5. Conclusions

Seasonal forecasts were developed for rainfall periods that coincide with important stages in the management, growth and development of wheat crops grown in rainfed environments of Pothwar. Strong correlations were found between the VARIMAX rotated principal components of SSTs (the climatic predictors) and well known SST anomalies associated with the El-Niño Southern Oscillation, Pacific Decadal Oscillation, Indian Ocean Dipole and the tropical Atlantic Ocean. The Generalized Additive Models revealed linear and non-linear relationships between agriculturally relevant rainfall periods and multiple time-lagged climatic predictors. Based on the statistical models developed, skilful forecasts of seasonal rainfall totals (continuous and categorical probabilistic forecasts) can be produced for Chakwal, Talagang and Islamabad. Forecasts such as the probability of exceeding median monsoon rainfall and volumetric rainfall with 95% confidence intervals can be provided with minimal financial investment alongside existing dynamical forecasts (http://www.pmd.gov.pk/rnd/rndweb/rnd_new/seasonal.php).

The relatively simple, statistical forecasts presented here can be easily implemented in the target region where specific, locally relevant forecasts are not readily available. A wide range of methods was employed to assess the true ability of the models to forecast rainfall. Such thorough evaluation is particularly important where the length of the available rainfall record is limited (i.e. about just sufficient for model development and some testing). While the addition of splines allows for more complex relationships between rainfall and SST to be modelled, it does require the user to either use less predictors in the model or limit use to longer datasets. As more seasonal rainfall data will become available, the models can be easily updated and improved. Skilful, but imperfect, forecasts can potentially assist decision-makers in managing climate related risks on-farm and the regional level. However, consideration of the value of the forecasts would require a research approach that engages with stakeholders and addresses agro-ecological and socio-economic constraints not included here.

Acknowledgements

We wish to thank the Higher Education Commission of Pakistan for the financial support provided, and two anonymous reviewers for their constructive comments.

Appendix

Table A1. Summary of climatic indices describing sea surface temperature (SST) anomalies related to various climate phenomena.

Index	Name	Region	Description	References
Nino1 + 2		10 to 0° S, 80–90° W	Mean monthly SST anomaly	Allan (2000)
Nino3		5° N to 5° S, 90–150° W	Mean monthly SST anomaly	
Nino3.4		5° N to 5° S, 120 to 170° W	Mean monthly SST anomaly	
Nino4		5° S to 5° N, 150° E to 160° W	Mean monthly SST anomaly	
IOD	Indian Ocean Dipole	Indian Ocean	Difference in SSTs between regions in western and south-eastern equatorial Indian Ocean	Saji *et al.* (1999)
TNA	Tropical Northern Atlantic index	5.5–23.5° N and 15–57.5° W	Mean monthly SST anomaly	Enfield *et al.* (1999)
TSA	Tropical Southern Atlantic index	Equator-20° S and 10 ° E to 30° W	Mean monthly SST anomaly	Enfield *et al.* (1999)
PDO	Pacific Decadal Oscillation	Pacific Ocean poleward of 20° N	Leading principal component of monthly SST anomalies in the North Pacific Ocean	Zhang *et al.* (1997) and Krishnan and Sugi (2003)

Table A2. Results for time-lagged relationships between VARIMAX rotated principal components (RC1, RC2 ... 10) of SSTs and six rainfall periods at three locations in Pothwar, Pakistan: p-values of the RC covariates selected in the Generalized Additive Models and values of forecast skill scores (estimates are rounded to two decimal places).

