Received: 19 January 2020 | Revised: 11 March 2021 | Accepted: 13 April 2021 | Published on: 19 May 2021

DOI: 10.1002/joc.7148

International Journal of Climatology RMetS

RESEARCH ARTICLE

Improved multi-model ensemble forecasts of Iran's precipitation and temperature using a hybrid dynamical-statistical approach during fall and winter seasons

Husain Najafi[1,2] | **Andrew W. Robertson**[3] | **Ali R. Massah Bavani**[2] | **Parviz Irannejad**[4] | **Niko Wanders**[5] | **Eric F. Wood**[6]

[1]Department of Computational Hydrosystems, Helmholtz Center for Environmental Research (UFZ), Leipzig, Germany

[2]Department of Irrigation and Drainage Engineering, University of Tehran, Tehran, Iran

[3]International Research Institute for Climate and Society (IRI), Columbia University, Palisades, New York

[4]Department of Space Physics, Institute of Geophysics, University of Tehran, Tehran, Iran

[5]Department of Physical Geography, Utrecht University, Utrecht, Netherlands

[6]Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey

Correspondence
Husain Najafi and Ali R. Massah Bavani, Department of Irrigation and Drainage Engineering, Aburaihan Campus, University of Tehran, Tehran, Iran.
Email: husain.najafi@ufz.de (H. N.) and armassah@ut.ac.ir (A. R. M. B.)

Abstract

Skillful seasonal climate forecasts can support decision making in water resources management and agricultural planning. In arid and semi-arid regions, tailoring reliable forecasts has the potential to improve water management by using key hydroclimate variables months in advance. This article analyses and compares the performance of two common approaches (empirical and hybrid dynamical-statistical) in seasonal climate forecasting over a drought-prone area located in Southwest Asia including Iran. Empirical models are framed as a baseline skill that hybrid models need to outperform. Both approaches provide probabilistic forecasts of precipitation and temperature using canonical correlation analysis to provide forecasts at 0.25° resolution. Empirical models are developed based on the large-scale observed atmosphere–ocean patterns for forecasting using antecedent climate anomalies as predictors, while the hybrid approach makes use of model output statistics to correct systematic errors in dynamical climate model forecast outputs. Eight state-of-the-art dynamical models from the North American Multi-Model Ensemble project are analysed. Individual models with the highest goodness index are weighted to develop seven different hybrid dynamical-statistical Multi-model Ensembles. In this study, (October–December) and (January–February) are considered as target seasons which are the most important periods within the water year for water resource allocation to the agriculture sector. The results show that the hybrid approach has improved performance compared to the raw general circulation models and purely empirical models, and that the performance of the hybrid models is season-dependent. Seasonal forecasts of precipitation (temperature) have a higher skill in OND (JFM). In addition, in most cases, Multi-model Ensemble (MME) is more skillful than the empirical models and outperforms individual dynamical models. However, the best individual model might be as skillful as the MME given the target season and region of interest.

KEYWORDS
hybrid, Iran, multi-model ensemble, North American Multi-Model Ensemble (NMME), Seasonal climate forecasting

1 | INTRODUCTION

Long-lead monthly and seasonal predictions can improve water management, especially in areas facing water scarcity. Continuous international efforts have been made during the past few decades to develop and improve long-range climate forecasting systems. Skillful seasonal climate forecasts are now available for many countries and regions which can help governments by providing valuable information for many disaster management decisions that add values into Regional Climate Outlook Forums (RCOFs) (Hellmuth *et al*., 2011; Mwangi *et al*., 2014; Shukla *et al*., 2016). Seasonal climate forecasts have assisted scholars and practitioners to tailor seasonal forecasting in sectoral applications such as streamflow prediction for water resources management (Golembesky *et al*., 2009; Li *et al*., 2009; Sankarasubramanian *et al*., 2009a, 2009b; Gobena and Gan, 2010; Oludhe *et al*., 2013; Robertson and Wang, 2013; Wang and Liu, 2013; Li *et al*., 2014; Robertson *et al*., 2014; Yu *et al*., 2014; Crochemore *et al*., 2016; Lu *et al*., 2017; Sahu *et al*., 2017), hydropower operation (Block, 2011; Block and Goddard, 2011), drought warning systems development (Mwangi *et al*., 2014; Sheffield *et al*., 2014), risk management in agriculture (Crane *et al*., 2011; Crane *et al*., 2010; Hansen *et al*., 2011), and water-food-energy nexus (Conway *et al*., 2015). Some recent studies have shown promising seasonal forecast skills (Kim *et al*., 2012; Weisheimer and Palmer, 2014; Dunstone *et al*., 2016) especially in the tropics (Barnston *et al*., 2012).

Seasonal forecasts can be constructed using empirical, and dynamic approaches. For building empirical seasonal climate forecasts, the coupled atmosphere–ocean processes and large-scale general circulation patterns are frequently used. Based on historical analysis, sources of regional predictability are identified in the first step. For example, anomalous hydroclimate conditions associated with the global impacts of El Niño-Southern Oscillation (ENSO) have been evaluated at global (e.g., Mason and Goddard, 2001; Sun *et al*., 2015; Emerton *et al*., 2017) and regional scales (Fraedrich and Müller, 1992; Moron and Plaut, 2003; Nazemosadat and Ghasemi, 2004; Brönnimann, 2007; Donat *et al*., 2014; Bichet *et al*., 2016; Rodríguez-Fonseca *et al*., 2016). Once the specific sources of predictability (land, atmosphere, ocean, and their interactions) are recognized, large-scale atmospheric-oceanic signals are correlated to observations to forecast regional hydroclimate conditions months or seasons in advance. For many years, empirical approaches have been applied as the primary approach towards the long-range prediction (Troccoli *et al*., 2008). In the dynamic approach, the skill of general circulation models (GCMs) is assessed for specific target seasons and different lead times. A recent analysis based on state-of-the-art coupled atmosphere–ocean GCMs has shown promising results at seasonal time scale over many parts of the world (Kim *et al*., 2012). Both methods provide decision-makers with relevant information for climate risk management given forecast models have appropriate resolution and reliable skill (National Council Report, 2010). Model Output Statistics (MOS) is applied to dynamical climate model outputs to remove systematic biases. Sometimes, MOS correction is referred to as a hybrid dynamical-statistical approach. Data-driven predictions are the most common methods which are applied as empirical methods for seasonal-to-interannual climate projections (Tippett and DelSole, 2013; Ciancarelli *et al*., 2014). Statistical models including linear regression techniques and Canonical Correlation Analysis (CCA) are also applied as MOS in several studies (Barnston and Tippett, 2017).

Since critical questions often arise during stakeholder engagement, on the current skill of state-of-the-art models in forecasting the hydroclimate indicators of relevance to decision-makers, the main idea of this study is to provide a framework for comparing the seasonal forecast skill of empirical and hybrid approaches which can be transferred to other regions. For this purpose, CCA is applied for empirical forecasting and developing hybrid models. The proposed framework uses the MOS of the North American Multi-Model Ensemble (NMME) predictions models (e.g., Kirtman *et al*., 2014). The NMME project includes both a retrospective forecast and a real-time prediction protocol, which enables bias correction, skill assessment, and forecast outputs in real-time. Note that extensive research has been conducted to verify monthly precipitation and temperature predictions from NMME. Some recent studies evaluated NMME seasonal forecast skill at global and regional scales mainly for drought onset (Yuan and Wood, 2013), streamflow, and soil moisture forecasting (Thober *et al*., 2015; Wood *et al*., 2015; Madadgar *et al*., 2016). NMME forecasts have been evaluated over many regions around the world, for example in the United States (Barbero *et al*., 2017; Khajehei *et al*., 2017), East Africa (Shukla *et al*., 2016), China (Ma *et al*., 2015), Central South West Asia

(Ehsan *et al*., 2017), and Europe (Mo and Lyon, 2015; Thober *et al*., 2015; Slater *et al*., 2017; Wanders *et al*., 2019).

One main source of climate predictability is the large-scale atmospheric circulation patterns and teleconnections. Their influence on surface climate variables (precipitation and temperature) has been investigated in a number of studies in south Asia (e.g., Ahmadi *et al*., 2019; Hoell *et al*., 2018; Nazemosadat and Cordery, 2000; Nazemosadat and Ghaedamini, 2010; Nazemosadat and Ghasemi, 2004; Ghasemi and Khalili, 2006; Raziei *et al*., 2013a; Irannejad *et al*., 2016; Pourasghar *et al*., 2012, 2015, 2019). The impacts of the warm and cold phases of ENSO on the inter-annual variability of Iran's precipitation and temperature have been investigated more than other teleconnections in the literature existing for southwest Asia domain including Iran. The average response of precipitation to El Niño and La Niña phases of ENSO is investigated in several studies (Barlow *et al*., 2002; Nazemosadat and Ghasemi, 2004; Hoell *et al*., 2017a; Alizadeh-Choobari *et al*., 2018). A consistent finding between previous studies is that wet and dry conditions are accompanied respectively by El Niño and La Niña (Nazemosadat and Ghasemi, 2004; Barlow *et al*., 2016; Rana *et al*., 2017, 2019; Hoell *et al*., 2017a, 2017b, 2018; Alizadeh-Choobari *et al*., 2018). The range of response is reported by Hoell *et al*. (2018) which concluded that precipitation of boreal cold season (November–April) in Southwest Asia is affected by central Pacific (CP) and eastern Pacific (EP) of El Niño and La Niña in Southwest Asia. The wintertime upper-tropospheric westerly jet stream over subtropical east Asia and the western Pacific (East Asia jet stream-EAJS) is influenced by Maritime Continent convection. It has been identified as a physical mechanism linking western Pacific Ocean conditions with central Southwest Asia winter precipitation (Tippett *et al*., 2003) in such a way that there is a strong correlation between EAJS and negative precipitation anomalies. Alizadeh-Choobari *et al*. (2018) discussed that 26 % of variance of Iran's precipitation is due to the ENSO signal. The CP El Niño has been argued to cause wet years in central, eastern, southwestern, southern and southeastern regions, explained by more amplified Rossby waves moving more slowly, while more rapid movement of Rossby waves during La Niña has resulted in less extreme weather which is anomalously dry (reference needed). Ahmadi *et al*. (2019) evaluated the effect of 34 teleconnection patterns (indices) on Iran's precipitation and their results demonstrated that Southern Oscillation Index (SOI) has a significant correlation with Iran's mean precipitation in three months of October, November, and December. Moreover, the combination of La Niña and predominant southeasterly wind on the Indian Ocean bring moisture from this region to the Middle East and Iran and the combination of La Niña and prevailing northwesterly wind on the Indian Ocean will increase the risk of drought in Iran. Large deviations are observed in time series of observed precipitation mainly linked to moderate/strong El Niño and La Niña events. For example, in some regions, extreme wet conditions (> two standard deviation) is observed during 1994–1995 El Niño. The first leading mode of winter precipitation (EOF-1) is significantly linked to ENSO characterized by a mono-sign pattern according to findings by Rana *et al*. (2019). Simulations based on a large ensemble of atmospheric models (three models, 42 members) forced by 1901–2012 time varying boundary conditions show the association of wet years with El Niño throughout the whole period and exceptionally dry conditions with La Niña after 1970 (Hoell *et al*., 2017a). According to their study, the long-term changes of precipitation in southwest Asia are influenced by a strong anomaly contrast between the western Pacific Ocean and the central Pacific Ocean. A review of regional drought mechanisms was provided by Barlow *et al*. (2016). The study found considerable potential predictability of drought linked to the tropical oceans, snowmelt and Madden Julian Oscillation (MJO) among other factors. In another study, extreme droughts during 1948–2012 were investigated using a physical model which found climate variability over the Pacific Ocean to be associated with several drought events (Hoell *et al*., 2017b). The Indian Ocean Dipole (IOD) has also been found to have a significant role in the inter-annual variability of precipitation in the southern part of Iran (Pourasghar *et al*., 2012).

