Seasonal precipitation forecast skill over Iran

Amin Shirvani[a,b*] and Willem A Landman[b,c]

[a] Department of Water Engineering, Oceanic and Atmospheric Research Center, College of Agriculture, Shiraz University, Shiraz, Iran.
[b] Department of Geography, Geoinformatics and Meteorology, University of Pretoria, Pretoria, South Africa.
[c] Council for Scientific and Industrial Research, Natural Resources and the Environment.

Abstract

This paper examines the skill of seasonal precipitation forecasts over Iran using one two-tiered model, three National Multi-Model Ensemble (NMME) models, and two coupled ocean-atmosphere or one-tiered models. These models are, respectively, the ECHAM4.5 atmospheric model that is forced with sea surface temperature (SST) anomalies forecasted using constructed analogue SSTs (ECHAM4.5-SSTCA); the IRI-ECHAM4.5-DirectCoupled, the NASA-GMAO-062012 and the NCEP-CFSv2; and the ECHAM4.5 Modular Ocean Model version 3 (ECHAM4.5-MOM3-DC2) and the ECHAM4.5-GML-NCEP Coupled Forecast System (CFSSST). The precipitation and 850 hPa geopotential height fields of the forecast models are statistically downscaling to the $0.5^\circ \times 0.5^\circ$ spatial resolution of the Global Precipitation Climatology Centre (GPCC) Version 6 gridded precipitation data, using model output statistics (MOS) developed through the canonical correlation analysis (CCA) option of the Climate Predictability Tool (CPT). Retroactive validations for lead times of up to three months are performed using the relative operating characteristic (ROC) and reliability diagrams, which are evaluated for above- and below-normal categories and defined by the upper and lower 75^{th} and 25^{th} percentiles of the data record over the 15-year test period of 1995/96 to 2009/10. The forecast models' skills are also compared with skills obtained by (a) downscaling simulations produced by forcing the ECHAM4.5 with simultaneously observed SST, and (b) the 850 hPa geopotential height NCEP-NCAR reanalysis data. Downscaling forecasts from most models generally produce the highest skill forecast at lead times of up to three months for autumn precipitation – the October-

* Corresponding author's email addresses: am_shirvani@hotmail.com; ashirvani@shirazu.ac.ir.

November-December (OND) season. For most seasons, a high skill is obtained from ECHAM4.5-MOM3-DC2 forecasts at a one-month lead time when the models' 850 hPa geopotential height fields are used as the predictor fields. For this model and lead time, the Pearson correlation between the area-averaged of the observed and forecasts over the study area for the OND, November-December-January (NDJ), December-January-February (DJF) and January-February-March (JFM) seasons were 0.68, 0.62, 0.42 and 0.43, respectively.

KEY WORDS: statistical downscaling, seasonal forecasting, GCMs, Iran.

1. Introduction

Seasonal forecasts are possible over certain areas of the globe as the seasonal-to-interannual variability of such regions is related to predictable components of the climate system such as the El Niño-Southern Oscillation (ENSO). ENSO is the climate mode that contributes the most to seasonal forecasts skill (Goddard and Dilley, 2005; Balmaseda and Anderson, 2009; Weisheimer *et al*., 2009). However, other parts of the globe, such as the Middle East, are not as strongly associated with the ENSO phenomenon (Ropelewski and Halpert, 1987; Ropelewski and Halpert, 1989). Notwithstanding, Nazemosadat and Coredy (2000) have documented an ENSO signal in Iran during the OND season, and Barlow *et al*. (2002) reported that severe droughts that occurred in central south Asia from 1998 to 2001 are related to a combination of the La Niña and the unusually warm SSTs in the west Pacific, which may have enhanced the regional dynamics of the warm pool. Nazemosadat and Ghasemi (2004) reported that warm ENSO phases reduce the intensity and probability of drought while the cold ENSO phases increase the intensity and probability of severe drought during OND, particularly over the southern parts of Iran. However, the response of the boreal winter precipitation to ENSO events is generally weak. The linear relationship between ENSO and seasonal precipitation is weak over most parts of Iran. Pourasghar *et al*. (2012) reported that the interannual precipitation variability in southern parts of Iran in autumn and early winter (late winter) is correlated with the Indian Ocean Dipole (the Mediterranean Sea). These studies suggest that this region of Iran is affected by a complicated combination of ENSO events and SSTs over the west Pacific and the Indian Ocean. Fallah-Ghalhary *et al*.

(2010) reported that adaptive neuro-fuzzy inference system (ANFIS) can predict spring rainfall over the northeast region of Iran with reasonable accuracy.

As a result of the complex associations between the Iranian seasonal rainfall variability and the global SST patterns, we will investigate the predictability of seasonal rainfall over the region by using state-of-the-art general circulation models (GCMs). GCMs have already been used extensively in seasonal forecasting globally (Barnston *et al*., 2003; Bartman *et al*., 2003; Tippett *et al*., 2003; Barnston *et al*., 2010; Barnston and Mason, 2011; Landman *et al*., 2012; Landman and Beraki, 2012; Doblas-Reyes *et al*., 2013; Landman, 2014; Sachindra *et al*., 2014a; Sachindra *et al*., 2014b). Such models have also been proven useful for Iran (Tippett *et al*., 2003; Tippett *et al*., 2005), but here we will expand on GCM modelling over the region by considering a wide range of dynamical forecasting systems.

GCMs do not tend to skillfully predict central southwest Asia's winter precipitation (Tippett *et al*., 2003). However, statistical correction of GCM output have improved forecasts in the region where the borders of Afghanistan, Pakistan and Tajikistan meet, but not much skill has been found along the southwest border of Iran (Tippett *et al*., 2003). Moreover, precipitation by the ECHAM4.5 GCM (Roeckner *et al*., 1996) had only produced low skill simulations over central southwest Asia, but when observed upper-level winds were used as predictors in a statistical model, the results were improved (Tippett *et al*., 2003).

Iran is an area subject to climatic extremes with high societal impact. For example, in 1999 the severe drought in Iran affected over 37 million people and the flooding in 2001 and 2002 affected about 1.5 million people (http://www.emdat.be/). These disasters were the cause of significant financial losses. However, very little evidence has thus far been presented in the literature on the possibility of predicting climatic extremes on a seasonal time scale over this region.

The effect of topography and lakes, which are important to local weather, may not be included in the GCMs. The role of the Middle East topography and neighbouring seas, – including the Persian Gulf and the Mediterranean, Black, Red and Caspian Seas, – are important to seasonal precipitation of the Middle East. Statistical methods are able to improve the accuracy and usefulness of GCMs simulations. In this research, the state-of-the-art atmospheric and ocean-atmosphere coupled models are employed to study the seasonal precipitation predictability over Iran. The forecasts from these models are empirically

improved by downscaling the large-scale forecasts from these models to gridded precipitation fields through MOS (Glahn and Lowry, 1972; Wilks, 2011). Two fields of the GCM forecasts, including precipitation and 850 hPa geopotential heights, are used for precipitation prediction for the seasons in which significant precipitation occurs over Iran. Advantages in statistically post-processing forecasts from GCMs and coupled GCMs are subsequently demonstrated. The NMME and coupled ocean-atmosphere models have not previously been considered in modelling studies over Iran.

2. Data, models and methods

2.1. Precipitation data

The GPCC Version 6 gridded monthly precipitation data with a spatial resolution of $0.5^\circ \times 0.5^\circ$ (Schneider *et al.*, 2011; http://www.esrl.noaa.gov/psd/data/gridded/data.gpcc.html) are used to calculate three-month seasonal precipitation totals for Iran (Figure 1). The three-month rainy seasons include OND, NDJ, etc. through to the February-March-April (FMA) season, which cover autumn, winter and spring. About 85% of the annual precipitation over the study area (Lat: $23.75^\circ \text{N to} 40.25^\circ \text{N}$, Long: $44.25^\circ \text{E to} 63.75^\circ \text{E}$) occurs from October to April (calculated over the period from 1982 to 2010 using GPCC v6 precipitation data). Figure 1 shows the climatological mean of the total precipitation over the study area from October to April. The seasonal mean ranges from 100mm in the central and eastern parts to 800mm in the northern parts. Although GPCC v6 precipitation data are available from January 1901 to December 2010, the period of 1982/83 to 2009/10 is considered throughout this research since this period is the common re-forecast period for all global models to be tested. All the data sets in this research were extracted from the data library of the International Research Institute for Climate and Society (IRI; http://iridl.ldeo.columbia.edu/).