	p-values for covariates										Skill scores						
SST	RC1	RC2	RC3	RC4	RC5	RC6	RC7	RC8	RC9	RC10	r^2	RMSE	pBIAS	S%	LEPS	BS	ROC
Chakwal, NDJ period																	
J	0.00	0.00	0.12	0.17	NA	NA	NA	0.01	0.00	NA	0.00	57.79	13.00	0.08	0.04	0.40	0.63
A	0.00	0.00	NA	0.04	0.01	NA	0.12	NA	0.34	NA	0.04	81.55	0.10	−0.13	−0.09	0.42	0.84
S	0.07	0.17	NA	0.00	NA	NA	NA	NA	0.85	0.29	0.00	66.74	−6.10	0.08	0.00	0.35	0.50
JA	0.01	0.01	NA	0.35	0.05	NA	0.54	0.14	0.01	0.15	0.05	70.15	12.70	−0.33	−0.09	0.50	0.99
AS	0.01	0.00	NA	0.00	0.05	NA	0.17	NA	NA	NA	0.04	67.83	−4.50	−0.25	−0.12	0.42	0.83
JAS	0.00	0.00	0.27	NA	0.00	0.09	0.02	0.15	0.00	0.01	0.16	77.33	3.60	−0.29	−0.20	0.50	1.00
Chakwal, FMA period																	
J	0.32	0.00	0.01	0.00	0.01	0.00	0.00	0.01	0.02	0.00	0.01	149.45	−8.10	0.29	0.10	0.30	0.05
A	0.02	0.00	NA	NA	0.55	0.06	0.00	0.18	0.01	0.23	0.06	126.13	−6.40	−0.08	0.11	0.36	0.25
S	0.01	0.00	0.38	NA	NA	0.06	0.01	0.07	0.02	0.15	0.05	106.69	−5.40	0.08	0.14	0.36	0.23
JA	NA	0.00	NA	NA	NA	0.23	0.00	0.40	NA	0.01	0.00	133.14	−4.70	−0.17	−0.01	0.42	0.65
AS	0.02	0.00	NA	NA	NA	0.01	0.00	0.04	0.01	0.13	0.05	123.42	−4.80	0.00	0.11	0.36	0.23
JAS	0.01	0.00	NA	0.28	NA	0.07	0.02	0.47	0.12	NA	0.05	134.21	−7.00	0.00	0.10	0.38	0.25
Chakwal, NDJFMA period																	
J	0.07	0.03	0.03	0.30	0.13	0.06	0.00	0.01	0.01	NA	0.00	146.44	0.20	0.00	0.04	0.36	0.48
A	0.00	0.00	NA	NA	0.17	0.05	0.03	0.09	0.00	NA	0.02	171.14	−5.20	0.00	−0.02	0.37	0.33
S	0.01	0.00	NA	0.17	NA	0.05	0.03	0.09	0.00	0.17	0.03	111.91	−1.10	0.17	0.09	0.32	0.22
JA	0.05	0.00	NA	NA	NA	0.22	0.02	0.14	0.03	NA	0.01	153.59	−0.90	0.00	−0.07	0.39	0.79
AS	0.01	0.00	NA	0.34	0.10	0.05	0.16	0.14	0.01	0.14	0.01	144.57	−3.90	0.13	0.04	0.31	0.19
JAS	0.01	0.01	NA	NA	NA	0.04	0.23	0.29	0.05	NA	0.00	158.85	−4.20	−0.04	0.02	0.36	0.49
Chakwal, monsoon periods JAS, JA, and AS (from top)																	
AM	0.00	NA	NA	NA	NA	0.02	NA	NA	0.00	0.28	0.37	118.22	0.60	0.22	0.31	0.28	0.01
AM	0.01	0.01	0.00	NA	0.00	0.00	0.00	0.00	NA	NA	0.08	142.47	6.30	0.27	0.20	0.25	0.00
AM	0.07	NA	0.76	NA	NA	0.05	NA	0.22	0.00	0.15	0.20	107.27	1.80	0.18	0.21	0.29	0.06
Talagang, NDJ period																	
J	0.02	0.00	NA	0.05	0.09	0.30	0.23	0.02	0.01	0.23	0.00	31.06	6.40	0.13	0.06	0.36	0.38
A	0.01	0.00	NA	0.01	0.31	NA	0.60	NA	0.05	NA	0.00	32.33	3.10	−0.29	0.02	0.41	0.80