For disseminating climate forecast information to end-users, the sources of uncertainties in GCMs should be quantified. A common strategy is to develop MMEs by capitalizing on the complementary strengths of multiple models and improving the estimate of the forecast distribution by increasing the ensemble size (Weigel *et al*., 2008; DelSole *et al*., 2013; ECMWF Working Group Report, 2016). Recent MME studies conducted have revealed the superior prediction skill of multiple models compared to individual single models (Batté and Déqué, 2011; Kirtman *et al*., 2014; Ma *et al*., 2015; Pegion, 2015; Shukla *et al*., 2016; Wanders and Wood, 2016; Ehsan *et al*., 2017; Vigaud *et al*., 2017).

Based on the existing literature, mostly weak and moderate skills have been found in seasonal forecasting of extratropics and Europe where the signal to noise ratio is low (see, for example, Frías *et al*., 2010; Ehsan *et al*., 2017; Manzanas *et al*., 2017). Considering barriers and challenges to implement tailored climate forecasts at the operational level (see Vogel and O'Brien, 2003; Bolson *et al*., 2013), reliable, and skillful forecast information can support sectoral applications with significant socioeconomic added values. In this study, a framework is proposed to evaluate seasonal climate forecast over

Iran as a semi-arid region located in Southwest Asia. It is expected that the suggested framework provides better insights into the potential opportunities and limitations of seasonal forecasting for tailoring in water resources management. Applying MOS to dynamic model outputs and linking the information from GCMs to climate data relevant for application sectors can help in hydroclimate modelling and all managing sectors that are impacted by interannual climate variability. The proposed framework has the potential for application to other regions and is capable of providing information at spatial resolutions appropriate for regional planning and to produce input data required for running hydrological and agricultural crop models. The next section provides an overview of the study area. Section 3 describes the methodological framework. Section 4 presents the main findings and mapping results followed by the conclusion of this article in Section 5.

2 | STUDY AREA

The study area is Iran located in Southwest Asia and some parts of adjacent countries (44° and 64° longitude and 24° and 40° latitude) which is shown as the inner box of Figure 1. The complex orography of Alborz and Zagros (two mountain ranges in the North and West of Iran) together with two desserts of Dasht-E-Kavir and Dasht-E-Lut in the central part of Iran plays an important role in climate variability in different parts of the region under study. An extensive part of area under study lies within the subtropical high pressure belt during summer, leading to hot dry summers in almost entire Iran. The area between the Alborz Mountains and the Caspian Sea experiences a humid subtropical climate (Molavi-Arabshahi *et al*., 2016), while other areas have mainly arid and semi-arid climates. The water year starts in the late September within the study region, and more than 70% of the annual Iranian precipitation falls from October to March, associated with eastward-propagating midlatitude baroclinic waves that form over the Mediterranean Sea in winter. Seasonal forecasts have the potential to help better plan and allocate water resources to the agriculture sector within the rainy season.

Trends of hydroclimate variables in addition to extreme event analysis have been the subject of recent studies targeted the area under study (e.g., Tabari *et al*., 2011; Soltani *et al*., 2016). Investigating extreme hydroclimate features provides a better understanding of changes in seasonal, interannual and inter-decadal variability significant for sectoral decision-making and regional planning. Analyses of the annual mean maximum and minimum temperature during 1975–2010 shows an increase (less than 0.06°C/decade) while for total precipitation, a considerable spatially coherent trend was not found for precipitation (Soltani *et al*., 2016). The spatial distribution of variance of precipitation and temperature in the region during the fall and winter seasons is provided in Figure 1. Note that Iran's precipitation variability is influenced largely by the mid latitude climate systems and complex orography of the two mountain chains of Alborz and Zagros positioned around the north and west of the country (Raziei *et al*., 2013b; Kiani *et al*., 2019). The highest seasonal precipitation variance is on winter and fall (Sabziparvar *et al*., 2015; Ahmadi *et al*., 2018). The pattern for precipitation variance is relatively similar in OND and JFM seasons showing a larger variance in the west of Iran. It is explained by topography along the Zagros Mountains slope (Alijani, 2008; Raziei *et al*., 2013b). The variability of temperature in the two seasons is different while the pattern for precipitation is relatively similar and explained by topography along the Zagros Mountains slope. The mean winter 250 hPa geopotential height anomaly over 50–60°E, 40–50°N minus that over 0–10°E, 45–55° N has found to have significant correlation with Iran's mean winter temperature (Irannejad *et al*., 2016). The larger variance in JFM mean near surface temperature is likely to be linked with the mentioned teleconnection.

3 | DATA AND METHODOLOGY

3.1 | Precipitation and temperature observed datasets (1983–2013)

Two observational monthly gridded precipitation and temperature datasets are used. Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks Climate Data Record (PERSIANN-CDR) is used as the reference data set for precipitation. PERSIANN-CDR is adjusted to the Global Precipitation Climate Project (GPCP) 0.25° monthly dataset during the merging procedure (Ashouri *et al*., 2015). Recent studies have shown a reasonable performance of monthly and annual precipitation of PERSIANN-CDR for Iran, based on a dense network of 2,100 precipitation gauges (Katiraie-Boroujerdy *et al*., 2016, 2017).

Data for temperature was obtained from the Climatic Research Unit (CRU), which provides a monthly gridded temperature on high spatial resolution ($0.5 \times 0.5°$) grids (Harris *et al*., 2014). In previous studies, CRU's monthly temperature has shown to have an accurate estimate when compared to 88 synoptic stations all over Iran (Miri *et al*., 2017). Miri *et al*. (2017) found that, except for the coastal line of the Caspian Sea in the north of the country, CRU performs well over Iran (RMSE ≤4°C, Pearson

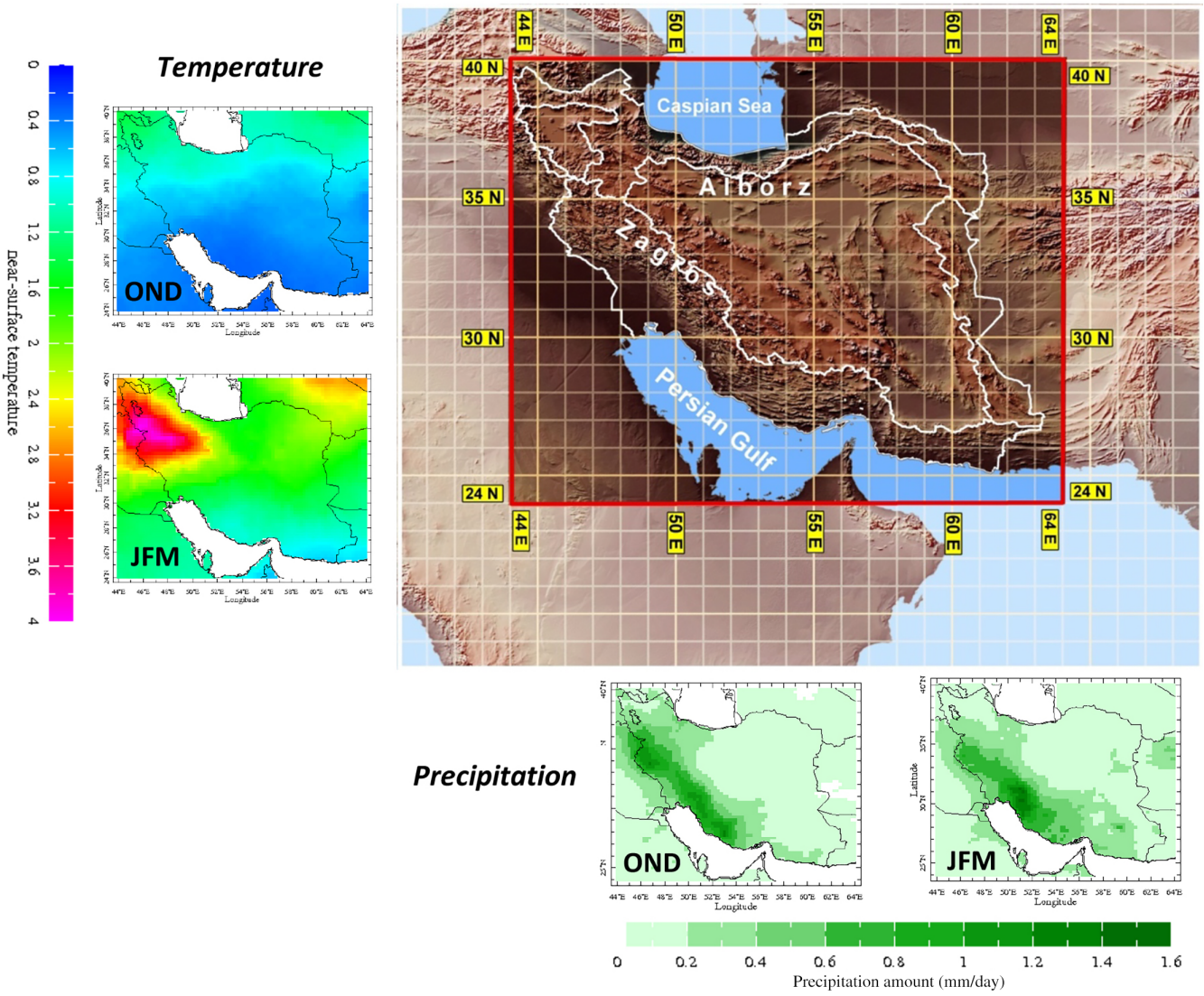

FIGURE 1 The area under study; inner box—Iran included, and time variance of temperature (CRU) and precipitation (PERSIANN-CDR) for OND and JFM seasons during 1983–2013 is also presented [Colour figure can be viewed at wileyonlinelibrary.com]

Correlation ≥.9). In this study, two seasons are selected as target seasons, October–December (OND) and January–March (JFM).

3.2 | Canonical correlation analysis

Empirical and hybrid forecasting models were constructed in this study using CCA. In CCA, the multitude of a considerable number of correlations in the correlation matrix is reduced to the small number of variables derived (Barnett and Preisendorfer, 1987). Linear combinations between variables x in set A and y in set B, are considered in the way that the correlation between their time series is maximized. Canonical correlations (ρ) are found for the linear combination (canonical variates) of $Z_t = \bar{a}_t x$ in set A and the linear combination of $W_t = \bar{b}_t y$ in set B from Step 1 to t. In each step, correlation is at its maximum value under the assumption that Z_t and W_t in step t are uncorrelated with Z_{t-1} and W_{t-1} (canonical variates are uncorrelated). For example, at the first step, $\rho_{1,1} = Corr(Z_1, W_1)$. In other words, given I components in set A (I>1) and J components in set B (J>1), t = min (I, J) and that the minimum (maximum) number of CCA modes cannot exceed the minimum (maximum) number of set A modes or that of B. Using eigenvectors in each step, t canonical correlations are obtained corresponding to t calculated canonical variates. Fitting CCA, the amount of noise is reduced by truncating subset of EOF coefficients (eliminating the higher EOF modes), which make it interesting for hydroclimate studies (e.g., seasonal forecasting in a

spatial domain) where t may be reasonably large. The minimum and the maximum allowable number of modes for both set A (predictor) and set B (predictand) are considered 1 and 15, respectively.

Forecasting models are trained using the cross-validated calculation. In this process, optimal truncation of principle components (PCs) and CCA modes are determined. The optimal forecasting model is constructed based on the highest goodness index given the optimum number of modes (for the predictor, predictand, and CCA). Cross-validated forecasts are produced for each year during the training period based on standardized anomalies that are calculated for precipitation and temperature. Goodness index (area-average Kendall's τ) is maximized through 1-year leave out cross-validation. This can be interpreted as the average correlation between the cross-validated forecasts and observations.

Precipitation data are transformed to a normal distribution. Probabilistic forecasts of precipitation and temperature are derived from the best-guess forecasts (regressions estimated from the CCA), and distribution of the errors using Climate Predictability Tool (Mason and Tippett, 2017). Seasonal precipitation forecasts are then transformed back to the original distribution by applying the inverse process. In this process, the cross-validated correlation is between the transformed forecasts and observations. Assuming that the errors are normally distributed, the variance of errors is defined by sampling errors in the regression parameters.

3.3 | Empirical approach

In this study, observed antecedent Sea Surface Temperature (SST), Mean Sea Level Pressure (MSLP), soil moisture and geopotential height at 850 hPa level were selected as predictor fields. Predictand is observed precipitation/temperature over the study area. For example, EOFs of the September SST values over several domains are used as the input of CCA (predictor) to forecast precipitation/temperature anomalies for target seasons. Table 1 provides a summary of variables and datasets used for the empirical approach.