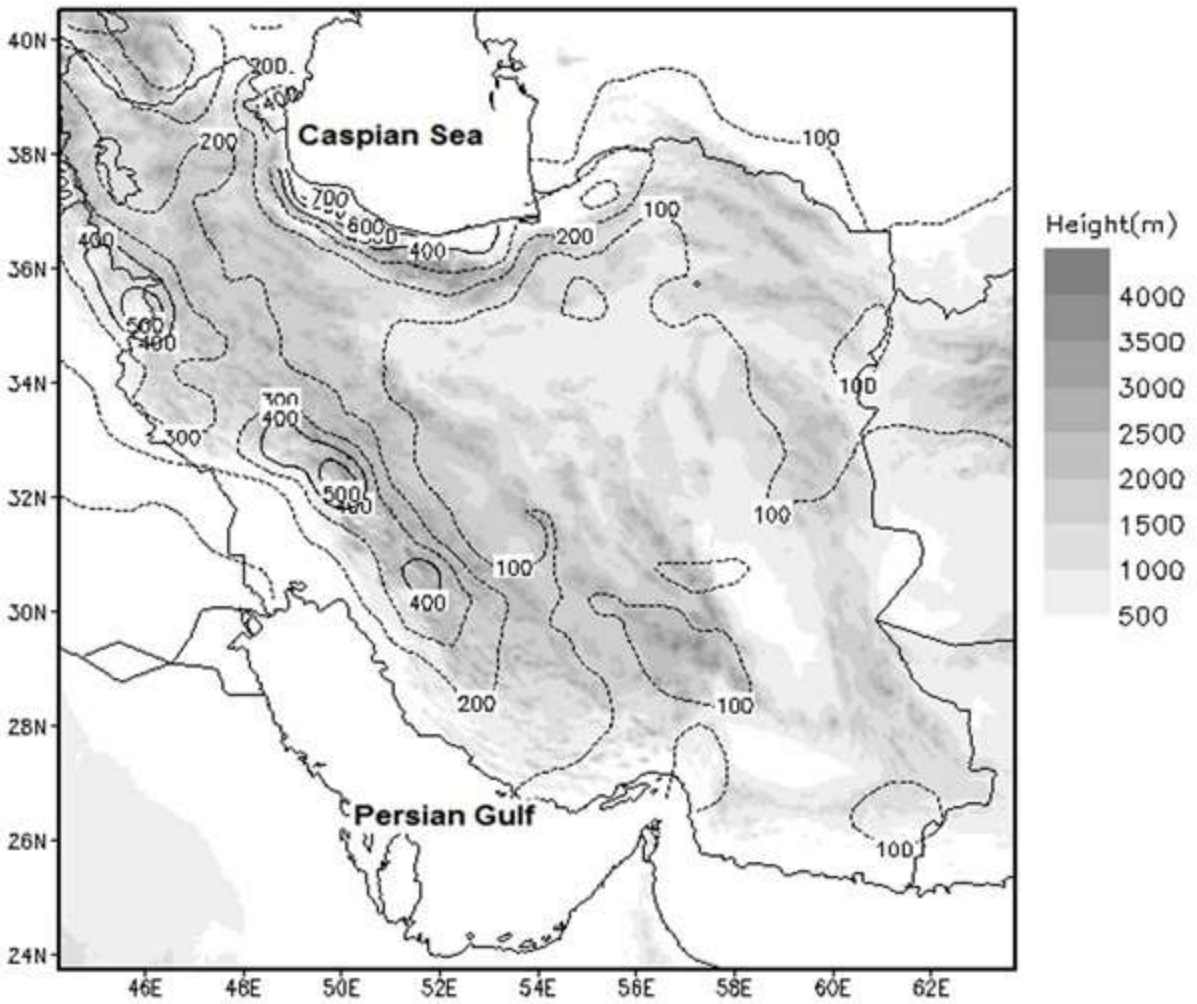


Fig. 1. Topography map of Iran and contour plot of the climatological mean of October to April total precipitation for the period 1982-2010. The GPCC precipitation data were used to calculate the climatological mean. The contours were plotted only where they were more than 100mm.

2.2. NCEP-NCAR reanalysis data

The 850 hPa geopotential height data representing the observed low-level atmospheric large-scale circulation over the study area are extracted from NCEP-NCAR reanalysis data (Kalnay *et al*., 1996; http://www.cpc.ncep.noaa.gov/products/wesley/reanalysis.html). The relationships between observed circulation and precipitation fields over the study area is assessed and tested against the simulated circulation of the global models in order to form an improved understanding of the physical processes driving observed and predicted seasonal variability and how the models are able to capture these processes. The spatial resolution of NCEP-NCAR data is 2.5° x 2.5°.

2.3. Atmospheric general circulation data

Two re-forecast or hindcast sets of the ECHAM4.5 (Roeckner *et al*., 1996) atmospheric general circulation model are used in this research. The first data set, available from January 1950 to the present, is produced by forcing the ECHAM4.5 with simultaneously observed SST and consists of 24 ensemble members. The second data set, available from January 1957 to Dec 2012, is produced by forcing the model with SST anomalies that are forecast using constructed analogue SSTs, denoted ECHAM4.5-SSTCA (Van den Dool, 2007), and consists of 24 ensemble members. The ECHAM4.5-SSTCA is a two-tiered model.

2.4. The hindcast data of the National Multi-Model Ensemble

The considered models from the NMME are the IRI-ECHAM4.5-DirectCoupled, NASA-GMAO-062012 and NCEP-CFSv2 (National Research Council, 2010; Kirtman *et al*., 2014; http://www.cpc.ncep.noaa.gov/products/NMME/). Only the hindcast precipitation data of these models are extracted for the period of 1982/83 to 2009/10.

2.5. Coupled general circulation models

The coupled ocean-atmospheric GCMs or one-tiered models to be considered are the ECHAM4.5 Modular Ocean Model version 3 (Pacanowski and Griffies, 1998) directly coupled to the ECHAM4.5 (denoted ECHAM4.5-MOM3-DC2; DeWitt, 2005) with archived hindcasts available from January 1982 to July 2012, and a slab mixed layer (denoted ECHAM4.5-GML-CFSSST) for the tropical western Pacific, Indian and Atlantic Oceans with prescribed SST forecasts from the National Centers for Environmental Prediction's (NCEP) Coupled Forecast System (CFS; Saha *et al*., 2006) in the central and eastern tropical Pacific, with hindcasts available from January 1982 to October 2012. These coupled model forecast sets each consist of 12 ensemble members.

For the ECHAM4.5 forced with predicted SST, the three NMME models and ECHAM4.5-GML-CFSSST, three forecast lead times are to be considered for model initialisation near the beginning of the month. For a one-month lead time, there are three weeks from the issuance of the forecast to the beginning of the forecast season. For example, a one-month lead-time forecast for OND is produced in the first week of August and two- and three-month lead-time

forecasts are produced at the beginning July and June, respectively. For the ECHAM4.5-MOM3-DC2 system used here, there are at least four weeks between the production of the forecast and the first month of the forecast season. For example, OND forecasts at a one-month lead time are produced at the end of July, two-month lead-time forecasts at the end of June, and three-month lead-time forecasts at the end of May.

The precipitation and 850 hPa geopotential height data sets of the coupled models, ECHAM4.5 and ECHAM4.5-SSTCA, are separately considered as the predictor field to construct retroactive downscaled forecasts (Landman *et al*., 2012).

2.6. Model output statistics

In this research, the method of MOS (Wilks, 2011) is used to derive statistical equations between global model hindcasts and precipitation over Iran. The MOS approach reduces model biases and corrects global model hindcasts towards more probable outcomes. The GPCC data set over the study area (Lat.: $23.75^{\circ}\,\mathrm{N}$ to $40.25^{\circ}\,\mathrm{N}$, Long.: $44.25^{\circ}\,\mathrm{E}$ to $63.75^{\circ}\,\mathrm{E}$) is considered as the predictand field in the MOS. This area is also considered as the predictor field when simulated or forecasted precipitation from the models is used in the MOS. However, a prediction area between $12^{\circ}\,\mathrm{N}$ and $60^{\circ}\,\mathrm{N}$ and from $22^{\circ}\,\mathrm{E}$ to $85^{\circ}\,\mathrm{E}$ is considered when observed, simulated or forecasted 850 hPa geopotential height is used. The domain for 850 hPa geopotential height is selected according to previous works such as that of Raziei *et al*. (2012, and references therein). CCA (Barnett and Preisendorfer, 1987) is used as the statistical technique to develop the MOS equations. It is employed to set up the statistical linkages between large-scale global model hindcasts and gridded precipitation over Iran. CCA has already been widely applied to perform statistical downscaling as a forecast technique (Landman *et al*., 2012, and references therein). Both predictor and predictand fields are to be pre-filtered by principal components analysis. Pre-filtering and orthogonalisation of predictors and predictands prior to CCA have been recommended by researchers such as Barnett and Preisendorfer (1987) and Barnston and Ropelewski (1992). The optimal number of modes to be retained in the CCA equations is determined using retroactive validation (Troccoli *et al*., 2008; Landman *et al*., 2012). To estimate true forecast skill, a set of data independent of the training period is predicted. Such

validation mimics true operational forecasting when no prior knowledge of future seasons is available. For example, for OND precipitation, the MOS models are trained using only the years from 1982 to and including 1995 (a 14-year training period), and then a prediction is subsequently performed for the next year (1996). The MOS models are then retrained using the second training period from 1982 to and including 1996 (a 15-year training period), and then a prediction is subsequently performed for 1997. This procedure is repeated until the last year (2010) of the data has been predicted using MOS models trained from 1982 to and including 2009, resulting in 15 years of independent forecast data.

Cross-validation using a five-year-out window is considered for the training period. The various cross-validation training periods are each associated with a different number of years. Error variances for each period are uniquely calculated and these variances are subsequently used to calculate the probabilistic downscaled hindcasts. The ROC (Mason, 1982; Marzban, 2004; Wilks, 2011) and reliability diagrams (Hamill, 1997; Wilks, 2011) are used for verification of the probabilistic hindcasts in order to determine skill. The ROC indicates whether or not the forecast probability was higher when an event occurred compared to when it did not (Troccoli *et al*., 2008).

Pearson's correlation coefficient and Kendall's tau are used to estimate the deterministic forecast skill parameters. Using bootstrap resampling, the significance of Pearson's correlation coefficient and Kendall's tau are evaluated for the independent test period. All these methods are used to find the association between large-scale atmospheric circulation and precipitation to improve our understanding of the physical mechanism responsible for seasonal precipitation over Iran.

The MOS equations and forecast skill are run using the CPT of the IRI (http://iri.columbia.edu/our-expertise/climate/tools/cpt/).