S	0.03	0.08	NA	0.00	NA	NA	NA	NA	0.09	0.36	0.00	36.15	−7.40	−0.17	−0.03	0.41	0.88
JA	0.03	0.00	NA	0.03	0.24	0.38	0.34	0.16	0.03	NA	0.00	31.75	6.30	−0.04	0.00	0.41	0.76
AS	0.01	0.02	NA	0.00	0.24	0.41	0.54	NA	0.12	NA	0.00	26.51	−0.30	−0.25	0.02	0.43	0.85
JAS	0.02	0.03	0.33	0.01	NA	NA	NA	NA	0.03	NA	0.07	34.51	2.50	−0.33	−0.17	0.51	1.00
Talagang, FMA period																	
J	0.06	0.00	NA	0.53	NA	0.08	0.00	NA	0.12	NA	0.02	52.86	1.40	0.00	0.02	0.37	0.37
A	0.02	0.00	NA	NA	0.09	0.02	0.00	NA	0.01	0.18	0.16	53.71	−6.90	0.33	0.25	0.25	0.00
S	0.00	0.00	0.20	NA	NA	0.10	0.00	0.27	0.00	0.03	0.21	42.51	−2.50	0.38	0.24	0.24	0.00
JA	0.08	0.00	NA	NA	NA	0.05	0.01	NA	0.03	0.22	0.09	52.02	−2.60	0.29	0.13	0.27	0.01
AS	0.02	0.00	NA	0.34	NA	0.03	0.01	NA	0.00	0.10	0.16	52.89	−5.50	0.42	0.23	0.25	0.00
JAS	0.02	0.00	NA	NA	NA	0.01	0.01	0.35	0.03	0.04	0.09	62.40	−6.80	0.29	0.16	0.27	0.01
Talagang, NDJFMA period																	
J	0.05	0.01	NA	NA	NA	0.04	0.01	NA	0.02	NA	0.02	53.65	2.10	0.13	0.03	0.29	0.07
A	0.01	0.00	NA	NA	NA	0.00	0.02	NA	0.00	0.18	0.08	61.80	−6.10	0.38	0.14	0.24	0.00
S	0.00	0.00	0.20	0.28	0.31	0.02	0.00	0.11	0.00	0.04	0.16	49.85	−0.90	0.58	0.20	0.15	0.00
JA	0.03	0.01	NA	NA	NA	0.01	0.07	NA	0.01	0.11	0.10	53.16	−1.70	0.50	0.13	0.20	0.00
AS	0.00	0.00	NA	NA	NA	0.00	0.02	0.25	0.00	0.04	0.19	51.43	−3.00	0.71	0.23	0.14	0.00
JAS	0.00	0.00	NA	NA	NA	0.00	0.01	0.80	0.00	0.08	0.09	59.19	−0.10	0.63	0.17	0.16	0.00
Talagang, monsoon periods JAS, JA, and AS (from top)																	
AM	0.01	NA	NA	NA	NA	NA	NA	NA	0.00	0.04	0.38	57.71	−0.20	0.27	0.23	0.28	0.00
AM	0.09	0.62	NA	NA	NA	0.55	NA	NA	0.04	NA	0.02	68.84	−1.40	0.10	0.10	0.31	0.12
AM	0.09	NA	NA	0.25	0.46	NA	NA	0.37	0.00	0.01	0.16	58.42	−0.20	0.10	0.15	0.34	0.22
Islamabad, NDJ period																	
J	0.00	0.05	0.08	0.00	0.15	0.09	0.04	0.01	NA	0.00	0.01	88.25	4.70	−0.13	−0.04	0.42	0.58
A	0.07	NA	NA	0.20	NA	NA	0.34	NA	NA	0.08	0.12	120.19	3.10	−0.17	−0.11	0.43	0.90
S	0.28	NA	NA	0.08	NA	0.11	0.11	0.08	0.43	0.30	0.02	86.14	−0.60	−0.08	−0.01	0.35	0.53
JA	0.00	NA	0.12	0.08	NA	0.02	0.32	NA	0.01	0.00	0.00	96.49	−1.50	−0.17	−0.03	0.42	0.65
AS	0.04	NA	NA	0.16	0.74	NA	0.31	NA	NA	0.07	0.00	72.52	−1.50	−0.21	−0.04	0.42	0.86
JAS	0.02	NA	NA	0.10	NA	0.20	NA	NA	NA	0.01	0.02	77.88	−4.40	0.04	0.05	0.38	0.50