3.4 | Hybrid model (dynamical-statistical approach)

Coupled atmospheric-ocean GCMs solve the Navier–Stokes equations to simulate physical mechanisms responsible for non-linear interactions between earth components. It is expected that the state-of-the-art dynamic seasonal forecasting systems represent the important mechanisms contributing to seasonal to inter-annual climate variability at the regional scale. MOS is applied to correct the GCMs' systematic errors and to derive post-processing equations that can be applied to future runs of similar forecast model (useful for real-time forecasting). Precipitation and temperature forecasts from this hybrid dynamical-statistical approach can then be used to force-land surface models for streamflow and soil moisture operational seasonal forecasting. In this study, a hybrid approach is developed by applying CCA to both individual seasonal forecasting systems presented in Section 3.4.1, in addition to their weighted combination provided in Section 3.4.2.

3.4.1 | NMME individual seasonal forecasting system

NMME is an experimental project based on collaboration and coordination of several modelling centres from the United States and Canada (Kirtman *et al*., 2014). Several atmosphere–ocean coupled models are contributed to the NMME project to produce monthly climate predictions with lead times up to 12 months ahead. The project is an effort to meet information accessible to all users for the specific tailored regional needs and supports required for real-time decision making. This article takes operational seasonal forecasting models participated in Phase-I of the project. The seasonal climate systems in the NMME project which are used in this study include NOAA National Centers for Environmental Prediction (NOAA/NCEP), NOAA's Geophysical Fluid Dynamics Laboratory (NOAA/GFDL), and models developed by the Environment Canada, and are provided in Table 2.

3.4.2 | MME

MME is an approach which helps to improve the skill and quantify uncertainties associated with seasonal forecasting systems. It is expected that both increased model diversity and greater ensemble members improve the quality of forecasts (Doblas-Reyes *et al*., 2000). The MMEs in this study consist of eight models from NMME (provided in Table 1). The ensemble mean from each model's retrospective forecasts is averaged over the period 1983–2013 and this model climatology is then subtracted from the year-to-year ensemble mean for each target season to correct for mean biases. Then, equal weights are assigned to the resulting anomalies of each model's

TABLE 1 Predictand variables for empirical forecasting models in this study

Variable	Description/Project	Center	Resolution (degree North and East)	References
Mean Sea Level Pressure (MSLP)	Monthly intrinsic MSL pressure from NCEP-NCAR CDAS-1: Climate Data Assimilation System I; NCEP-NCAR Reanalysis Project	US Weather Service—National Met. Center	2.5	Kalnay *et al.* (1996)
Geopotential Height		US Weather Service—National Met. Center	2.5	Kalnay *et al.* (1996)
SST	Extended Reconstructed Sea Surface Temperatures (ERSST) from NOAA NCDC ERSST version3b	NOAA NCDC	2	Smith *et al.* (2008); Xue *et al.* (2003)
Soil Moisture	NOAA NCEP CPC Global Monthly Soil Moisture: (Global monthly high-resolution soil moisture)	Climate Prediction Center	0.5	Huang *et al.* (1996); and van den Dool *et al.* (2003)

TABLE 2 Brief descriptions of the NMME individual seasonal forecasting systems used in the present study

Model name	Atmospheric component resolution (Degree)	Ensemble member	Forecast Lead time (month)	Center	Reference
CFSv2	0.9	24	0.5–9.5	National Centers for Environmental Prediction (NOAA/NCEP)	Saha *et al.* (2014)
CMC1-CanCM3	2.5	10	0.5–11.5	Environment Canada's Canadian Meteorological Centre (CMC)	Merryfield *et al.* (2013)
CMC2-CanCM4	2.5	10	0.5-11.5	Environment Canada's Canadian Meteorological Centre (CMC)	Merryfield et al. (2010)
CM2.1-aer04	2	11	0.5–11.5	NOAA/GFDL	Zhang *et al.* (2007)
CMp2.5 (FLOR_A06)	0.5	11	0.5-11.5	NOAA/GFDL	Zhang et al. (2007)
CMp2.5 (FLOR_B01)	0.5	11	0.5-11.5	NOAA/GFDL	Zhang et al. (2007)
RSMA-CCSM4	0.9 × 1.25	11	0.5–11.5	COLA/NCAR	Gent *et al.* (2010)
NASA-GMAO	1 × 1.25	12	0.5–8.5	NASA	Vernieres *et al.* (2012)

ensemble mean output to make the MME ensemble mean. CCA is then applied to this MME output to forecast precipitation and temperature.

3.5 | Verification scores

In this study, continuous and categorical skill maps are presented. Continuous scores include the Spearman correlation, Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). In addition to Spearman correlation, model performance is also provided based on area-averaged Kendall's τ rank correlation. Both Spearman and Kendall's τ correlations measure monotonicity relationships. However, Kendall's τ is selected as the goodness index in this study, since the discriminatory power of the forecasts is an important aspect for decision-making especially when seasonal forecasts are provided in categories. It is not sensitive to the distribution of the data and is more tractable in cases when data have tied ranks (Gilpin, 1993). Kendall's τ is interpretable as the percentage of pairs of data points that show a positive correlation.

Suggested by WMO's Standardized Verification System for Long-Range Forecasts (SVSLRF) verification, the area under Relative Operating Characteristics curve (ROC) is also calculated for each grid point as a certain measure for both hybrid and empirical forecasting approaches. ROC gives the proportion of times that a specific condition (e.g., above-normal/below normal conditions) is distinguished successfully from the other categories. Threshold values define the above- and below-normal categories calculated as the terciles from the climatological period. For time series of predictand (observations) and predictors, standardized anomalies are calculated for every grid cell during 1983–2013.

4 | RESULTS

4.1 | Empirical approach

4.1.1 | Forecasting precipitation

As discussed in previous sections, seasonal precipitation in the study area during fall and winter season is greatly influenced by Pacific and Indian Ocean SST. The next section constructs empirical forecasting models for OND and JFM precipitation.

a. OND

Several lags and combinations of predictor fields are tested for developing empirical relations between Iran's precipitation and large-scale patterns. The association between west Asia precipitation, ENSO teleconnection together with tropical and extra-tropical SST have been reported in previous studies, most strongly during October to December (Nazemosadat and Shahgholian, 2017). In agreement with previous studies, observed SST in August–September is found to be the best independent variable among other predictors for forecasting OND precipitation in the empirical approach. Selecting SST among other fields leads to the highest skill of empirical models. Figure 2 shows the leading two CCA modes of SST (Lat: 48 N–26S Lon: 84E–90 W) and Iran's precipitation which is used to construct the empirical forecast model for OND precipitation. More than 50% of total variance of OND precipitation is explained by the first mode. The correlation between precipitation average in study area and SST in the Tropical Pacific (not shown) during 1983–2013 is found to be significant ($r = .60$). Figure 6a shows that the Spearman correlation is between .3 and .45 in almost all grid points except those located in the southeastern part of the study area (in 5% significance level [$\alpha = .05$], the absolute critical value for Spearman's ρ is .36). Except for the northwestern part, OND forecast skill is high ($\rho \geq .45$ and ROC ≥ 0.7) in the northern part of the country (see Figure 6a–c).

b. JFM

The results show that the empirical approach generally have lower skill in forecasting precipitation in JFM (Figure 6d–f) than OND (Figure 6a–c) over Iran, with skill confined to the southwest of the domain. The average of MSLP for two months of November and December is recognized as the best predictor (Kendal's $\tau = 0.253$) in JFM. Table 3 illustrates how Kendal's τ and CCA modes change by predictor and domain selection in constructing empirical models for forecasting JFM precipitation. Calculated RMSE is between 0.3 and 0.7 mm/day for most of the grid points in both target seasons. The calculation method of the modes provided in Table 3 for X, Y and CCA is explained in Section 3.2. Temporal scores for JFM is shown in Figure 3.

4.1.2 | Forecasting temperature

a. OND

Average SST in September is selected as a predictor for forecasting OND temperature. The leading CCA modes for predictor and predictand in addition to temporal scores are shown in Figure 4. Kendall's τ of the best empirical model to forecast is 0.209. The Spearman correlation is less than .3 in almost all grid cells except the southwestern part (Figure 6g). The skill in terms of ROC is high in the Southwest part (Figure 6h,i).

b. JFM

For JFM, observed SST in November and December is the best-selected predictor among others (see Figure 5). Positive winter temperature has been reported by Walker and Bliss (1932) to be associated with El Niño. Kendall's τ of the best empirical model to forecast JFM temperature is 0.315. The Spearman correlation (Figure 6j) is higher in JFM (between .3 and .45 countrywide and more than .60 in southern Iran) in comparison to OND. The skill of empirical models for forecasting temperature looks more similar between the OND and JFM seasons than in the case of precipitation.

To test the stability and robustness of the CCA results beyond the leave-one-out cross-validation, we run CCA (1) by removing the trend component and (2) by additional split-in-two sensitivity tests. For the

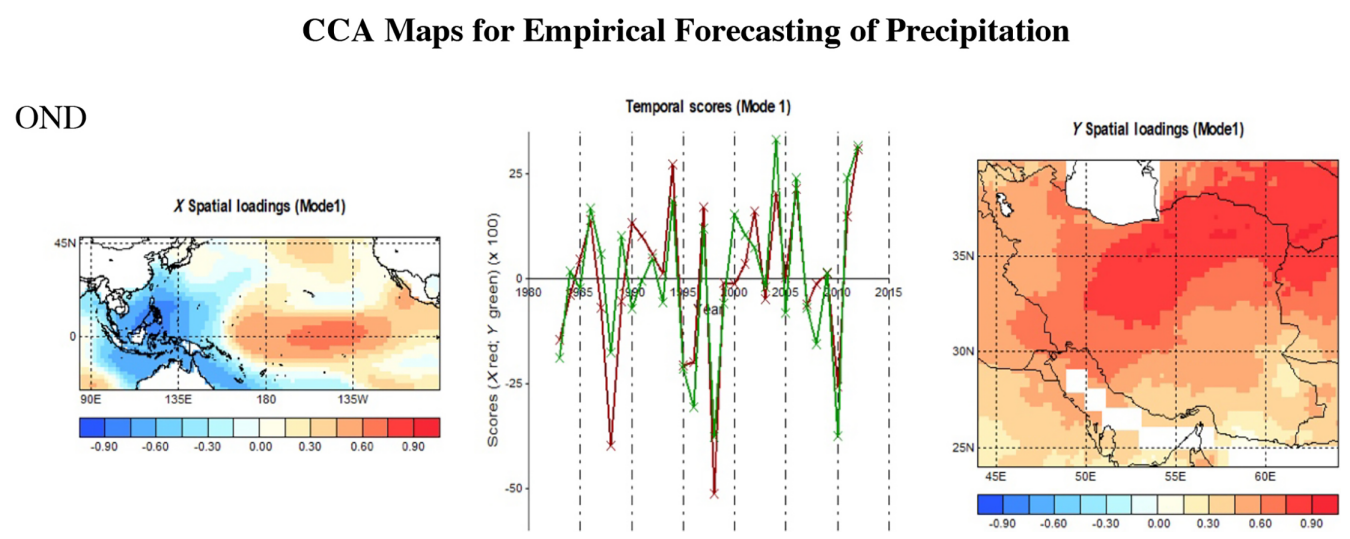

FIGURE 2 Spatial loading of predictor (averaged SST for August–September) in left, temporal scrores of the fiest CCA mode (middle) and predictand (OND precipitation) in right [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Performance of the effect of different predictor selection in precipitation forecasting results of empirical models

Predictand	Predictor	Domain	Modes (*X*, *Y*, CCA)	Selected month(s) for the predictor	Goodness index
JFM Precipitation (PERSIANN-CDR)	MSLP	Whole globe	7, 1, 1	November–December	0.213
			8, 1, 1	December	0.224
		Lat: 30 N–30 S, Lon: 60E–155 W	1, 1, 1	November–December	0.253
			1, 1, 1	December	0.223
	SST	Whole globe	5, 2, 1	November–December	0.184
			10, 1, 1	December	0.166
	Geopotential Height (850 HPa)	Whole globe	10, 1, 1	November–December	0.205
			9, 1, 1	December	0.199
	MSLP and Geopotential Height (850 HPa)	Whole globe	6, 1, 1	November–December	0.208

Note: The goodness index is considered as the average Kendall's τ correlation between the cross-validated empirical forecasts and observations.