3. Results

3.1. ROC scores

The downscaled simulations of atmospheric ECHAM4.5 GCM and forecasts systems are assessed over the region to indicate the strengths and weaknesses in the considered GCMs, and to determine which season is the most predictable by using the forecast systems described here. ROC scores are calculated for five three-month seasons and for all GCMs.

The verification results for the above- and below-normal categories are used since there is usually little skill to be derived from predicting the near-normal category (Van den Dool and Toth, 1991). The above- and below-normal categories are defined by the upper and lower 75^{th} and 25^{th} percentiles of the data records, respectively. Table 1 and Figure 2 indicate the ROC scores for the above- and below-normal categories for the OND, NDJ, DJF, JFM and FMA seasons for ECHAM4.5 GCM and GCM forecasts for the 15-year retroactive test period (1995/96 to 2009/10) when the models' precipitation field is considered as the predictor field. The skill of the forecasts of ECHAM4.5-SSTCA and ECHAM4.5-GML-CFSSST is low in comparison with the others. All models have been found to produce high scores for both categories for OND at lead times of up to three months with ECHAM4.5-SSTCA being the exception. Therefore, the seasonal OND precipitation over Iran is more predictable than the other seasons. A possible explanation is the significant linear correlation between Niño3.4 SST and OND precipitation over the larger part of Iran (Figure 3). Figure 3a indicates that there is a significant positive correlation between Niño3.4 SST and OND precipitation over the central, north-eastern and western areas of Iran. The linear association between Niño3.4 SST and precipitation of the other seasons is weak and insignificant for the most part. The ROC scores for the below-normal category for the FMA season for the ECHAM4.5-SSTCA model, which is based on the SST forecasting at one lead time, is high because of increasing correlation values between Niño3.4 SST and FMA precipitation in comparison with the DJF and JFM seasons. Figure 4 shows the composite maps of moisture transport and streamlines at 850 hPa geopotential height for the seasons considered. One of the reasons for the increasing correlation values between Niño3.4 SST and FMA precipitation is that moisture transport to the northern half of Iran increases during April compared with the earlier two months. The main source of the supplied moisture is mostly located over the eastern parts of the Mediterranean Sea (Figure 4).

Table 1. Roc scores of different GCMs when precipitation field used as the predictor.

Seasons / Models		ONDb	NDJb	DJFb	JFMb	FMAb	ONDa	NDJa	DJFa	JFMa	FMAa
AGCM		0.71	0.58	0.59	0.64	0.62	0.62	0.55	0.55	0.61	0.59
NMME NASA	L1	0.73	0.62	0.46	0.55	0.68	0.71	0.57	0.43	0.49	0.60
	L2	0.74	0.62	0.57	0.44	0.57	0.72	0.58	0.55	0.56	0.53
	L3	0.68	0.60	0.55	0.57	0.45	0.65	0.66	0.50	0.53	0.49
NMME Direct Coupled	L1	0.71	0.64	0.59	0.57	0.59	0.71	0.63	0.53	0.55	0.47
	L2	0.66	0.64	0.59	0.50	0.54	0.67	0.60	0.58	0.51	0.46
	L3	0.67	0.60	0.53	0.52	0.68	0.62	0.52	0.54	0.48	0.63
Coupled MOM3DC2	L1	0.72	0.62	0.56	0.54	0.58	0.70	0.62	0.51	0.53	0.48
	L2	0.71	0.59	0.58	0.50	0.52	0.68	0.62	0.57	0.49	0.44
	L3	0.72	0.55	0.61	0.44	0.64	0.62	0.54	0.61	0.52	0.61
NMME NCEP	L1	0.74	0.50	0.62	0.58	0.53	0.66	0.54	0.65	0.61	0.54
	L2	0.69	0.56	0.53	0.51	0.54	0.62	0.54	0.52	0.45	0.43
	L3	0.72	0.60	0.53	0.51	0.48	0.69	0.70	0.46	0.53	0.45
Coupled GML	L1	0.63	0.57	0.41	0.49	0.56	0.53	0.60	0.40	0.56	0.50
	L2	0.68	0.61	0.58	0.49	0.53	0.60	0.54	0.54	0.45	0.49
	L3	0.64	0.61	0.49	0.56	0.52	0.54	0.60	0.53	0.54	0.48
ECHAM4.5 SSTCA	L1	0.63	0.51	0.50	0.50	0.66	0.55	0.50	0.53	0.46	0.60
	L2	0.66	0.52	0.54	0.58	0.49	0.62	0.55	0.49	0.60	0.45
	L3	0.46	0.59	0.47	0.47	0.58	0.47	0.50	0.41	0.57	0.53

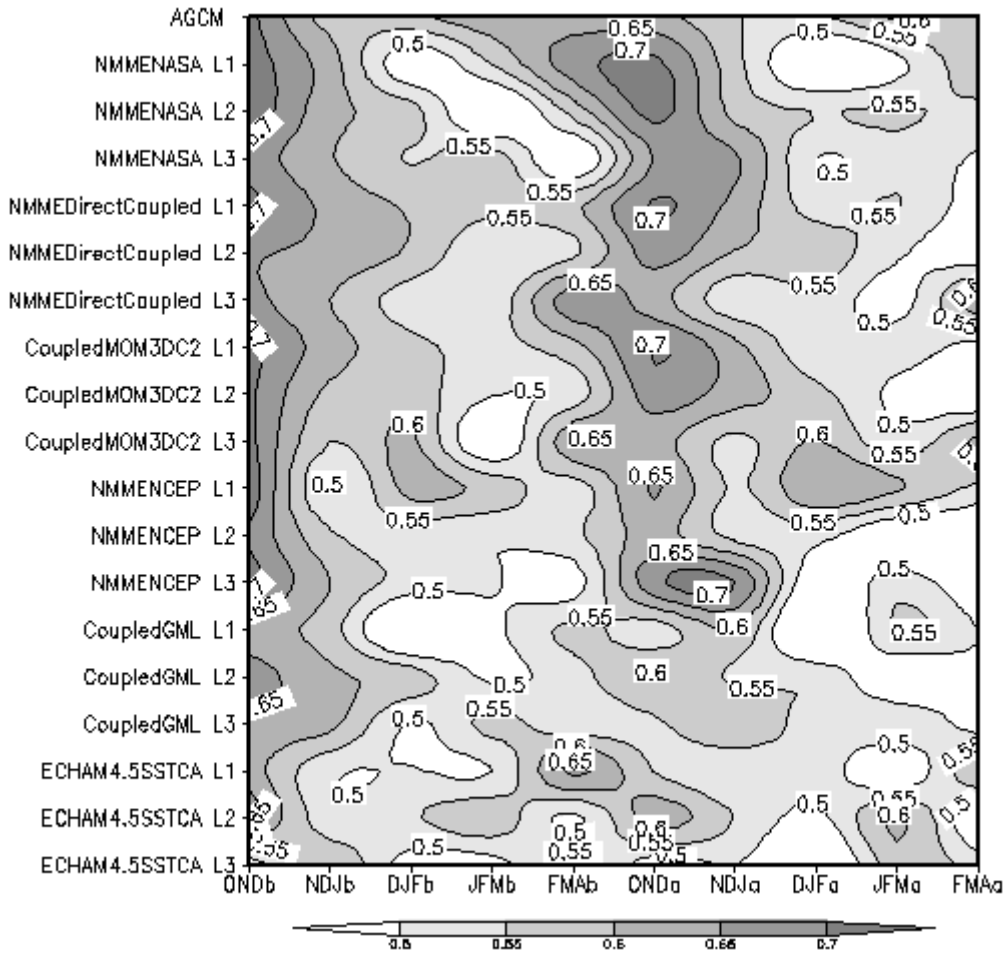


Fig. 2. ROC scores of different GCMs when precipitation is used as the predictor field. The b refers to below-normal and a to above-normal (e.g. JFMb represents below-normal JFM). The L refers to lead-times (e.g. L2 represents a 2-month lead-time).