Table A2. Continued

	p-values for covariates										Skill scores						
SST	RC1	RC2	RC3	RC4	RC5	RC6	RC7	RC8	RC9	RC10	r^2	RMSE	pBIAS	S%	LEPS	BS	ROC
Islamabad, FMA period																	
J	NA	0.22	0.00	NA	0.09	0.00	0.08	0.00	0.10	NA	0.05	158.83	0.90	0.13	0.12	0.33	0.26
A	0.41	0.01	0.04	NA	0.01	0.11	0.03	0.00	0.08	0.15	0.02	183.74	14.00	0.25	0.20	0.31	0.13
S	NA	0.02	0.08	NA	NA	0.37	0.06	0.06	0.00	NA	0.16	122.10	−4.90	0.33	0.23	0.25	0.01
JA	NA	0.03	0.01	0.21	0.18	NA	0.03	0.00	0.02	NA	0.01	195.42	−1.60	0.17	0.14	0.31	0.09
AS	NA	0.00	0.32	NA	NA	0.10	0.00	0.00	0.00	NA	0.12	127.02	−2.40	0.21	0.23	0.28	0.04
JAS	NA	0.01	0.05	NA	0.23	NA	0.07	0.00	0.02	NA	0.09	154.62	1.70	0.29	0.21	0.27	0.02
Islamabad, NDJFMA period																	
J	0.16	0.25	0.09	NA	NA	0.01	0.11	0.01	0.06	NA	0.00	197.67	1.70	−0.04	−0.09	0.40	0.72
A	0.37	0.06	NA	NA	0.12	0.07	0.05	0.05	0.12	0.03	0.00	218.02	1.90	0.00	0.01	0.38	0.76
S	NA	0.09	0.20	NA	NA	0.18	0.16	0.25	0.02	NA	0.00	167.07	2.10	0.00	−0.02	0.36	0.56
JA	0.24	0.05	NA	NA	0.44	0.15	0.05	0.04	0.10	0.36	0.01	254.21	−5.50	0.04	−0.03	0.38	0.69
AS	0.23	0.07	NA	NA	NA	0.32	0.05	0.12	0.02	NA	0.02	165.68	−2.00	−0.08	−0.03	0.43	0.83
JAS	0.30	0.06	NA	NA	NA	0.33	0.11	0.15	0.01	NA	0.00	223.59	2.70	−0.17	−0.09	0.44	0.86
Islamabad, monsoon periods JAS, JA, and AS (from top)																	
AM	NA	0.10	NA	0.15	0.30	0.36	0.74	NA	0.46	NA	0.05	267.05	−0.20	−0.27	−0.14	0.48	1.00
AM	NA	0.03	NA	0.07	0.08	0.23	0.09	NA	NA	NA	0.03	201.70	−1.80	0.35	0.14	0.25	0.02
AM	NA	NA	NA	NA	0.22	0.25	NA	NA	NA	NA	0.11	177.28	−1.10	−0.18	−0.06	0.35	0.80

Individual months in the forecast period are denoted by their first letter.

References

Allan RJ. 2000. ENSO and climatic variability in the past 150 years. In *El Niño and the Southern Oscillation: Multiscale Variability and its Impact on Natural Ecosystems and Society*, Diaz HF, Markgraf V (eds). Cambridge University Press: Cambridge; 3–55.

Allan R, Chambers D, Drosdowsky W, Hendon H, Latif M, Nicholls N, Smith I, Stone R, Tourre Y. 2001. Is there an Indian Ocean dipole, and is it independent of the El Niño – Southern Oscillation? *CLIVAR Exchanges* **6**(3): 18–22.

Arif M, Malik MA. 2009. Economic feasibility of proposed cropping patterns under different soil moisture regimes of the Pothwar plateau. *Int. J. Agric. Biol.* **11**: 27–32.

Ashok K, Guan ZY, Saji NH, Yamagata T. 2004. Individual and combined influences of ENSO and the Indian Ocean Dipole on the Indian summer monsoon. *J. Clim.* **17**: 3141–3155.

Ashok K, Saji NH. 2007. On the impacts of ENSO and Indian Ocean dipole events on sub-regional Indian summer monsoon rainfall. *Nat. Hazards* **42**: 273–285.

Barlow M, Cullen H, Lyon B. 2002. Drought in central and southwest Asia: La Niña, the warm pool, and Indian Ocean precipitation. *J. Clim.* **15**: 697–700.

Barlow M, Nigam S, Berbery EH. 2001. ENSO, Pacific decadal variability, and US summertime precipitation, drought, and stream flow. *J. Clim.* **14**: 2105–2128.

Boulanger JP, Leloup J, Penalba O, Rusticucci M, Lafon F, Vargas M. 2005. Observed precipitation in the Paraná-Plata hydrological basin: long-term trends, extreme conditions and ENSO teleconnections. *Clim. Dynam.* **24**: 393–413.

Bracco A, Kucharski F, Molteni F, Hazeleger W, Severijns C. 2005. Internal and forced modes of variability in the Indian Ocean. *Geophys. Res. Lett.* **32**: L12707, DOI: 12710.11029/12005GL02 3154.

Byerlee D, Husain T. 1993. Agricultural research strategies for favored and marginal areas - the experience of farming systems research in Pakistan. *Exp. Agric.* **29**: 155–171.