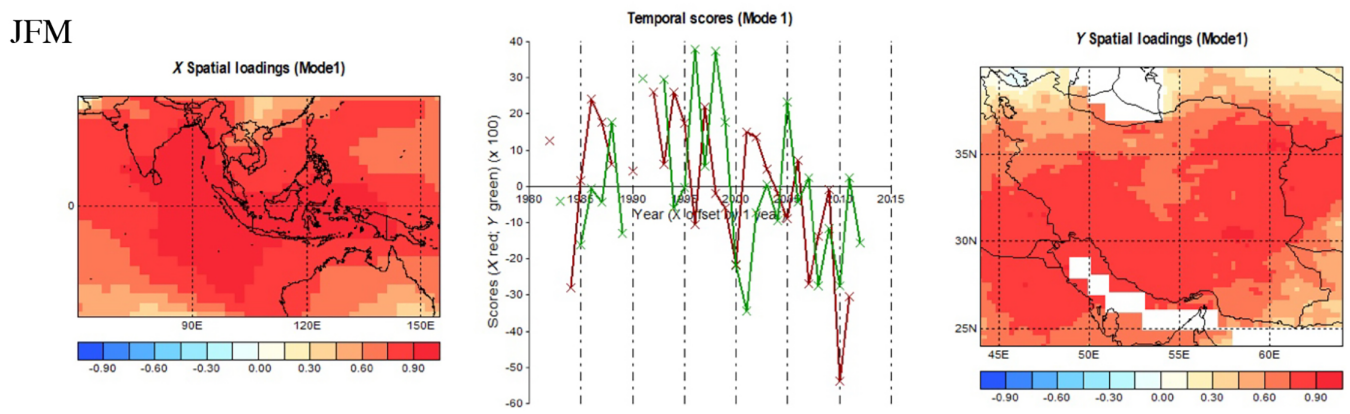

FIGURE 3 Spatial loading of predictor (averaged MSLP for November–December) in left, temporal scrores of the fiest CCA mode (middle) and predictand (JFM precipitation) in right [Colour figure can be viewed at wileyonlinelibrary.com]

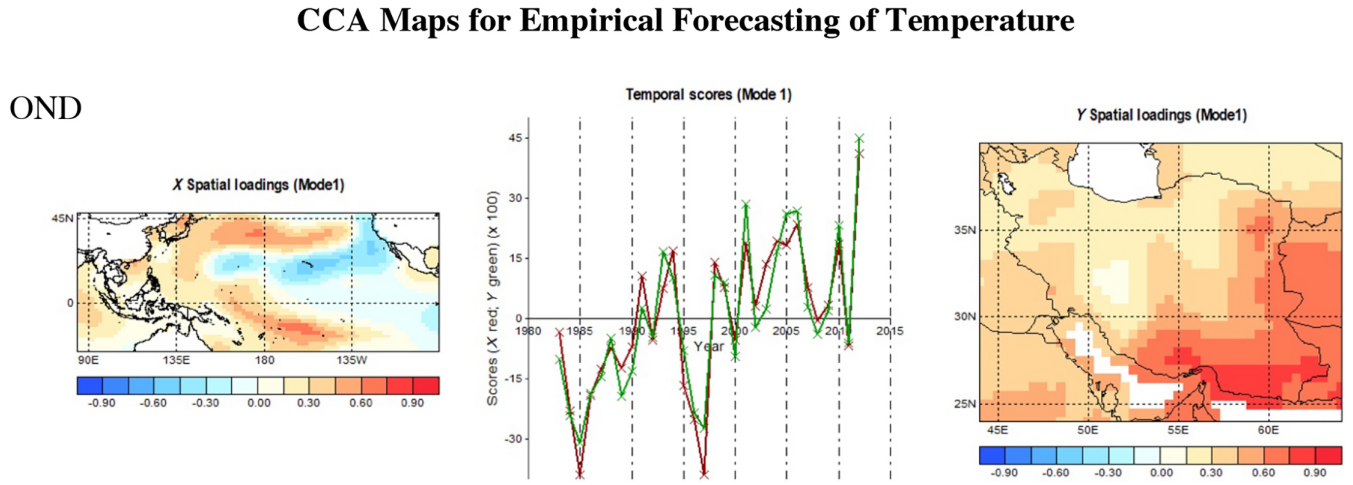

FIGURE 4 Spatial loading of predictor (September SST) in left, temporal scrores of the fiest CCA mode (middle) and predictand (OND temperature) in right [Colour figure can be viewed at wileyonlinelibrary.com]

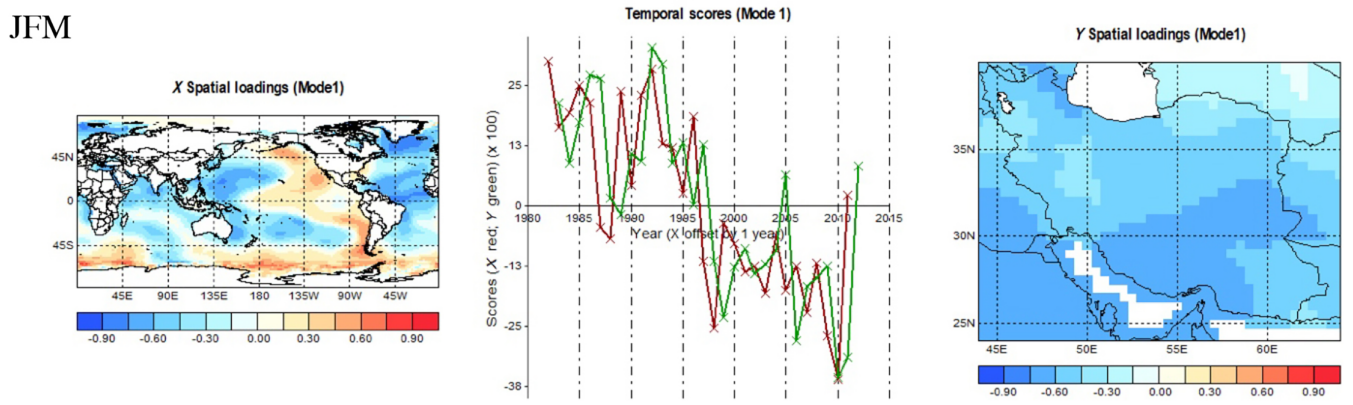

FIGURE 5 Spatial loading of predictor (November–December SST) in left, temporal scrores of the fiest CCA mode (middle), and predictand (JFM temperature) in right [Colour figure can be viewed at wileyonlinelibrary.com]

trend analysis, it was found that trend component in empirical approach has contributed to high skill for temperature forecast. By detrending, the value of goodness index in empirical forecast of near-surface air temperature is decreased from 0.209 to 0.155 for OND and from 0.315 to 0.146 for JFM. However, a significant impact in terms of precipitation forecasting skill was not found by removing trend (see figures in the Supporting Information). Therefore, trend component could contribute partially to empirical forecast of air temperature. We also apply CCA for the first and last halves of the record, separately. Based on additional analysis, CCA performance is considered acceptable by capturing the variability for the first period (1983–1997) and second period (1998–2013). In addition, it was found that spatial patterns of the leading modes for the predictor (X) is quite similar for two halves.

4.2 | Hybrid approach

4.2.1 | Raw GCM output validation

To demonstrate the usefulness of MOS within the framework of developing hybrid models, we provide one example of comparing the skill of the GCM raw model output before and after applying the hybrid approach. It would be instructive to know the extent to which the forecast skill is improved by applying hybrid approach. Previous studies show that it improves climate model outputs by removing the systematic error (spatial pattern, mean, and amplitude) from forecast grid (Mass *et al.*, 2008). Therefore, Figure 7 shows a comparison of Spearman rank correlation and ROC (below and above normal) for CFSv2. The skill of CFSv2 is calculated for OND precipitation initialized in September. Based on Figure 7,

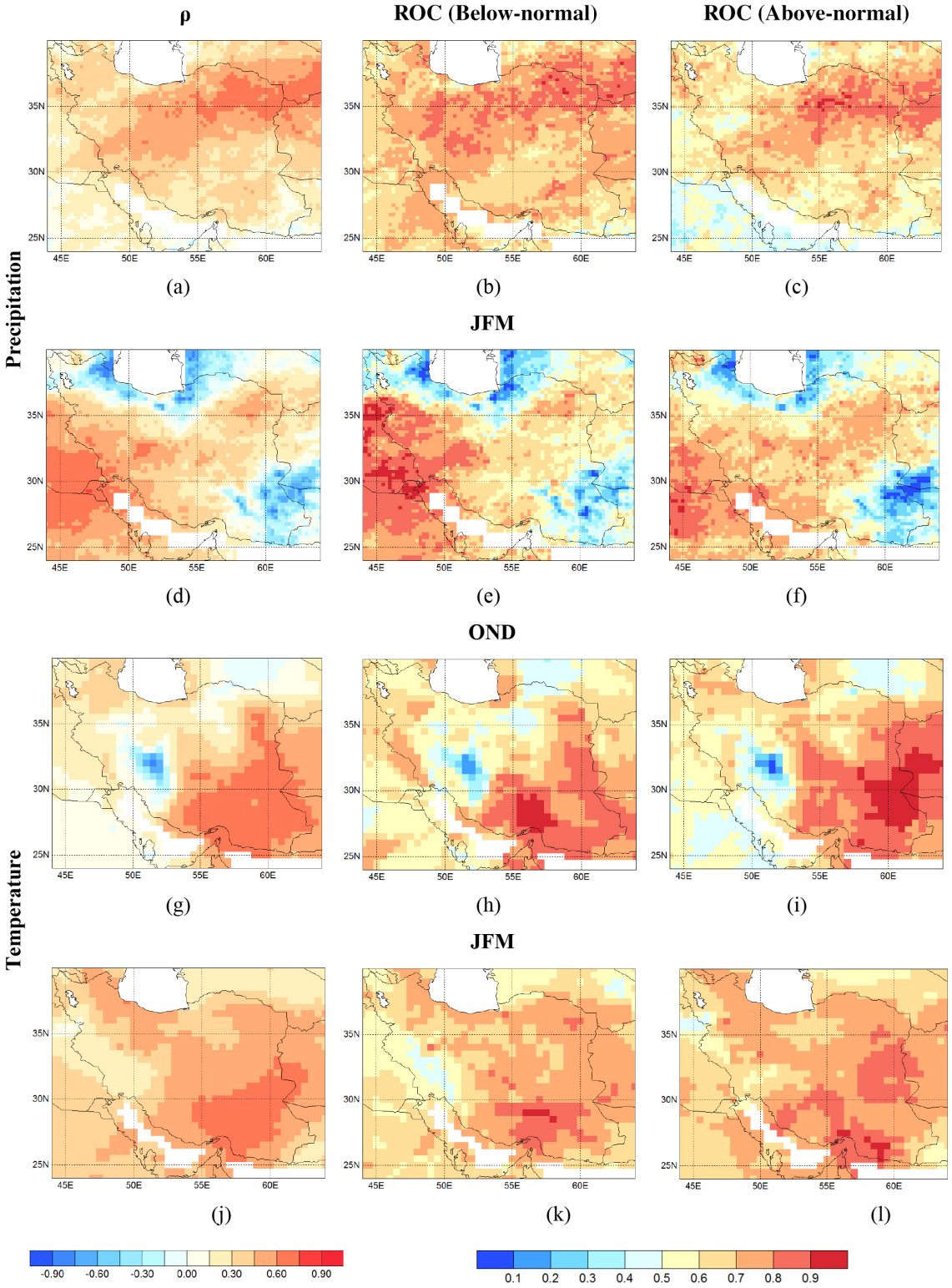

FIGURE 6 Empirical forecast skill of precipitation and temperature for OND and JFM. Predictands are PERSIANN-CDR for precipitation and CRU for temperature. Predictors are average SST for August and September (upper panel), average MSLP for November and December (second panel), average SST for September (third panel), and December (lower panel). Tercile categories are set at 33% and 66% based on the climatological distribution for below- and above normal, respectively [Colour figure can be viewed at wileyonlinelibrary.com]

Spearman correlation of CFSv2-corrected forecast (Figure 7d) shows improvement when compared to its raw model output (Figure 7a) especially in the eastern part. The skill improvement in terms of ROC is also significant as for all grid points, (ROC ≥0.5). Sections 4.2.1 and 4.2.2 provide results of the hybrid approach for individual and MME forecasts at 0.5 and 1.5-month lead times.