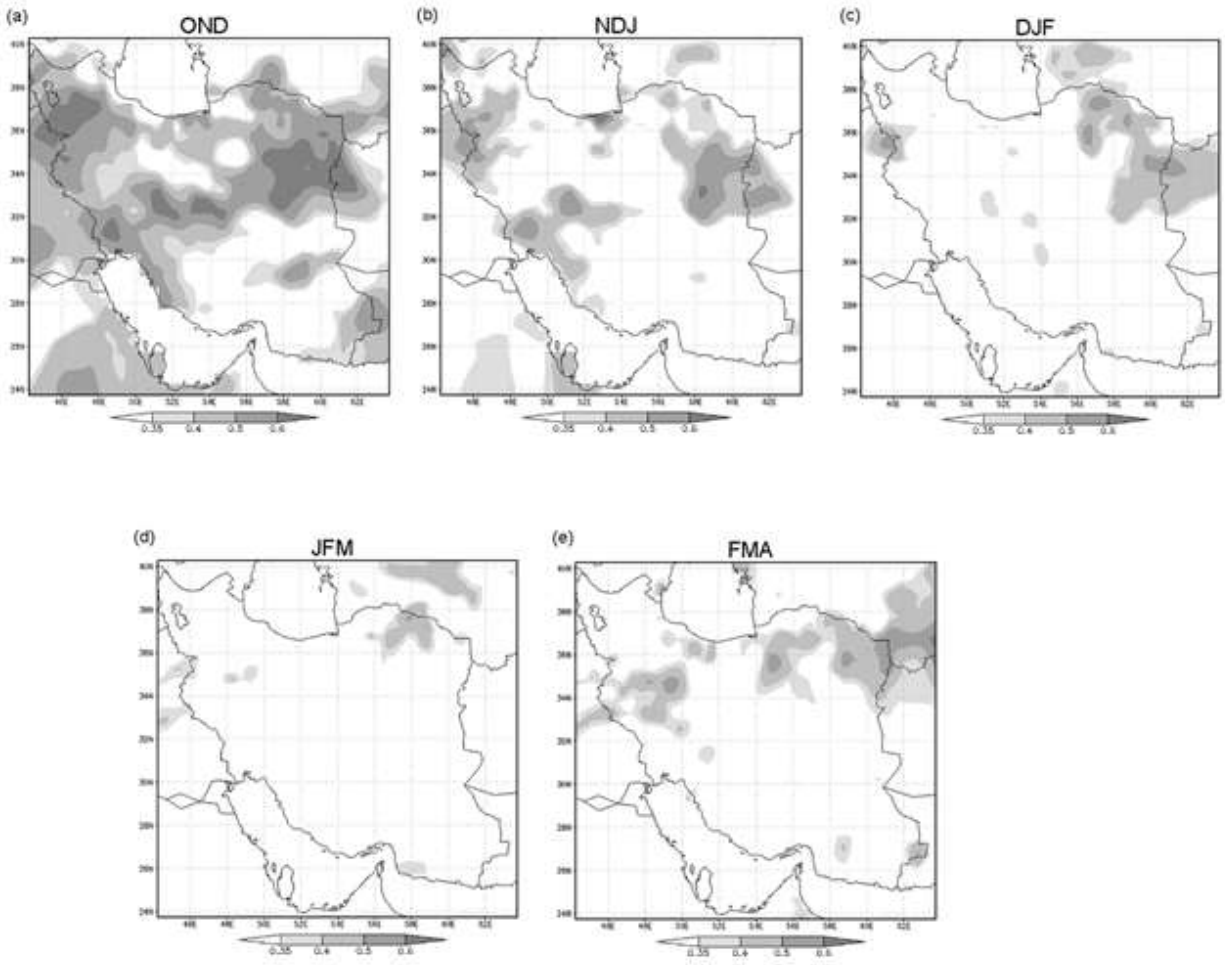
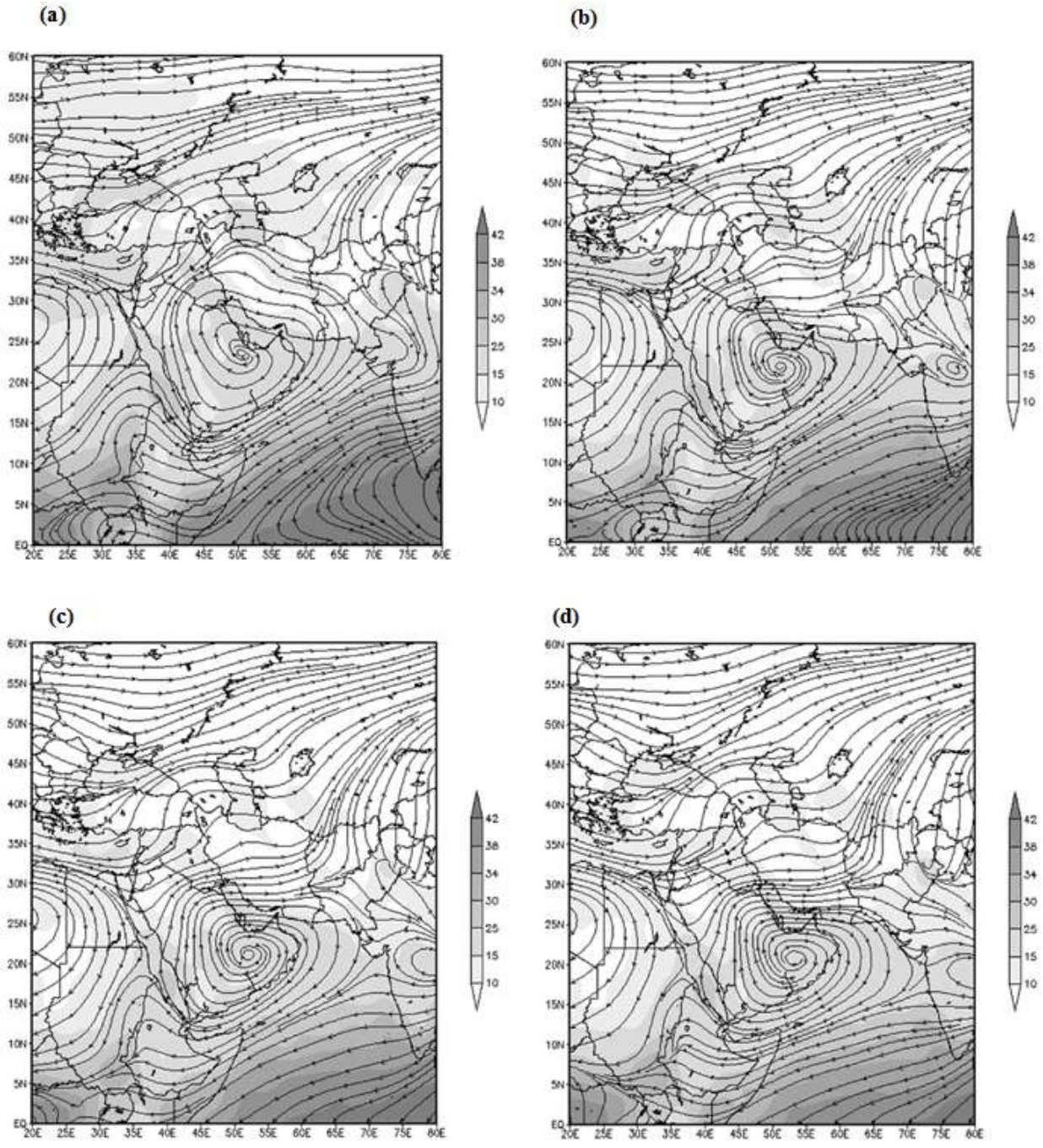


Fig. 3. Correlation coefficient between Niño 3.4SST and the GPCCv6 gridded precipitation over the period 1982- 2009. (a) OND, (b) NDJ, (c) DJF, (d) JFM, and (e) FMA season. Gray color shows significant correlation coefficient at 95% level.


(a)
(b)
(c)
(d)
60N
55N
50N
45N
40N
35N
30N
25N
20N
15N
10N
5N
EQ
20E
25E
30E
35E
40E
45E
50E
55E
60E
65E
70E
75E
80E
42
38
34
30
25
15
10

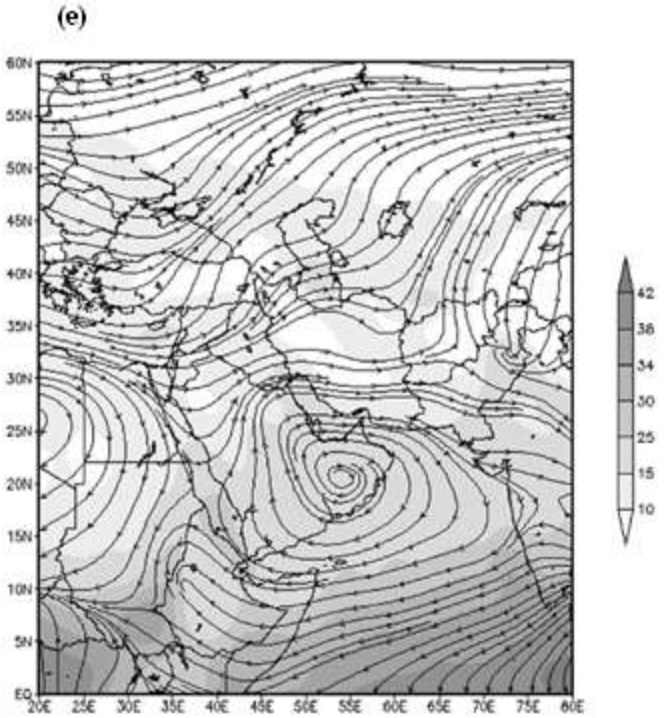


Fig. 4. Moisture transport (g s^{-1}) and Streamlines; (a) OND, (b) NDJ, (c) DJF, (d) JFM, and (e) FMA season. Climatological means over 1982-2010.

Figure 5 shows ROC score differences between ECHAM4.5 simulation and six forecast systems for five three-month seasons. This figure indicates that most of the models are better able to discriminate below-normal seasons from the rest of the seasons than they are at discriminating above-normal rainfall from the rest of the seasons, since most of the differences are positive. During the 15-year test period, at the end of the predicted calendar years, the Niño3.4 SST is less than -0.5℃ for seven of the 15 years for all seasons, and more than $+0.5$℃ for two, three and five years for the FMA, JFM and first three seasons, respectively – therefore, there were more La Niña years during the test period. A possible explanation for the higher ROC scores for the below-normal seasons is the association between cold event years and observed dry seasons over the study area, similar (although in opposite sense) to what has been found for Australian and South African rainfall variability (Power *et al*., 2006; Landman *et al*., 2012). For DJF, NCEP-CFSv2 has been able to produce the highest ROC scores among all models at a one-month lead time (Figure 5). Figure 5 indicates that for both categories of the JFM season, the ECHAM 4.5 simulation has been able to produce higher ROC scores than all forecast systems. However, for the NDJ season

for above- and below-normal categories, IRI-ECHAM4.5-DirectCoupled, NASA-GMAO-062012, NCEP-CFSv2, and ECHAM4.5-MOM3-DC2 have been able to produce higher ROC scores than the ECHAM4.5.