Cane M. 2000. Understanding and predicting the world's climate system. In *Applications of Seasonal Climate Forecasting in Agricultural and Natural Ecosystems – The Australian Experience*, Hammer GL, Nicholls N, Mitchell C (eds). Kluwer Academic Publishers: Dordrecht; 29–50.

Chang C, Harr P, Ju J. 2001. Possible roles of Atlantic circulations on the weakening Indian monsoon rainfall-ENSO relationship. *J. Clim.* **14**: 2376–2380.

Chang P, Yamagata T, Schopf P, Behera SK, Carton J, Kessler WS, Meyers G, Qu T, Schott F, Shetye S, Xie SP. 2006. Climate fluctuations of tropical coupled systems - the role of ocean dynamics. *J. Clim.* **19**: 5122–5174.

Clark CO, Cole JE, Webster PJ. 2000. Indian Ocean SST and Indian summer rainfall: predictive relationships and their decadal variability. *J. Clim.* **13**: 2503–2519.

Copas JB. 1983. Regression, prediction and shrinkage. *J. R. Stat. Soc., Ser. B* **45**: 311–354.

D'Arrigo R, Villalba R, Wiles G. 2001. Tree-ring estimates of Pacific decadal climate variability. *Clim. Dyn.* **18**: 219–224.

D'Arrigo R, Wilson R. 2006. On the Asian expression of the PDO. *Int. J. Climatol.* **26**: 1607–1617.

Dommenget D, Latif M. 2002. A cautionary note on the interpretation of EOFs. *J. Clim.* **15**: 216–225.

Drosdowsky W, Chambers LE. 2001. Near-global sea surface temperature anomalies as predictors of Australian seasonal rainfall. *J. Clim.* **14**: 1677–1687.

Enfield DB, Mestas AM, Mayer DA, Cid-Serrano L. 1999. How ubiquitous is the dipole relationship in tropical Atlantic sea surface temperatures? *J. Geophys. Res.* **104**: 7841–7848.

Everingham YL, Muchow RC, Stone RC, Inman-Bamber NG, Singels A, Bezuidenhout CN. 2002. Enhanced risk management and decision-making capability across the sugarcane industry value chain on seasonal climate forecasts. *Agric. Syst.* **74**: 459–477.

Fischer RA. 2011. Wheat physiology: a review of recent developments. *Crop Pasture Sci.* **62**: 95–114.

Fischer G, van Velthuizen H, Shah M, Nachtergaele F. 2002. *Global Agro-ecological Assessment for Agriculture in the 21st Century: Methodology and Results*. International Institute for Applied Systems Analysis, Food and Agriculture Organization of the United Nations: Laxenburg.

Goddard L, Mason SJ, Zebiak SE, Ropelewski CF, Basher R, Cane MA. 2001. Current approaches to seasonal-to-interannual climate predictions. *Int. J. Climatol.* **21**: 1111–1152.

Goswami BN, Madhusoodanan M, Neema C, Sengupta D. 2006. A physical mechanism for North Atlantic SST influence on the Indian summer monsoon. *Geophys. Res. Lett.* **33**: L02706_02701–L02706_02704.

Graham MH. 2003. Confronting multicollinearity in ecological multiple regression. *Ecology* **84**: 2809–2815.

Hannachi A, Jolliffe IT, Stephenson DB. 2007. Empirical orthogonal functions and related techniques in atmospheric science: a review. *Int. J. Climatol.* **27**: 1119–1152.

Hansen JW. 2002. Realizing the potential benefits of climate prediction to agriculture: issues, approaches, challenges. *Agric. Syst.* **74**: 309–330.

Hansen JW. 2005. Integrating seasonal climate prediction and agricultural models for insights into agricultural practice. *Philos. Trans. R. Soc. London, Ser. B* **360**: 2037–2047.

Hansen JW, Indeje M. 2004. Linking dynamic seasonal climate forecasts with crop simulation for maize yield prediction in semi-arid Kenya. *Agric. Forest. Meteorol.* **125**: 143–157.

International Research Institute for Climate and Society. 2012a. IRI/LDEO climate data library. Kaplan Extended v2 ssta: SST anomaly data. http://iridl.ldeo.columbia.edu/SOURCES/.KAPLAN/.EXTENDED/.v2/.ssta/.