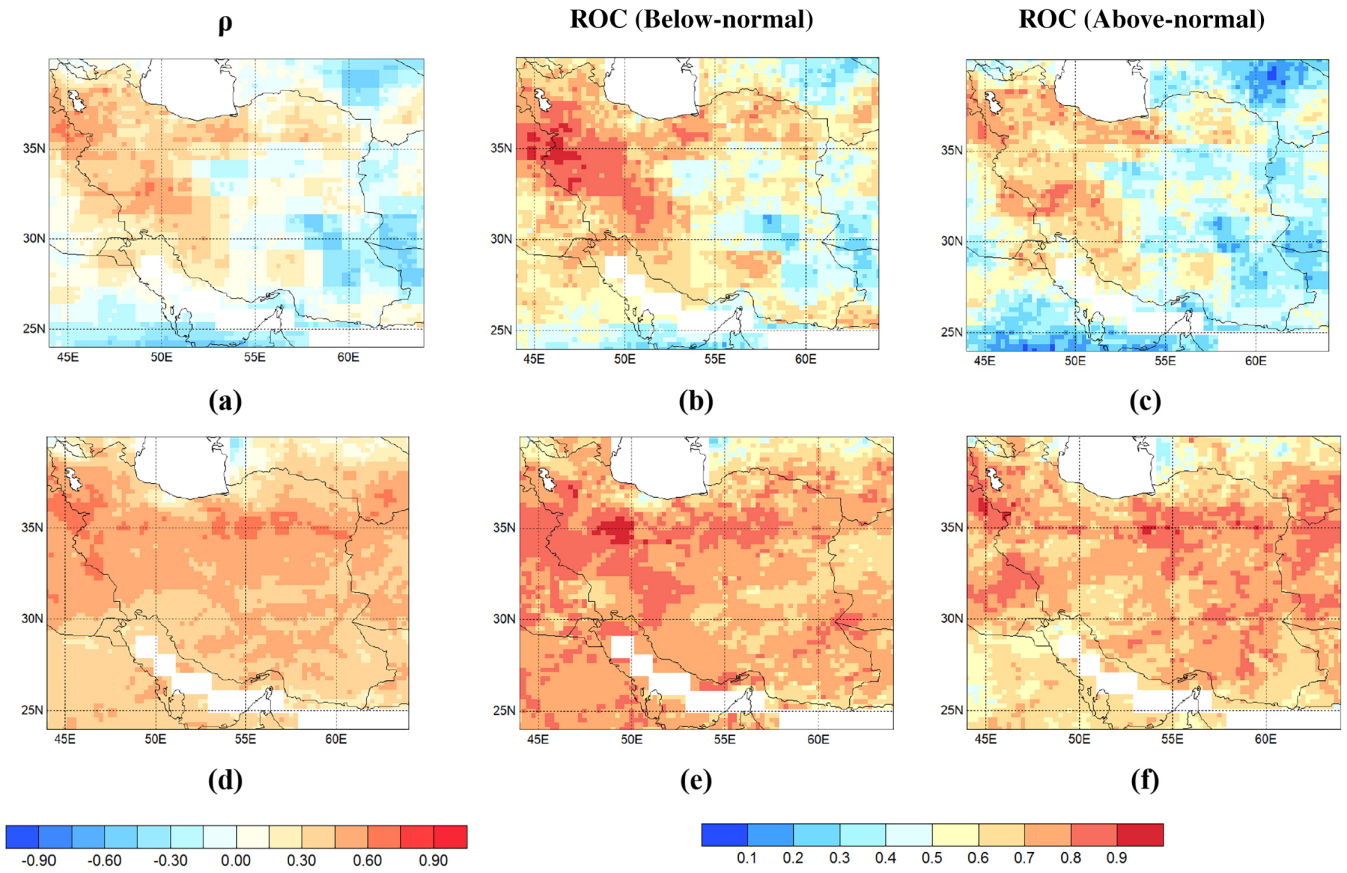

FIGURE 7 CFSv2 skill comparison for forecasting OND precipitation; raw model skill maps are shown in the top panel, hybrid skill maps are shown in the bottom panel [Colour figure can be viewed at wileyonlinelibrary.com]

4.2.2 | Individual models

Precipitation

a. OND

In general, the hybrid approach has successfully captured interannual variability of precipitation in both below and above normal conditions in OND (Figure 8). However, the magnitude for some extreme wet years is still underestimated. The precipitation skill is shown for the two best individual NMME models in Figure 8. The CMC2 (forecasts initialized in early September, Kendall's $\tau = 0.397$) and GFDL-aer04 (forecasts initialized in early October, Kendall's $\tau = 0.388$) have the highest skills among all individual models for forecasting OND precipitation (Figure 8a–c, and Figure 8d–f). Based on Figure 8b,e, both models perform better in forecasting below-normal conditions considering the number of grid cells where ROC $\geq$0.8. GFDL-aer04 performs the best in distinguishing below normal conditions within this area, which suggests the added value of utilizing seasonal forecasting systems to cope with meteorological droughts. The Spearman correlation of GFDL-aer04 in the west of Iran is spatially similar in 0.5 and 1.5-month lead times whereas it is significantly lower for the CMC2 all over the country. In the southwest of Iran, forecasting precipitation is highly essential for agriculture planning and hydropower generation. As this area contributes to more than one-third of annual surface water resources, the high skill of seasonal forecasting systems can significantly help decision-makers in water resources management. It is found that the skill of four models (CMC2, CFSv2, GMAO-06212, and CMC1) is higher for forecasts initialized in early September. For all models investigated in this study, time series of standardized anomalies for every grid cell were calculated and compared between observations and cross-validated hindcasts during 1983–2013. Large deviations are observed in time series of precipitation mainly linked to moderate/strong El Niño and La Niña events. For example, in some regions, extreme wet conditions (> two standard deviation) is observed

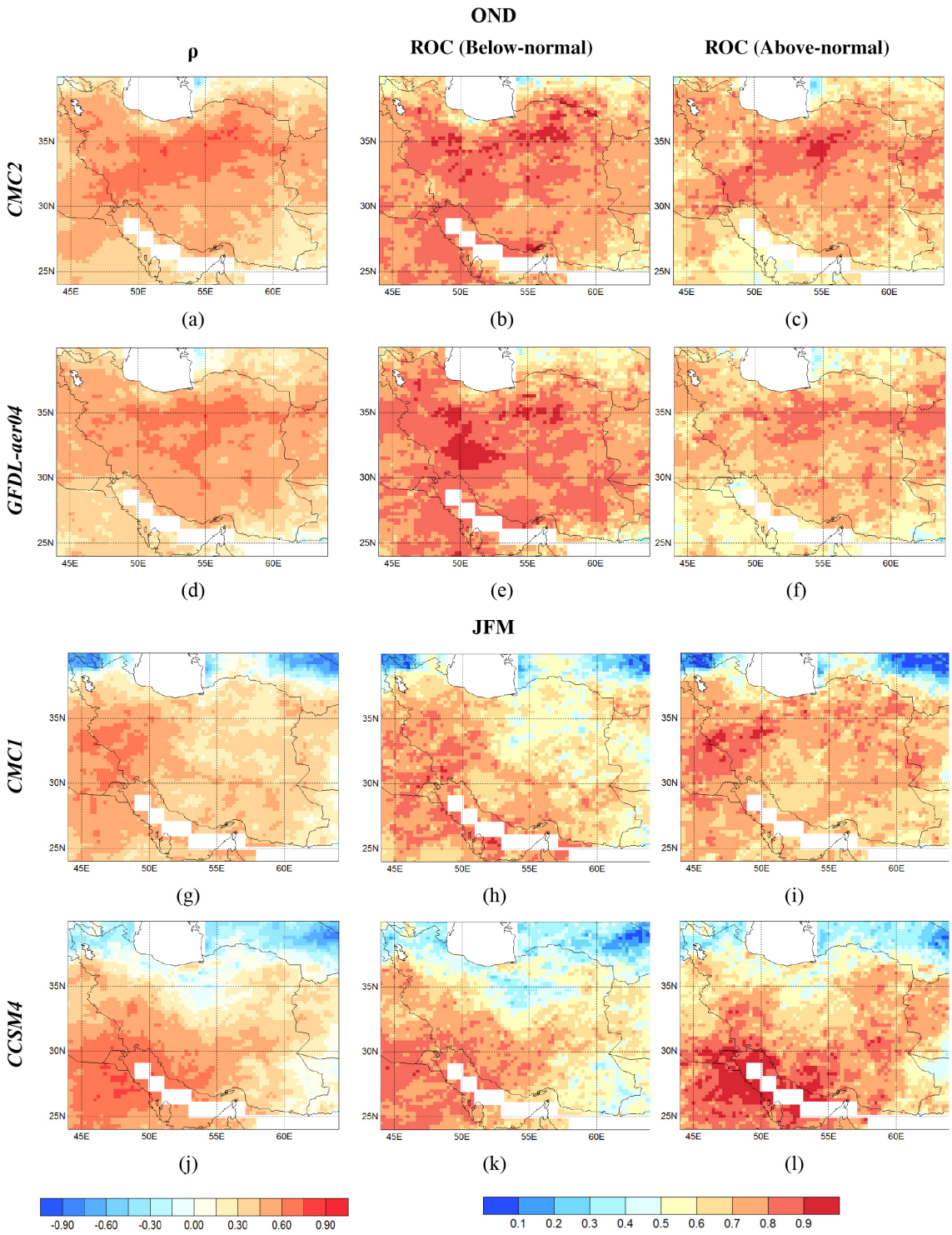

FIGURE 8 Model Skill (Spearman correlation and area under ROC curve) for the two best individual NMME models having the highest skill in forecasting OND and JFM's precipitation. Initializations for CMC2 (a–c) and GFDL-aer04 (d–f) are made in early September and early October, respectively. Initializations for CMC1 (g–i) and CCSM4 (j–l) are made in early January [Colour figure can be viewed at wileyonlinelibrary.com]

during 1994–1995 El Niño. In general, hybrid approach has successfully captured interannual variability of precipitation in both below and above normal conditions. However, the magnitude for some extreme wet years is underestimated by hybrid approach.

b. JFM

Accumulated three-month precipitation in JFM is also important for agricultural water allocation in Iran. When a water year starts with a dry spell (below normal conditions in OND), wet conditions in the second rainy season (JFM) might compensate for water deficits. Therefore, any reliable forecast of JFM's precipitation provides valuable information for stakeholders. However, the results suggest that JFM's precipitation is not as skillful as that of OND in the study area. Unlike OND, NMME's individual model skill is not high neither at 1.5 nor 0.5-month lead times for JFM target season. Spearman correlation of all models is less than .45 for forecasts made in the early December (except for CCSM4 forecasts in the south and southwestern parts). It means the skill of NMME models in JFM is comparable to that of OND target season.

After applying CCA, the highest goodness index in forecasting JFM precipitation at 1.5-month lead time is achieved for CMC1, CCSM4, and CFSv2 models. The performance of CMC1 is better than the other two models in the northern parts. For this model, the Spearman rank correlation (Figure 8g) is less than .6 ($.3 \leq \rho \leq .6$) in the west of the study area between 45 and 51°E (Zagros region). CCSM4 has the same spatial pattern regarding the Spearman correlation to that of CFSv2 in areas close to the Persian Gulf also showing high skill (Spearman correlation ≥.75) in areas located in 25–30°N and 45–50°E (Figure 8j). CFSv2 has skill in the area extended from the southwest to the central parts of Iran (Spearman correlation between .6 and .75 not shown here). Among all models, CFSv2 performs the best in forecasting above-normal precipitation anomalies (ROC ≥0.8 in the southern part of the study area).