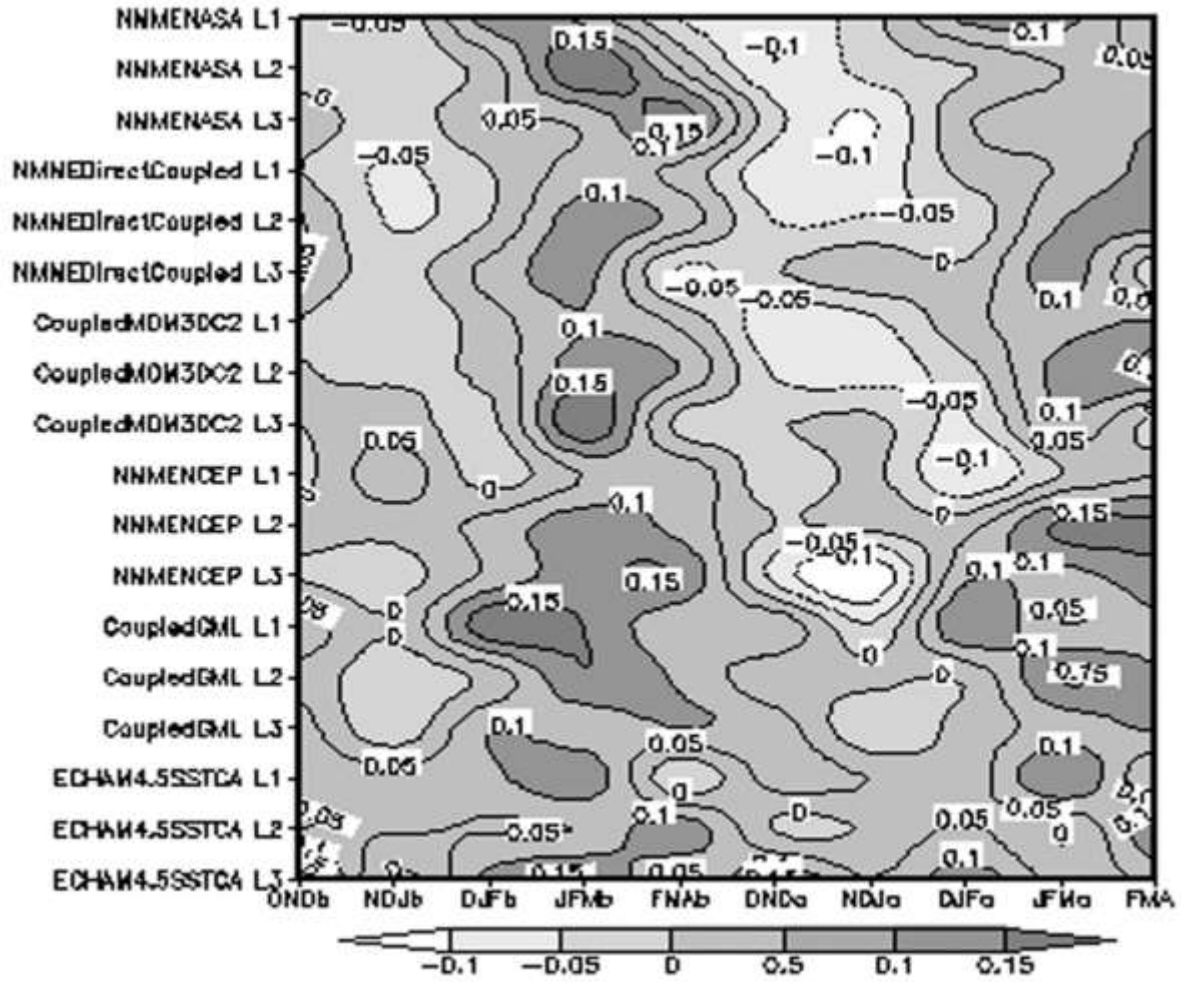


Fig. 5. ROC scores differences between ECHAM4.5 GCM simulations and models when precipitation is used as the predictor field.

Table 2 and Figure 6 indicate the calculated ROC scores for five three-month seasons and for NCEP-NCAR data, ECHAM4.5, ECHAM4.5-MOM3-DC2, ECHAM4.5-GML-CFSSST, and ECHAM4.5-SSTCA when 850 hPa geopotential heights are considered as the predictor field for the 15-year retroactive test period (1995/96 to 2009/10). The OND season is again found to be the most predictable. The ECHAM4.5-MOM3-DC2 produced the highest ROC scores at a one-month lead time for all seasons except for the FMA season. Figure 7 shows the difference between NCEP-NCAR data and ECHAM4.5-MOM3-DC2, ECHAM4.5-

GML-CFSSST and ECHAM4.5-SSTCA for all lead times and seasons considered. The forecast systems have produced scores close to NCEP-NCAR for both categories of NDJ, in particular for below-normal. Additionally, the coupled models have no success for the FMA season, particularly for the above-normal category. Figure 7 confirms that ECHAM4.5-MOM3-DC2 is generally better than the others except for below-normal FMA. For this reason, ECHAM4.5-MOM3-DC2 can be considered as the best model.

Table 2. Roc scores of NCEP-NCAR and different GCMs when the 850 hPa geopotential heights is used as the predictor field.

Seasons / Models		**ONDb**	**NDJb**	**DJFb**	**JFMb**	**FMAb**	**ONDa**	**NDJa**	**DJFa**	**JFMa**	**FMAa**
NCEP-NCAR		0.83	0.60	0.66	0.59	0.77	0.75	0.66	0.71	0.66	0.71
AGCM		0.77	0.64	0.61	0.63	0.64	0.74	0.64	0.61	0.59	0.58
Coupled MOM3DC2	L1	0.75	0.56	0.61	0.60	0.54	0.66	0.59	0.61	0.70	0.51
	L2	0.79	0.58	0.51	0.49	0.48	0.57	0.53	0.50	0.47	0.41
	L3	0.75	0.57	0.51	0.45	0.53	0.63	0.56	0.54	0.48	0.52
Coupled GML	L1	0.67	0.61	0.57	0.51	0.62	0.55	0.54	0.55	0.48	0.54
	L2	0.70	0.57	0.61	0.47	0.49	0.59	0.54	0.48	0.47	0.43
	L3	0.71	0.52	0.55	0.53	0.59	0.59	0.53	0.53	0.48	0.52
ECHAM4.5 SSTCA	L1	0.65	0.62	0.58	0.57	0.59	0.53	0.61	0.60	0.50	0.50
	L2	0.53	0.63	0.57	0.48	0.52	0.55	0.50	0.59	0.48	0.41
	L3	0.48	0.57	0.59	0.51	0.54	0.50	0.55	0.53	0.52	0.40

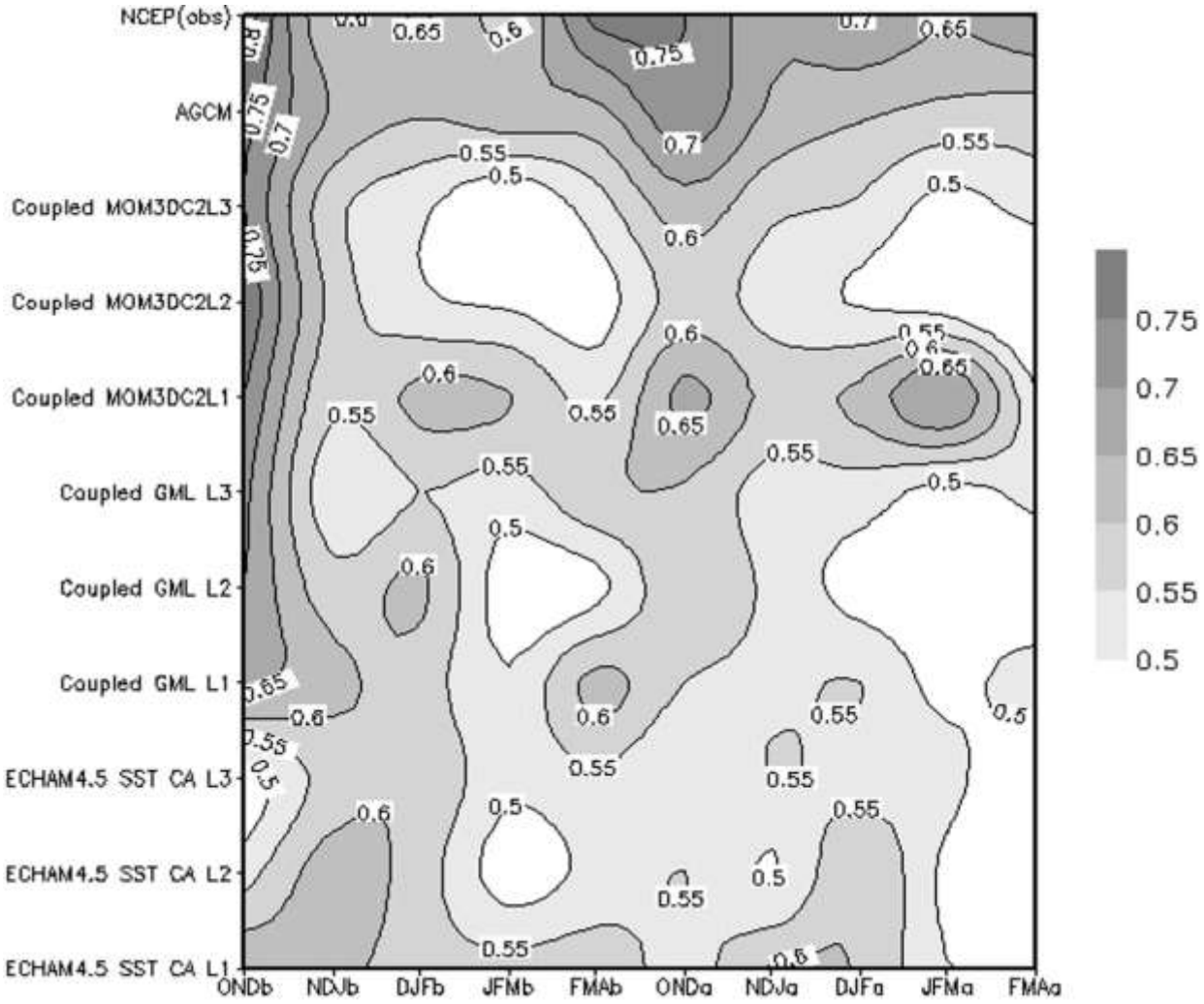


Fig. 6. ROC scores of NCEP-NCAR and different GCMs when the 850 hPa geopotential height is used as the predictor field.