International Research Institute for Climate and Society. 2012b. IRI/LDEO climate data library. Statistical analysis tutorial. http://iridl.ldeo.columbia.edu/dochelp/StatTutorial/SVD/ (accessed 9 July 2013).

Jolliffe IT. 1987. Rotation of principal components: some comments. *J. Climatol.* **7**: 509–512.

Jolliffe IT. 2002. *Principal Component Analysis*. Springer: New York, NY.

Jolliffe IT, Stephenson DB. 2003. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley & Sons Ltd.: West Sussex.

Kaplan A, Cane M, Kushnir Y, Clement A, Blumenthal M, Rajagopalan B. 1998. Analyses of global sea surface temperature 1856–1991. *J. Geophys. Res.* **103**: 567–589.

Khan MA, Shah SAA. 2010. Food insecurity in Pakistan: causes and policy response. *J. Agric. Environ. Ethics* , DOI: 10.1007/s10806-010-9274-2.

Krishnan R, Sugi M. 2003. Pacific decadal oscillation and variability of the Indian summer monsoon rainfall. *Clim. Dyn.* **21**: 233–242.

Kucharski F, Bracco A, Yoo J, Molteni F. 2007. Low-frequency variability of the Indian monsoon-ENSO relationship and the tropical Atlantic: the "weakening" of the 1980s and 1990s. *J. Clim.* **20**: 4255–4266.

Kucharski F, Bracco A, Yoo JH, Molteni F. 2008. Atlantic forced component of the Indian monsoon interannual variability. *Geophys. Res. Lett.* **35**: L04076, DOI: 10.1029/2007GL033037.

Kumar KK, Rajagopalan B, Hoerling M, Bates G, Cane M. 2006. Unraveling the mystery of Indian monsoon failure during El Nino. *Science* **314**: 115–119.

Li T, Zhang YS, Chang CP, Wang B. 2001. On the relationship between Indian Ocean sea surface temperature and Asian summer monsoon. *Geophys. Res. Lett.* **28**: 2843–2846.

Lu R, Dong B, Ding H. 2006. Impact of the Atlantic Multidecadal Oscillation on the Asian summer monsoon. *Geophys. Res. Lett.* **33**: L24701.

Meinke H, Nelson R, Kokic P, Stone R, Ramasamy S, Baethgen W, Selvaraju R. 2007. Actionable climate knowledge: from analysis to synthesis. *Clim. Res.* **33**: 101–110.

Meinke H, deVoil P, Hammer GL, Power S, Allan R, Stone RC, Folland C, Potgieter A. 2005. Rainfall variability at decadal and longer time scales: signal or noise? *J. Clim.* **18**: 89–96.

Meyers G, McIntosh P, Pigot L, Pook M. 2007. The years of El Nino, La Nina, and interactions with the tropical Indian ocean. *J. Clim.* **20**: 2872–2880.

Moeller C, Asseng S, Berger J, Milroy SP. 2009. Plant available soil water at sowing in Mediterranean environments – Is it a useful criterion to aid nitrogen fertiliser and sowing decisions? *Field Crop Res.* **114**: 127–136.

Moeller C, Smith I, Asseng S, Ludwig F, Telcik N. 2008. The potential value of seasonal forecasts of rainfall categories – case studies from the wheatbelt in Western Australia's Mediterranean region. *Agric. Forest Meteorol.* **148**: 606–618, DOI: 610.1016/j.agrformet.2007.1011.1004.

National Oceanic and Atmospheric Administration. (2012). Climate indices: monthly atmospheric and ocean time series. http://www.esrl.noaa.gov/psd/data/climateindices/list/.

NCAR - Research Application Program. 2010. Verification: forecast verification utilities, R package version 1.31 ed. http://CRAN.R-project.org/package=verification.

Newman M, Compo GP, Alexander MA. 2003. ENSO-forced variability of the Pacific Decadal Oscillation. *J. Clim.* **16**: 3853–3857.

Nicholls N. 2000. Opportunities to improve the use of seasonal climate forecasts. In *Applications of Seasonal Climate Forecasting in Agricultural and Natural Ecosystems The Australian Experience*, Hammer GL, Nicholls N, Mitchell C (eds). Kluwer Academic Publishers: Dordrecht; 309–327.

van Ogtrop FF, Vervoort RW, Heller G, Stasinopoulos D, Rigby R. 2011. Long-range forecasting of intermittent streamflow. *Hydrol. Earth Syst. Sci.* **15**: 3343–3354.