Temperature

a. OND

The skills of hybrid approach applied to individual NMME models are shown in Figure 9 for OND temperature. In the northeast and south of Iran, they are significantly higher than other parts of the study area. The Spearman correlation does not exceed .45 for most models at lead times of 0.5 and 1.5 (month). Temperature forecasts are more skillful in the Strait of Hormoz compared to other parts. NMME models perform better in forecasting above-normal temperature anomalies considering the area under ROC curve. FLOR-B01 and CMC2 have the highest goodness index among others (0.344 and 0.306). The MAE is less than 1 and RMSE is between 0.6 and 1.1°C all over the country. For model initializations on the early September, FLOR-B01 (Figure 9a–c) and CMC2 (Figure 9d–f) have higher skill than the others (Kendal's τ >0.2). At the 0.5-month lead time, CFSv2 forecasts standardized temperature anomalies better than the rest of the models (not shown here). Some models (CMC1, CMC2, GFDL-aer04, and GMAO-06212) also have a negative Spearman correlation in the west of Iran, where temperature variation has a high impact on hydropower plants and crop growing.

b. JFM

The skills of JFM's temperature forecasts are higher than OND. NMME models perform better at the 0.5-month lead time for forecasting JFM's temperature. The skill of CMC2 and GMAO-06212 is significantly higher at the 1.5-month lead time compared to 0.5-month lead time. The optimal goodness index of four models (CFSv2, CCSM4, GMAO-06212, and CMC1) is higher than 0.4. Figure 9 shows Spearman correlation and ROC (above and below normal) for two individual NMME models that have the highest Kendall's τ (CFSv2 and CCSM4). After applying CCA, the ROC (above-normal) of CFSv2 exceeds 0.8 in a large number of grid points (Figure 9i). For this model, $.75 \leq \rho \leq .9$ in the central and eastern parts of the study area (Figure 9g). For forecasts made in the early January, Spearman correlations of CCSM4 and NASA-GMAO-06212 are greater than 0.45 for the entire domain. However, CCSM4 has higher skill in the north and northeastern parts of Iran (between 0.75 and 0.9). Comparatively, the Spearman correlation is higher in the south of the study area (Strait of Hurmoz) for NASA-GMAO-06212. On the longer lead times (forecasts made in early December), the maximum of Kendall's τ is calculated for FLOR-A06, and CMC1 equal to 0.355 and 0.336, respectively. By applying MOS for spatial correction, CFSv2 outperforms GFDL-FLOR-B01 in the east and central parts of the area under study. The calculated RMSE is less than 1.96°C for all models. Similar to OND, models perform better in forecasting above normal temperature anomalies.

4.2.3 | MME

We select several model combinations to compare the performance of the hybrid approach (individual models and MMEs) to the raw outputs of seasonal climate forecasting systems. The aim is to evaluate how MME can improve the forecast skill compared to the results provided in the previous section. Comparison of models is mainly based on the

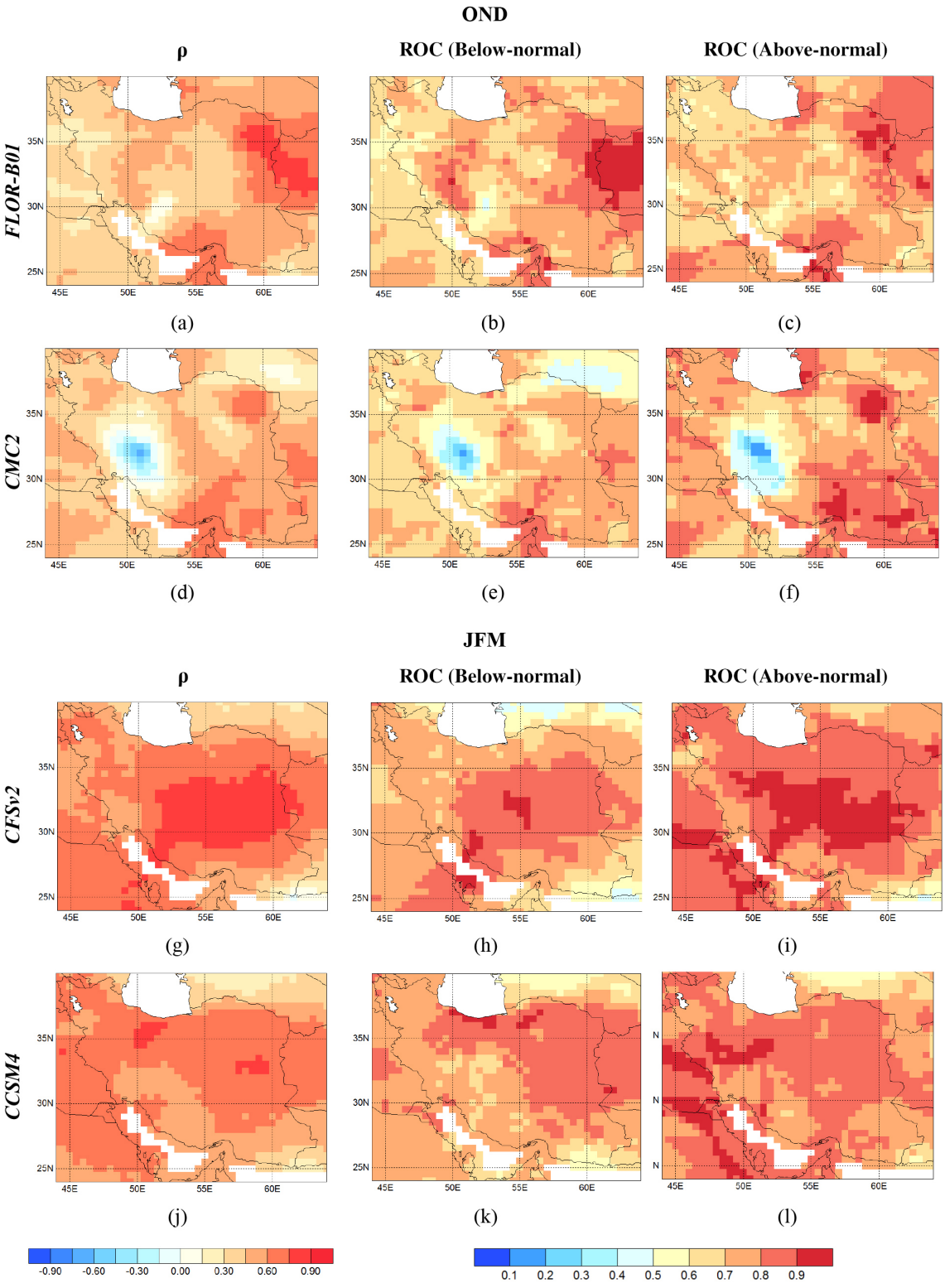

FIGURE 9 Same as Figure 8. But for temperature. For (a–g), initializations are made in early September (1.5-month lead time). For (g–l), initializations are made in early January (0.5-month lead time) [Colour figure can be viewed at wileyonlinelibrary.com]

goodness index although other skill performances are also discussed. To understand the improvements by performing the hybrid, the skill of dynamic models (raw output) is calculated before applying CCA (not shown) and it is found that the precipitation and temperature forecasts are improved under a hybrid-approach. Then, the rank of each individual dynamic model is given based on model performance using the goodness index. Table 4 provides an

TABLE 4 Individual model skill improvement by using hybrid approach and goodness index comparison to raw model outputs for forecasting OND precipitation in 0.5-month lead-time

	Raw models	Hybrid approach	
Model/Skill	Goodness index (Kendall τ)		Rank
GFDL-aer04	0.10	0.39	1
CCSM4	0.12	0.37	2
FLOR_A01	0.05	0.33	3
CFSv2	0.23	0.33	4
FLOR_B06	0.09	0.31	5
CMC2	0.12	0.30	6
CMC1	0.16	0.30	7
NASA	0.18	0.28	8

example of the skill improvement for OND precipitation forecast (0.5-month lead-time) in the target area as an illustration. Based on Table 4, the improvement is achieved for all individual models which suggest CCA has been successfully applied to remove systematic biases. It provides the insights on how correcting systematic error have resulted in overall improvement compared to uncorrected forecast.

To form MMEs, equal weights are assigned to the ensemble mean of individual models (raw outputs). Then CCA was applied to calculate calibrated MME probabilistic retrospective forecasts. Model selection for each MME is based on the rank of each (individual) model in terms of calculated goodness index. The models which perform better (having higher goodness index) are added one by one to form seven different MMEs, consisting of two to eight individual models. Then CCA is applied to remove the systematic error of the MMEs. Another approach is applying MOS to each individual model at the first step and then averaging tercile-category probabilities of MOS corrected models. We used the first approach in the present study.

For the ease of reading, each specific MME is called by a code (e.g., MME_XXX) which includes a digit number (XXX). Each digit corresponds to one specific seasonal forecasting system. The number of digits (after the acronyms of 'MME_') shows the number of individual models that are weighted within that specific MME (MME_XXXXX encompass five different individual models). The model naming/description is presented in Figures 10 and 12. For instance, MME_3481 means that the MME is developed by assigning equal weights (in this case 0.25) to each of the four NMME individual models (CMC2, CM2.1-aer04, GMAO-06212, and CFSv2). While equal weights are assigned to develop each MME, the order of digits shows the models which have the highest goodness index one after the other.

Precipitation

a. OND

In OND, six MMEs out of eight outperform the best individual model (CMC2). Among all models, MME_12348 has the highest goodness index ($\tau = 0.433$). RMSE and MAE for this model are less than 1.10 and 0.91 mm/day. Note that, there are some parts that performances of other model combinations are better according to Spearman correlation and area under ROC curves (e.g., MME_34 in the western part of Iran and the area along the coasts of Persian Gulf). Based on the analysis, MME_34812576 has the lowest goodness index ($\tau = 0.337$), suggesting that adding individual models with a lower skill does not necessarily provide any added value to the MME and might even result in decreasing model skill b. JFM In JFM, MME has higher skill at 0.5-month lead time and outperforms that of 1.5-month lead time. MME_2718, MME_27184, and MME_271843 are outperforming the best individual model ($\tau = 0.323$). Adding CFSv2, GMAO-06212, GFDL-aer04, and CMC2 to the MME_27 increases the Spearman correlation in the northeastern parts. Since the skill of the MME is close to zero in the northern boundary of the study area negative anomaly correlation is calculated as a result of cross-validation bias (Barnston and van den Dool, 1993).

Figure 11 summarizes the highest goodness index for OND and JFM at 0.5 and 1.5-month lead times for individual and MMEs. It shows the skill of individual models against MMEs. Based on Figures 8 and 9, the skill of MMEs decrease when more models are added for OND, and for the two last individual models (GFDL_FLOR_A06 and FLOR_A06) for JFM. It means, adding more models to the MME does not necessarily result in higher skill and is influenced by the skill of which an individual is picked. This result is partially in line with previous findings in similar studies (e.g., Robertson *et al*., 2004).