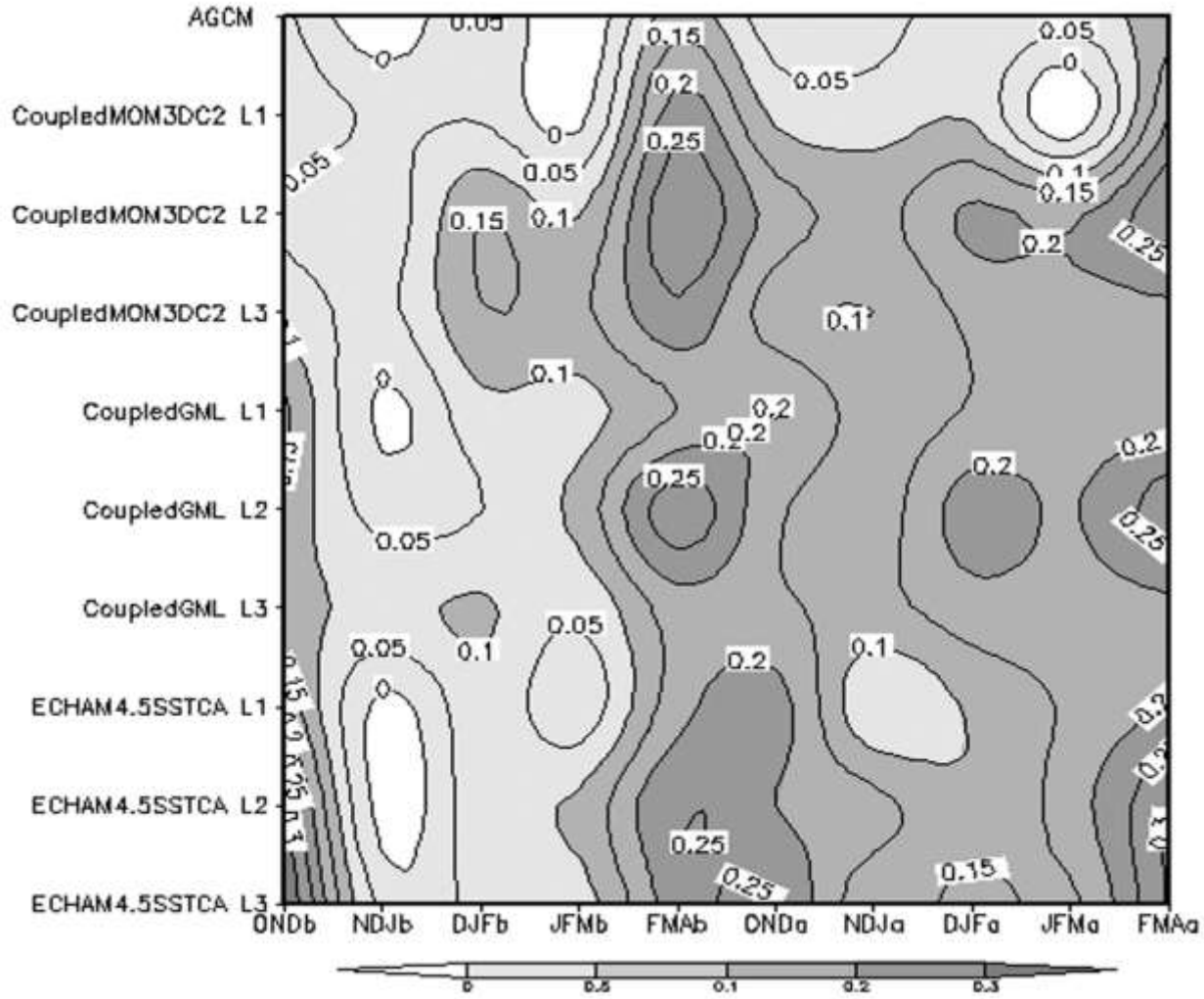


Fig. 7. ROC scores differences between NCEP-NCAR and models when the 850 hPa geopotential height is used as the predictor field.

3.2. Reliability diagram

The obtained ROC scores indicate that high scores are generally obtained from ECHAM4.5-MOM3-DC2 forecasts for a one-month lead time when using the 850 hPa geopotential height fields as predictor field. OND and FMA are the seasons with the highest and lowest predictability, respectively. For further discussion, we will consider this coupled model for OND, DJF and JFM when using the 850 hPa geopotential height field as predictor and NDJ when precipitation is used as predictor at a one-month lead time.

Figure 8 shows reliability diagrams at a one-month lead time for these seasons. The weighted least square regression lines for two categories (above- and below-normal) are also shown.

The frequencies histograms of wet-dry forecasts are presented in probability intervals of 10% starting at 5%, and explain how strongly and frequently the issued forecasts depart from the climatology probabilities. The mean of the forecasts probabilities is represented in the x-axis and the observed relative frequencies is represented in the y-axis.

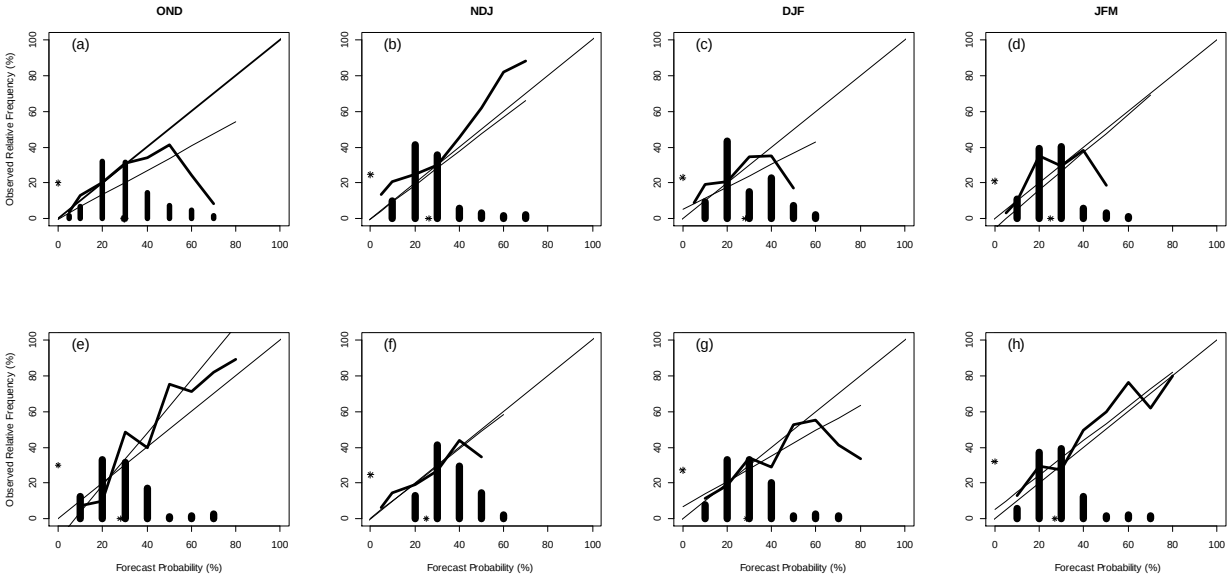


Fig. 8. Reliability diagram and frequency histogram for above (upper Fig.s) and below (lower Fig.s) normal seasonal precipitation 1-monthe lead forecasts produced by downscaling the ECHAM4.5 MOM3DC2 coupled models to Iranian gridded precipitation. The thin lines are the weighed least square regression lines of the reliability curve. The star marked on the x-axis and y-axis shows the means of the forecasts probabilities and observed relative frequencies, respectively.

In Figure 8a, for above-normal OND rainfall, the forecast probabilities are higher than the observed relative frequencies, implying that observed wet OND precipitation occurs less frequently than predicted (over-forecasting).

Moreover, for OND dry conditions, Figure 8e indicates under-confidence. In this case the forecast has high resolution but poor reliability (Troccoli *et al*., 2008). In Figures 8a and 8e, the frequencies histogram indicate some sharpness since the peak of histogram for the wet and dry OND season are in the intervals 25–35% and 35–45%, which are near to the climatology probability of 25%. The mean of above-normal precipitation occurred at 20% while the mean for forecast probabilities was 29%; and the mean of below-normal precipitation occurred at 30% while the mean for forecast probabilities was 28%

The forecast probabilities of the NDJ season are over-confident and the observed relative frequencies are in the intervals of 25–35% and 35–45%, which indicate some sharpness

(Figure 8b). The mean of above-normal precipitation for the NDJ season occurred at 24.7%, while the mean for forecast probabilities was 26.3%.

Figure 8f indicates very slight over-confidence for below-normal NDJ and the slope is close to unity so that the mean of the below-normal precipitation season and forecast probabilities are 24.7% and 25%, respectively.

Figures 8c and 8g indicate that the forecast probabilities of the DJF season are more than the observed relative frequencies, implying that observed wet-dry DJF precipitation occurs less frequently than predicted (over-forecasting or wet bias) and the frequency histogram of dry DJF indicates some sharpness since the histogram of dry DJF are in intervals 25–35% and 35–45%, which are near the climatology probability of 25%. The mean of above normal precipitation occurred in 22.9% while the mean for forecast probabilities was 28.2%; and the mean of below-normal precipitation occurred in 27.2% while the mean for forecast probabilities was 28.9%.