Passioura JB, Angus JF. 2010. Improving productivity of crops in water-limited environments. *Adv. Agron.* **106**: 37–75.

Power S, Casey T, Folland C, Colman A, Mehta V. 1999. Inter-decadal modulation of the impact of ENSO on Australia. *Clim. Dyn.* **15**: 319–324.

Power S, Colman R. 2006. Multi-year predictability in a coupled general circulation model. *Clim. Dyn.* **26**: 247–272.

Punjab Barani Commission. 1976. *Report of the Punjab Barani Commission*. Government Punjab: Lahore.

R Development Core Team. 2011. R: a language and environment for statistical computing. http://www.R-project.org (accessed 9 July 2013).

Revelle W. 2011. psych: Procedures for Personality and Psychological Research Northwestern University, Evanston. http://personality-project.org/r/psych.manual.pdf.

Reynolds RW, Smith TM. 1994. Improved global sea surface temperature analyses. *J. Clim.* **7**: 929–948.

Richman MB. 1986. Rotation of principal components. *J. Climatol.* **6**: 293–335.

Richman MB. 1987. Rotation of principal components: a reply. *J. Climatol.* **7**: 511–520.

Saji NH, Goswami BN, Vinayachandran PN, Yamagata T. 1999. A dipole mode in the tropical Indian Ocean. *Nature* **401**: 360–363.

Saji NH, Yamagata T. 2003. Possible impacts of Indian Ocean Dipole mode events on global climate. *Climate Res.* **25**: 151–169.

Schott FA, Xie SP, McCreary JP. 2009. Indian Ocean circulation and climate variability. *Rev. Geophys.* **47**: RG1002.

Smith TM, Reynolds RW, Peterson TC, Lawrimore J. 2008. Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006). *J. Clim.* **21**: 2283–2296.

Stewart JI. 1991. Principals and performance of response farming. In *Climatic Risk in Crop Production: Models and Management for the Semiarid Tropics and Subtropics*, Muchow RC, Bellamy JA (eds). CAB International: Wallingford; 361–382.

Stone RC, Meinke H. 2005. Operational seasonal forecasting of crop performance. *Philos. Trans. R. Soc. Ser. B: Biol. Sci.* **360**: 2109–2124.

Suleman S, Wood MK, Shah BH, Murray L. 1995. Development of a rainwater harvesting system for increasing soil moisture in arid rangelands of Pakistan. *J. Arid Environ.* **31**: 471–481.

Terray P, Dominiak S. 2005. Indian Ocean sea surface temperature and El Niño-Southern Oscillation: a new perspective. *J. Clim.* **18**: 1351–1368.

Ummenhofer CC, Gupta AS, Li Y, Taschetto AS, England MH. 2011. Multi-decadal modulation of the El Niño – Indian monsoon relationship by Indian Ocean variability. *Environ. Res. Lett.* **6**: 1–8, DOI: 10.1088/1748-9326/6/034006.

Weaver CP, Lempert RJ, Brown C, Hall JA, Revell D, Sarewitz D. 2013. Improving the contribution of climate model information to decision making: the value and demands of robust decision frameworks. *Wiley Interdiscipl. Rev.-Clim. Change* **4**: 39–60.

Westra S, Brown C, Lall U, Koch I, Sharma A. 2010. Interpreting variability in global SST data using independent component analysis and principal component analysis. *Int. J. Climatol.* **30**: 333–346.

Westra S, Sharma A. 2010. An upper limit to seasonal rainfall predictability? *J. Clim.* **23**: 3332–3351.

Wilks DS. 2011. *Statistical Methods in the Atmospheric Sciences*. Elsevier: Oxford.

Wood SN. 2003. Thin plate regression splines. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **65**: 95–114.

Wood SN. 2006. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC: Boca Raton, FL.

Wood SN. 2008. Fast stable direct fitting and smoothness selection for generalized additive models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **70**: 495–518.

Zambrano-Bigiarini M. 2011. hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series. R package version 0.3-2 ed. http://CRAN.R-project.org/package=hydroGOF.

Zhang R, Delworth TL. 2006. Impact of Atlantic multidecadal oscillations on India/Sahel rainfall and Atlantic hurricanes. *Geophys. Res. Lett.* **33**: L17712.

Zhang Y, Wallace JM, Battisti DS. 1997. ENSO-like interdecadal variability: 1900–1993. *J. Clim.* **10**: 1004–1020.