Temperature

a. OND

In Figure 12, Spearman correlation is provided for different combinations of NMME individual models (seven MMEs) for forecasting temperature in OND (1.5-month lead-time) and JFM (0.5-month lead time). In OND, calculated Kendall's τ is more significant for all MMEs at the 1.5-month lead time compared to 0.5-month. Among all models, MME_63 (consists of FLOR_B01 and CMC2) has the highest goodness index ($\tau = 0.327$). Spearman

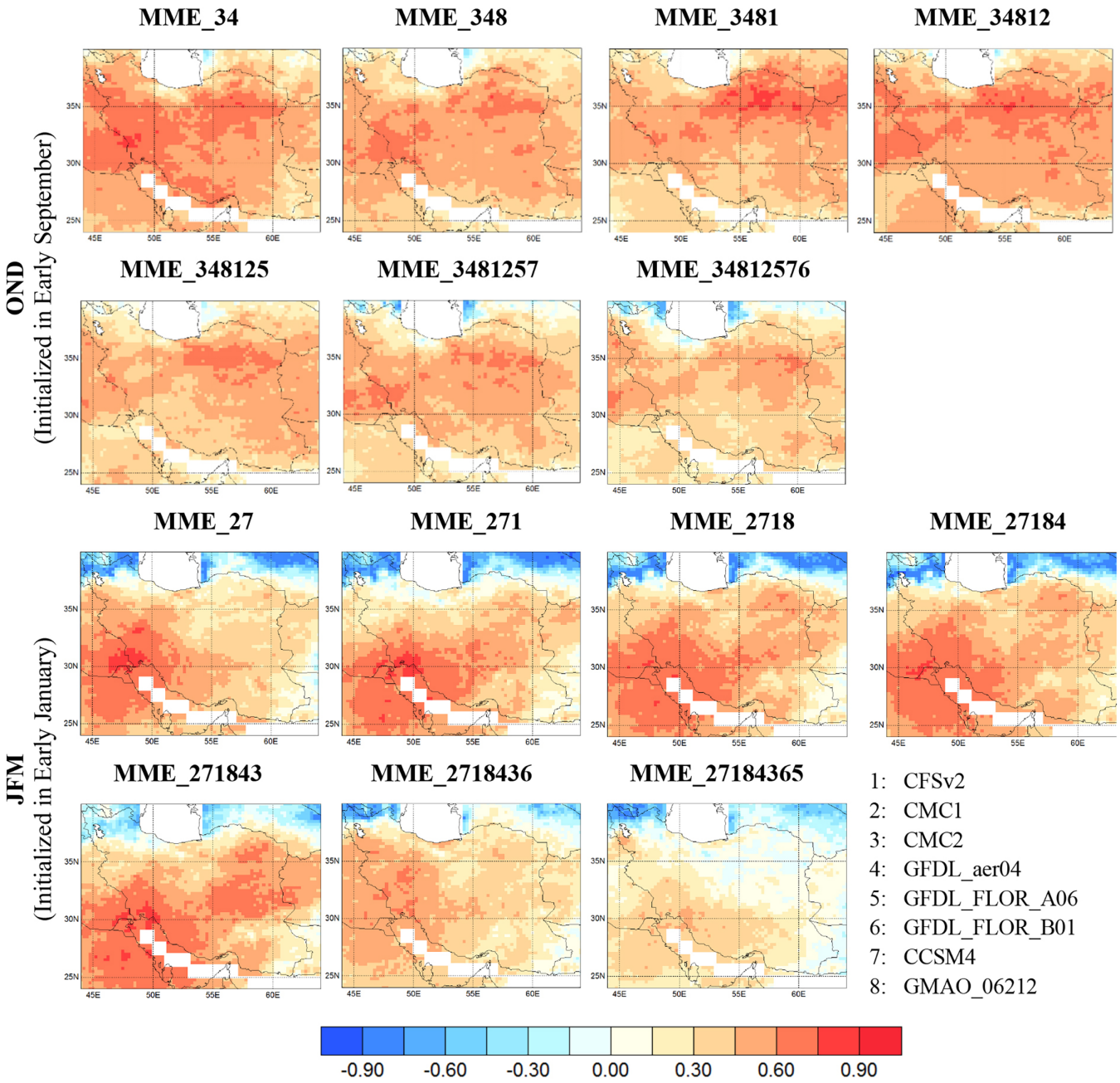

FIGURE 10 Seven different combinations of NMME individual models (MMEs) for precipitation forecasting in OND (1.5-month lead time) and JFM (0.5-month lead time) target seasons. Equal weights are assigned to individual models to develop each MME. The basis for combining individual models is their ranks according to calculated goodness index (Kendall's τ correlation) [Colour figure can be viewed at wileyonlinelibrary.com]

correlation of MME_63 ($.45 \leq \rho \leq .60$) in the northwestern parts is higher than other models.

b. JFM

According to Fig 129, all MMEs perform better in JFM in comparison with OND temperature forecasting. Spearman correlation exceeds 0.6 in the eastern parts. Some models also have $\rho \geq 0.6$ in the coastal line of the Caspian Sea. While the general pattern of Spearman correlation is more or less the same, some models show higher skill in the western part (e.g., MME_178). MME consisted of CFSv2 and CCSM4 (MME_17) has the highest skill compared to other models. The maximum RMSE and MAE of this model are 1 and 0.86 degree Celsius, respectively. Fig 130 summarizes the highest goodness index for OND and JFM temperature forecast at 0.5 and 1.5- lead times for individual and MMEs.

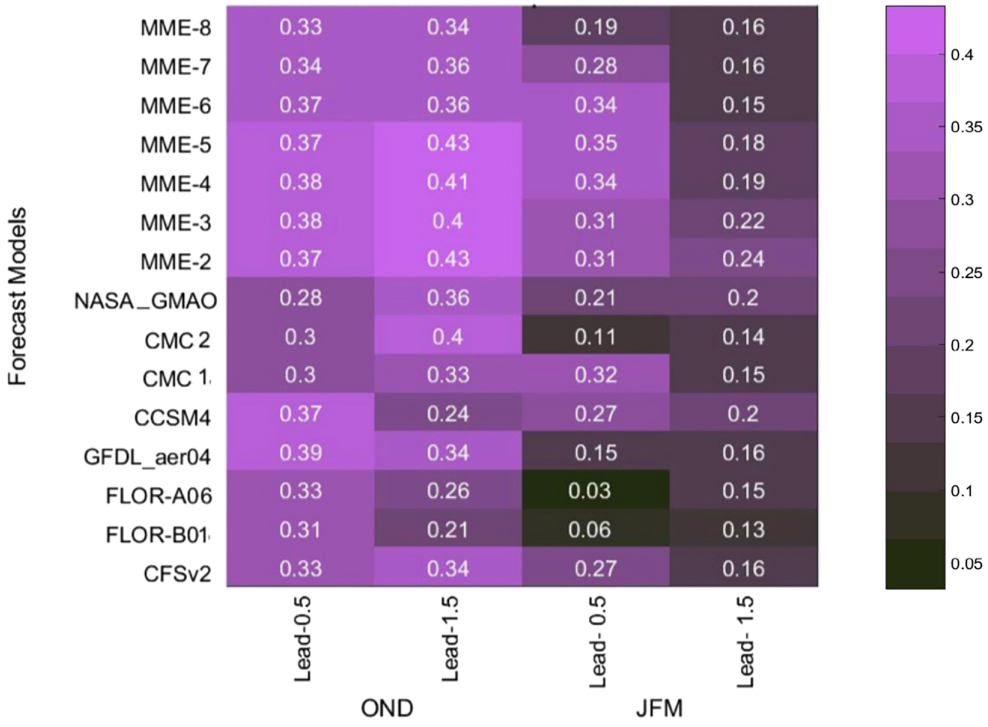

FIGURE 11 Average rank correlation (Kendal's τ) between the cross-validated forecasts (individual and MME) and PERSIANN-CDR for OND and JFM′ precipitation (1983–2013). The number after the term 'MME' is the number of individual models that take part in the Multi-model Ensemble. MME-8 is developed based on all individual models [Colour figure can be viewed at wileyonlinelibrary.com]

4.3 | Comparison of empirical and hybrid approaches

Given the recent advances in numerical modelling of the earth system and global observation network, one assumption is that a physics-based model provides a better forecast compared to that of empirical models. If physics-based models perform better, one reason might be the ability of GCMs to solve dynamical interactions between the land-atmosphere–ocean.

The analyses provided in previous sections reveal that empirical models have a moderate skill; however, hybrid models have higher correlation on average in both target seasons. The skill is season-dependent as regional climate predictability is influenced by large atmospheric-oceanic general circulations and regional impacts of teleconnections on spatiotemporal regional pattern of climate variability (e.g., ENSO, NAO) as demonstrated in a number of studies (e.g., see Hurrell, 1996; Cullen *et al*., 2002; Nazemosadat and Ghasemi, 2004; Ghasemi and Khalili, 2006; Nazemosadat *et al*., 2006; Dezfuli *et al*., 2010; Sabziparvar *et al*., 2011).

For comparison of empirical and hybrid approaches, the areas under the ROC curve (below and above normal) are provided for models with the highest calculated goodness index in Figure 14. In general, models do not have a similar performance in all parts of the study area, and the skill varies from one region to the other (except for regions with high climate predictability). For a specific target season, almost all forecasting models have skill regardless of the selected approach which means the climate is more predictable in specific parts of the region.

In a hybrid approach, developing MMEs has resulted in improved the skill of individual models. The improvement is sometimes significant. For precipitation, there is at least one hybrid model that outperforms empirical models and has a higher skill in terms of Spearman correlation. The MME consisted of five models (CMC1, CMC2, GMAO-06212, GFDL-aer04, and CFSv2) forecast correctly both below and above normal precipitation conditions in more than 70% of times. Areas across the Caspian Sea (north of Iran) in addition to the southeastern part of the study area have a negative rank correlation in JFM.

For temperature, the Spearman correlation (ROC area) of the empirical approach is negative (less than 0.5) in the southwest of Iran. MMEs are not able to outperform the best individual model (FLOR-B01) for forecasting OND temperature. However, all MMEs (consists of two to eight models) and some individual models have higher goodness index ($\tau_{\text{hybrid}} > \tau_{\text{empirical}} = 0.209$). Hybrid approach is significantly better than empirical in forecasting temperature for JFM target season. In MME with the highest goodness index (MME-5), more than 70% of grid points have $\rho \geq .6$. Approximately, 25% of all grid points within the study area have $\rho \geq .75$. For most parts of the study area in JFM target season, MME has a high discriminatory power which is provided by ROC skill maps ($ROC_{\text{above-normal}} \geq 0.8$. Also $ROC_{\text{below-normal}} \geq 0.7$).

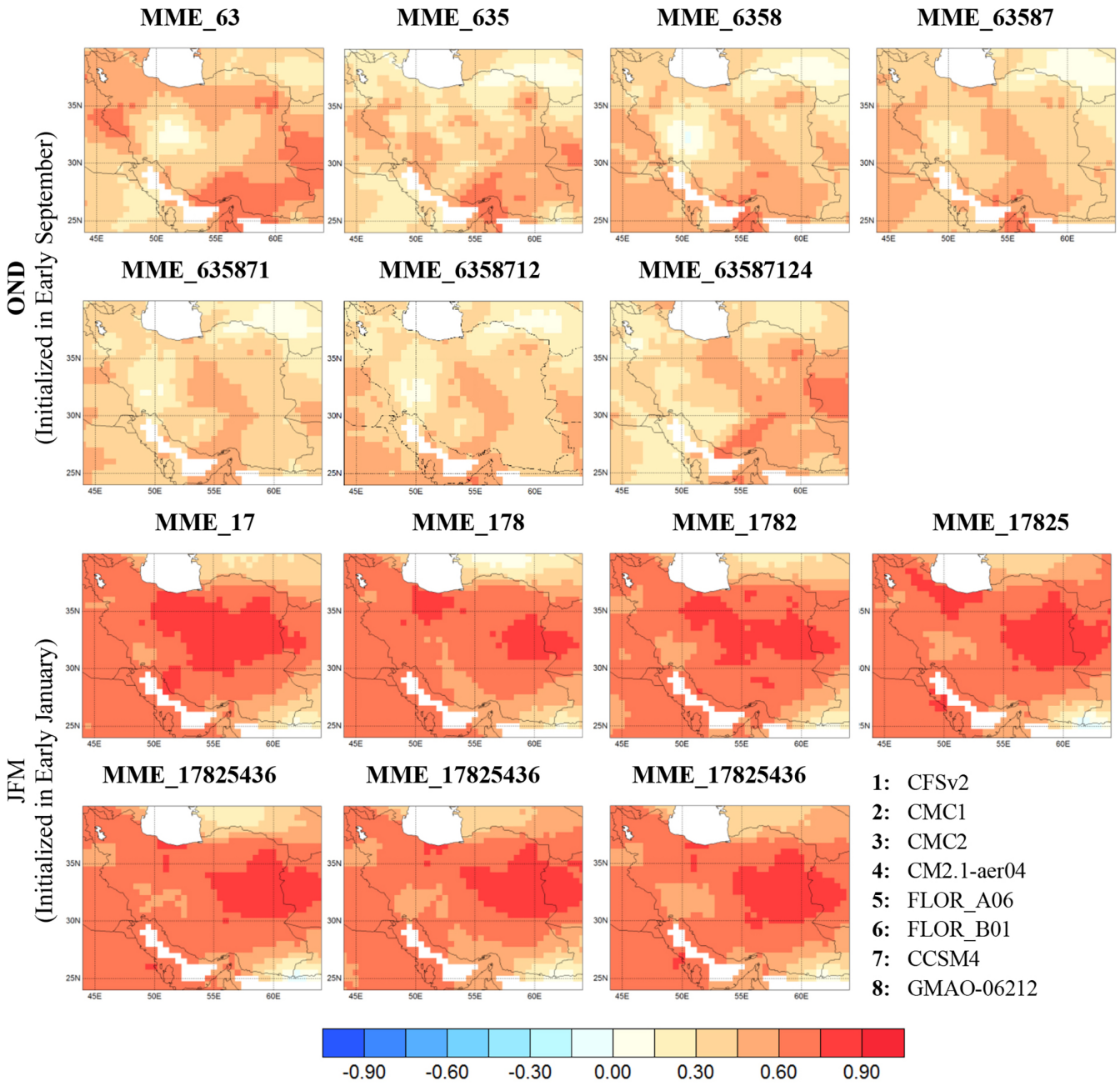

FIGURE 12 Seven different combinations of NMME individual models (MMEs) for forecasting temperature in OND (1.5-month lead time) and JFM (0.5-month lead time) target seasons. Equal weights are assigned to each model. Models which have the highest average Kendall's τ correlation after CCA is applied are added one by one to the MME. Figure 12 summarizes Kendal's τ between the cross-validated forecasts (individual and MME) and CRU at 0.5 and 1.5- month lead times for temperature (OND and JFM target seasons). According to Figure 12., temperature forecasts have higher skill in JFM in comparison to OND [Colour figure can be viewed at wileyonlinelibrary.com]