Figure 8d indicates that the forecast probabilities of wet JFM are more than the observed relative frequencies, implying that observed wet JFM precipitation occurs less frequently than predicted (over-forecasting or wet bias). The mean of above-normal precipitation for the JFM season occurred at 21% while the mean for the forecast probabilities was 25%.

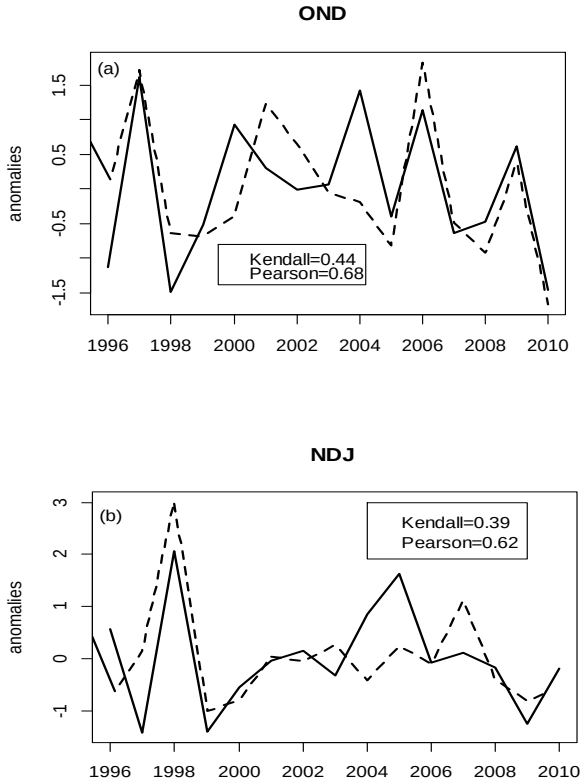
On the contrary, Figure 8h indicates that the forecast probabilities of dry JFM season are less than the observed relative frequencies, implying that observed dry JFM precipitation occurs more frequently than predicted (under-forecasting). The mean of below-normal precipitation for JFM occurred at 32% while the mean for forecast probabilities was 27%. Additionally, the histograms in Figures 8d and 8h indicate some sharpness.

Therefore, the reliability results are in general agreement with the ROC scores.

3.3. Non-probabilistic forecast skill

Figure 9 shows the area-averaged and then normalised observed OND, NDJ, DJF, JFM precipitation indices over the study area, versus one-month lead-time retroactive forecasts obtained by downscaling the ECHAM4.5-MOM3-DC2 predictions to Iranian gridded precipitation. The Pearson correlation and Kendall's tau as non-probabilistic skill estimates indicate the association between the area-averaged time series of the observed and forecasts (Figure 9). The bias-corrected and accelerated bootstrap confidence interval, or BCa intervals

(Efron and Tibshirani, 1993; Wilks, 2011) for 90% and 95% for the Pearson correlation and Kendall's tau for the OND, NDJ, DJF and JFM seasons are shown in Table 3. The BCa indicates that the Pearson correlation is statistically significant for the OND, NDJ and JFM seasons (the corresponding lower confidences interval are more than zero). The BCa of the Kendall correlation is statistically significant for the OND and JFM seasons.



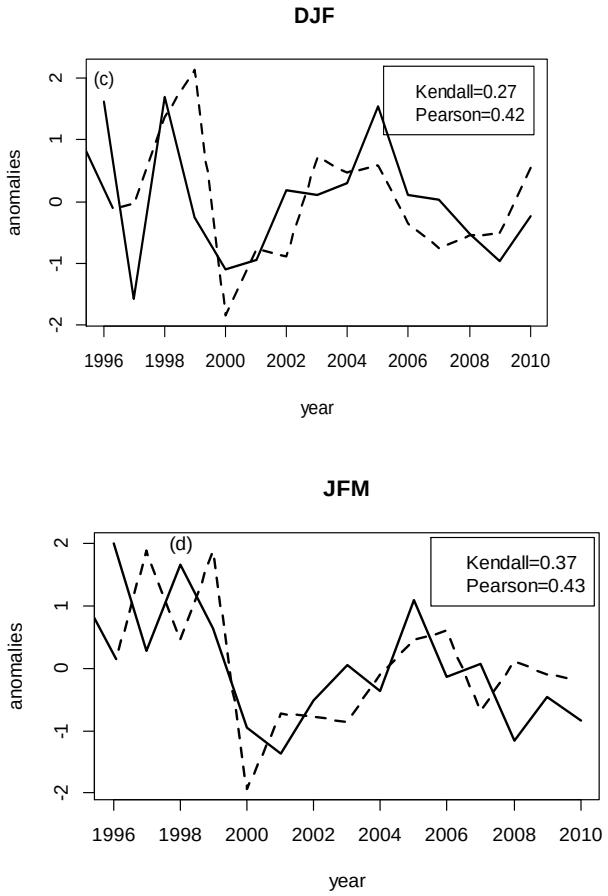


Fig. 9. Area-average (a) OND, (b) NDJ, (c) DJF and (d) JFM precipitation index (OBS) over the study area vs retroactive forecast obtained by downscaling the ECHAM4.5-MOM3DC2 one month lead predictions to gridded precipitation. Solid (observed) and Dash (predict). The Pearson and Kendall correlations between the area averaged observed and predicted are brought in the box.

Table 3. The bias-corrected bootstrap confidence interval for 90% and 95% for Pearson correlation and Kendall's tau for five seasons.

	Pearson Correlation				**Kendall's tau**			
Season\ BCs	**0.025**	**0.05**	**0.95**	**0.975**	**0.025**	**0.05**	**0.95**	**0.975**
OND	0.33	0.40	0.87	0.90	0.11	0.17	0.65	0.69
NDJ	0.08	0.16	0.88	0.92	-0.13	-0.04	0.73	0.79
DJF	-0.24	-0.10	0.77	0.80	-0.18	-0.08	0.66	0.70
JFM	0.10	0.16	0.61	0.64	0.03	0.10	0.57	0.61

The seasonal area-averaged NDJ, DJF and JFM observed precipitation in 2000, 2001 and 2008–2010 as dry years are associated with La Niña events. The sign of the area-averaged precipitation during these 12 cases (three seasons x four La Niña years) are captured by ten of the forecasts, indicating some skill during the La Niña years for winter boreal variability

over Iran. On the contrary, the skill of the forecasts for DJF and JFM is low during the weak La Niña and neutral years of 1996 and 1997, respectively.

3.4. CCA maps

The 850 hPa geopotential height data of NCEP-NCAR is used as the predictor in place of the forecasted low-level height fields of the ECHAM4.5-MOM3-DC2 to check whether this model reproduces the relationship between the geopotential height data of NCEP-NCAR and precipitation. For the comparison of NCEP-NCAR data and ECHAM4.5-MOM3-DC2, only the first principal component is considered as the mode to be retained in the CCA equations in the retroactive test for both ECHAM4.5-MOM3-DC2 and NCEP-NCAR.

Figure 10 shows the predictor spatial loading (XCCA) of the first CCA mode for four seasons for NCEP-NCAR data and ECHAM4.5-MOM3-DC2. Figures 10a and 10b indicate that the spatial loadings are similar for both NCEP-NCAR and ECHAM4.5-MOM3-DC2, and the reason for the skill of this coupled model in predicting precipitation during OND is because the model is able to reproduce the NCEP-NCAR leading pattern.

By comparing Figures 10c and 10d, different spatial loading over the north-west is shown, which suggests that the ECHAM4.5-MOM3-DC2 coupled model is unable to reproduce NCEP-NCAR leading pattern and the degrading forecast skill for NDJ.

Also, the ECHAM4.5-MOM3-DC2 coupled model is not really able to reproduce NCEP-NCAR leading pattern for DJF and JFM (Figures 10e–h).

4. Discussion and conclusions

Seasonal precipitation forecast skill over Iran has been investigated using forecasts from six forecast systems for five seasons of three months. These models are statistically downscaled and compared over a retroactive period of 15 years. For most models, the season with the highest skill forecast at lead times of up to three months is OND. One reason for this predictability is that OND precipitation is linearly associated with ENSO events, particularly cold events, over most of Iran. A low skill is produced from most models for all lead times for the remaining seasons except for below-normal FMA when the models' precipitation is considered as the predictor field. However, NCEP-CFSv2 has been able to produce the highest ROC scores among all models for winter precipitation (DJF) at one-month lead time

because the ROC scores of this model are higher than those found for ECHAM4.5. For most seasons, a high skill is obtained from ECHAM4.5-MOM3-DC2 forecasts at a one-month lead time when models' 850 hPa geopotential heights are used as the predictor field. The 850 hPa geopotential heights are generally a good predictor for most seasons. The skill of all forecasts for the below-normal category is more than that found for the above-normal category, which indicates some skill during the La Niña years. Although the response of DJF and JFM precipitation to ENSO events is weak over most of Iran, ECHAM4.5-MOM3-DC2 indicates high skill for these months at a one-month lead time. For this model at a one-month lead time when models' 850 hPa geopotential heights are used as the predictor field, there is a significant Pearson correlation between the area-averaged of the observed and forecasts over the study area for the OND, NDJ, DJF and JFM seasons that are 0.68, 0.62, 0.42 and 0.43, respectively. This suggests that such a coupled model is able to improve Iranian precipitation forecast. The ECHAM4.5-MOM3-DC2 coupled model correctly reproduces the relationship between the heights and precipitation for OND. However, this coupled model is not really able to reproduce the physical mechanism of the NCEP-NCAR model for the NDJ, DJF and JFM seasons – therefore, the highest skill is restricted to OND.

The paper has demonstrated that seasonal rainfall over Iran is predictable at lead times of a few months. However, predictability is mainly restricted to OND when ENSO states have been found to be strongly linked to Iranian rainfall variability. Notwithstanding, the OND season is of great significance for various agricultural practices, such as dry-land cropping – since dry-land farmers know that to survive financially they most likely would have to take advantage of good OND rainfall seasons to offset dry seasons. Using forecasts from systems such as those presented and tested here should be able to aid such agricultural practices.

Acknowledgements

The authors greatly appreciate the two anonymous reviewers for their useful comments and suggestions. Sections of this work were prepared when the first author was a scientific visitor at the Council for Scientific and Industrial Research. This work was funded by the Fars Regional Water Organization.

References

Balmaseda M, Anderson DLT. 2009. Impact of initialization strategies and observations on seasonal forecast skill. *Geophysical Research Letters* **36**: L01701. doi:10.1029/2008GL035561.

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

Barnett TP, Preisendorfer R. 1987. Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis. *Monthly Weather Review* **115**: 1825–1850.

Barnston AG, Ropelewski CF. 1992. Prediction of ENSO episodes using canonical correlation analysis. *Journal of Climate* **5**: 1316–1345.

Barnston AG, Mason SJ, Goddard L, Dewitt DG, Zebiak SE. 2003. Multimodel ensembling in seasonal climate forecasting at IRI. *Bulletin of the American Meteorological Society* **84**(12): 1783–1796.

Barnston AG, Li S, Mason SJ, DeWitt G, Goddard L, Gong X. 2010. Verification of the first 11 years of IRI's seasonal climate forecasts. *Journal of Applied Meteorology and Climatology* **49**(3): 493–520.

Barnston AG, Mason SJ. 2011. Evaluation of IRI's seasonal climate forecasts for the extreme 15% tails. *Weather and Forecasting* **26**(4): 545–554.

Bartman AG, Landman WA, Rautenbach CJdeW. 2003. Recalibration of general circulation model output to austral summer rainfall over southern Africa. *International Journal of Climatology* **23**: 1407–1419.

DeWitt DG. 2005. Retrospective forecasts of interannual sea surface temperature anomalies from 1982 to present using a directly coupled atmosphere–ocean general circulation model. *Monthly Weather Review* **133**: 2972–2995.

Doblas-Reyes FJ, García-Serrano J, Lienert F, Pintó Biescas A, Rodrigues LRL. 2013. Seasonal climate predictability and forecasting: status and prospects. *WIREs Climate Change* **4**: 245–268. doi:10.1002/wcc.217.

Efron B, Tibshirani RJ. 1994. *An introduction to the bootstrap*. Chapman & Hall: New York.

Fallah-Ghalhary GA, Habibi-Nokhandan M, Mousavi-Baygi M, Khoshhal, J. 2010. Spring rainfall prediction based on remote linkage controlling using adaptive neuro-fuzzy inference system (ANFIS). *Theoretical and Applied Climatology* **101**(1–2): 217–233.

Glahn, HR, Lowry DA. 1972. The use of model output statistics (MOS) in objective weather forecasting. *Journal of Applied Meteorology* **11**: 1203–1211.

Goddard L, Dilley M. 2005. El Niño: catastrophe or opportunity. *Journal of Climate* **18**(2): 651–665.

Hamill TM. 1997. Reliability diagrams for multicategory probabilistic forecasts. *Weather and Forecasting* **12**: 736–741.

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 KC, Ropelewski C, Wang J, Jenne R, Joseph D. 1996. The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society* **77**: 437–471.

Kirtman BP, Min D, Infanti JM, Kinter III JL, Paolino DA, Zhang Q, Van den Dool H, Saha S, Mendez MP, Becker E, Peng P, Tripp P, Huang J, DeWitt DG, Tippett MK, Barnston AG, Li S, Rosati A, Schubert SD, Rienecker M, Suarez M, Li AE, Marshak J, Lim YK, Tribbia J, Pegion K, Merryfield WJ, Denis B, Wood EF. 2014. The North American Multi-Model Ensemble (NMME): phase-1 seasonal-to-interannual prediction; phase-2 toward developing intra-seasonal prediction. *Bulletin of the American Meteorological Society* **95***:* 585–601.

Landman WA, Beraki A. 2012. Multi-model forecast skill for mid-summer rainfall over southern Africa. *International Journal of Climatology* **32**(2): 303–314.

Landman WA, DeWitt D, Lee DE, Beraki A, Lötter D. 2012. Seasonal rainfall prediction skill over South Africa: one- versus two-tiered forecasting systems. *Weather and Forecasting* **27**(2): 489–501.

Landman WA. 2014. How the International Research Institute for Climate and Society has contributed towards seasonal climate forecast modelling and operations in South Africa. *Earth Perspectives* **1**: 1–22.

Marzban C. 2004. The ROC curve and the area under it as performance measures. *Weather and Forecasting* **19**: 1106–1114.

Mason IB. 1982. A model for assessment of weather forecasts. *Australian Meteorological Magazine* **30**: 291–303.

National Research Council. 2010. *Assessment of Intraseasonal to Interannual Climate Prediction and Predictability*. The National Academies Press: Washington, DC.

Nazemosadat M, Coredy I. 2000. On the relationships between ENSO and autumn rainfall in Iran. *International Journal of Climatology* **20**: 47–61.

Nazemosadat M, Ghasemi AR. 2004. Quantifying the ENSO-related shifts in the intensity and probability of drought and wet periods in Iran. *Journal of Climate* **17**(20): 4005–4018.

Pacanowski RC, Griffies SM. 1998. *MOM 3.0 manual*. NOAA/Geophysical Fluid Dynamics Laboratory: Princeton, NJ, 608pp.

Pourasghar F, Tozuka T, Jahanbakhsh S, Sarraf BS, Ghaemi H, Yamagata T. 2012. The interannual precipitation variability in the southern part of Iran as linked to large-scale climate modes. *Climate Dynamics* **39**(9–10): 2329–2341.

Power S, Haylock M, Colman R, Wang X. 2006. The predictability of interdecadal changes in ENSO activity and ENSO teleconnections. *Journal of Climate* **19**: 4755–4771.

Raziei T, Mofidi A, Santos JA, Bordi I. 2012. Spatial patterns and regimes of daily precipitation in Iran in relation to large-scale atmospheric circulation. *International Journal of Climatology* **32**(8): 1226–1237.

Roeckner E, Arpe K, Bengtsson L, Christoph M, Claussen M,Dümenil L, Esch M, Giorgetta M, Schlese U, Schulzweida U. 1996. *The atmospheric general circulation model ECHAM4: model description and simulation of present-day climate.* Max-Planck-Institut für Meteorologie Rep. 218: Hamburg, 90pp.

Ropelewski CF, Halpert MS. 1987. Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation (ENSO). *Monthly Weather Review* **115**: 1606–1626.

Ropelewski CF, Halpert MS. 1989. Precipitation patterns associated with the high index phase of the Southern Oscillation. *Journal of Climate* **2**: 268–284.

Sachindra DA, Huang F, Barton A, Perera BJC. 2014a. Statistical downscaling of general circulation model outputs to precipitation—part 1: calibration and validation. *International Journal of Climatology* **34**(11): 3264–3281.

Sachindra DA, Huang F, Barton A, Perera BJC. 2014b. Statistical downscaling of general circulation model outputs to precipitation—part 2: bias-correction and future projections. *International Journal of Climatology* **34**(11): 3282–3303.

Saha S, Nadiga S, Thiaw C, Wang J, Wang W, Zhang Q, Van den Dool HM, Pan HL, Moorthi S, Behringer D, Stokes D, Peña M, Lord S, White G, Ebisuzaki W, Peng P, Xie P. 2006. The NCEP Climate Forecast System. *Journal of Climate* **19**: 3483–3517.

Schneider U, Becker A, Finger P, Meyer-Christoffer A, Rudolf B, Ziese M. 2011. GPCC Full Data Reanalysis Version 6.0 at 0.5°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data. doi:10.5676/DWD_GPCC/FD_M_V6_050.

Tippett MK, Barlow M, Lyon B. 2003. Statistical correction of central southwest Asia winter precipitation simulations. *International Journal of Climatology* **23**: 1421–1433.

Tippett MK, Goddard L, Barnston AG. 2005. Statistical-dynamical seasonal forecasts of central-southwest Asian winter precipitation. *Journal of Climate* **18**(11): 1831–1843.

Troccoli A, Harrison M, Anderson DLT, Mason S. 2008. *Seasonal Climate: Forecasting and Managing Risk, NATO Science Series on Earth and Environmental Sciences*, Vol. 82. Springer: 467pp.

Van den Dool HM, Toth Z. 1991. Why do forecasts for "near normal" often fail? *Weather Forecasting* **6**: 76–85.

Van den Dool HM. 2007. *Empirical Methods in Short-Term Climate Prediction*. Oxford University Press: Oxford, 215pp.

Weisheimer A, Doblas-Reyes FJ, Palmer TN, Alessandri A, Arribas A, D´equ´e M, Keenlyside N, MacVean M, Navarra A, Rogel P. 2009. ENSEMBLES: A new multi-model ensemble for seasonal-to-annual predictions—Skill and progress beyond DEMETER in forecasting tropical Pacific SSTs. *Geophysical Research Letters* **36**: L21711. doi:10.1029/2009GL040896.

Wilks, DS. 2011. *Statistical Methods in the Atmospheric Sciences* (third edition). Academic Press: San Diego, CA.