5 | SUMMARY AND DISSCUSSION

5.1 | Summary

In this article, two main approaches for seasonal forecasting of precipitation and temperature are compared in an area located in Southwest Asia including Iran. An empirical forecast approach is considered as a benchmark, and the inter-comparison analysis is conducted to see if MOS corrections of state-of-the-art coupled GCM outputs have higher skill outperforming that of the empirical approach. The same statistical method (CCA) is used in both the empirical and hybrid approaches. 1-year leave out cross-validation is applied to fit CCA within the period of 1983–2013 using predictor data. A variety of continuous performance scores are calculated to evaluate

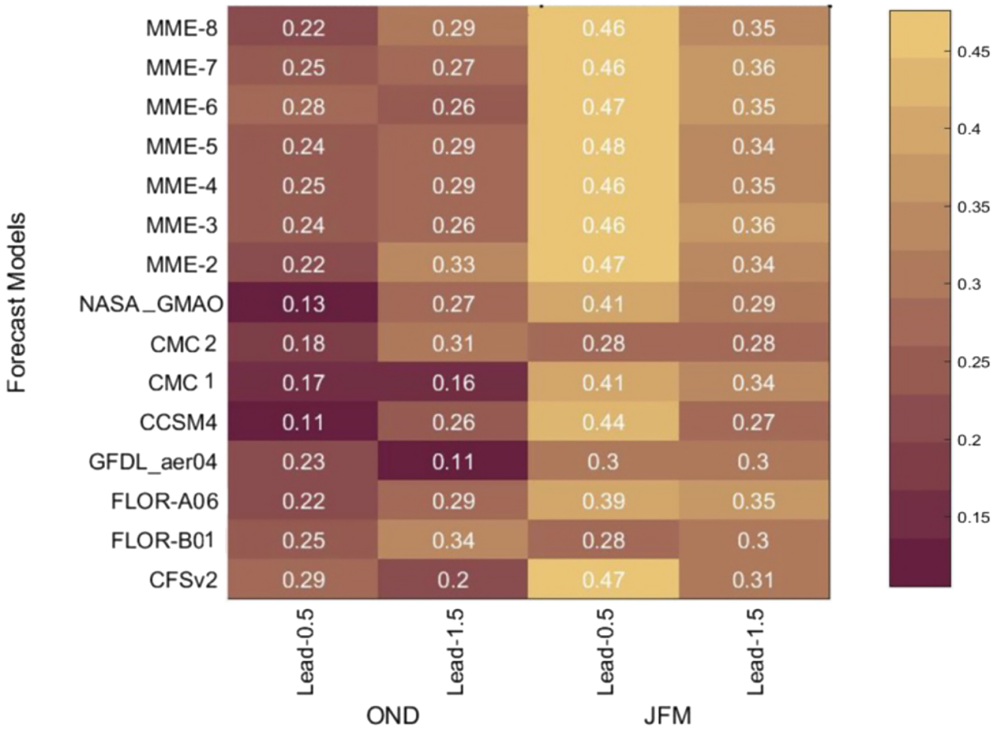

FIGURE 13 Average rank correlation (Kendal's τ) between the cross-validated forecasts (individual and MME) and CRU at 0.5 and 1.5- month lead times for temperature (OND and JFM target season over 1983–2013). The number after the term 'MME' is the number of individual models that take part in the Multi-model Ensemble. MME-8 is developed based on all individual models [Colour figure can be viewed at wileyonlinelibrary.com]

the forecasts including MAE, RMSE, Spearman, and Kendall's τ rank correlations. In addition, the ROC statistics for above and below normal conditions are calculated in accordance with the WMO's SVSLRF for all grid points. The main aims of the article have been to investigate the performance of the hybrid approach compared to the empirical one, and to test possible improvements by means of removing systematic biases and developing multi-model ensembles. Some of the main findings are:

- Empirical models have moderate skill in seasonal climate forecasts of precipitation and temperature (Figure 6).
- For both seasons and variables, the hybrid approach beats the raw GCM by means of improving model skill (e.g., see Figure 7 for CFSv2 comparison).
- Based on Figure 8 and Figure 9, hybrid models have a higher skill in forecasting OND precipitation compared to JFM. For temperature, hybrid model exhibits higher skill in JFM compared to OND.
- In forecasting precipitation using hybrid multi-model ensemble, higher skill is found for OND target season at the 1.5-month lead time (compared to 0.5- month), whereas in forecasting JFM, the skill is significantly dropped from 0.5-month lead-time to 1.5 month (Figure 11).
- Some individual NMME models have a high skill in the area under study for forecasting OND precipitation (Figure 8) and JFM temperature (Figure 9).
- Forecasting skill among different multi-model ensembles corresponds to the number of climate forecasting systems, their individual performance and systematic biases (Figure 10 and Figure 12).
- Developing MME can increase model skill compared to individual climate forecasting systems for precipitation (Figure 11) and temperature (Figure 13).
- Overall, hybrid models outperform empirical models (Figure 14).

5.2 | Discussion

While a large proportion of previous studies have focused on interannual variability during the boreal winter (November–April), this study considered the OND and JFM seasons separately, motivated by hydrologic user demands. Forecasting information for OND can help in early water resources planning in the beginning of water year. In cases of observed dry conditions in OND, skillful forecast of hydroclimate variables in JFM could help in drought risk management for the rest of water year.

In current study, several predictors are considered to build empirical seasonal forecasting of precipitation and temperature over Iran. Considering the link between Pacific Ocean and interannual variability in study area found in previous studies (e.g., Barlow *et al*., 2016), a physical reasoning was made for choosing predictor domains for empirical approach. Precipitation in the study region is mainly influenced by several teleconnections. Among several studies which have investigated the relation between Iran's precipitation

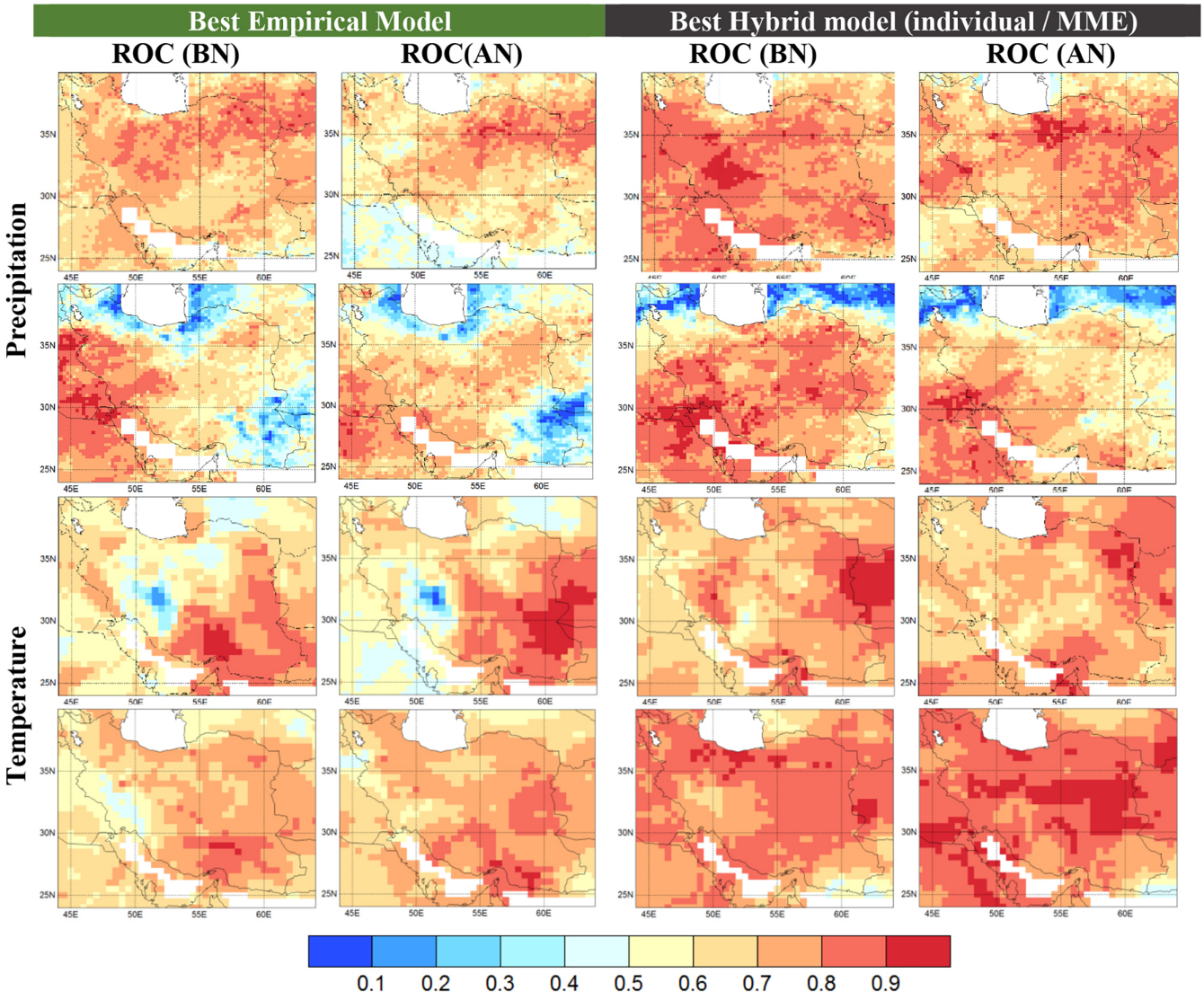

FIGURE 14 Comparison between empirical and hybrid approaches for OND (first and third rows) and JFM (second and fourth rows). Skill maps are based on the area under the ROC curve (below and above normal) for precipitation and temperature. Hybrid approach (individual/MME) having the highest goodness index (after cross-validation) are presented. The term 'Best' in this figure corresponds to the model which has the maximum Kendall's τ on average in target season of interest. For hybrid approach, the model which has the highest skill in all lead time (0.5 and 1.5- months) is provided. All measures are calculated over 1983–2013 [Colour figure can be viewed at wileyonlinelibrary.com]

and several teleconnections, most have been agreed that ENSO, IOD, NAO, and MJO are as the most influential teleconnection patterns (see e.g., Nazemosadat and Ghasemi, 2004; Pourasghar *et al*., 2015, 2019; Rana *et al*., 2017, 2019; Ahmadi *et al*., 2019). ENSO and IOD influenced OND precipitation while JFM precipitation variability is mainly associated with (NAO) based on teleconnection research studies over the region. Low NMME skill in forecasting winter precipitation over the study area has been assigned to the low predictability of NAO in seasonal time scale found by Ehsan *et al*. (2017). In this study, antecedent SST was found to provide the highest overall skill in terms of the goodness index for temperature (OND, JFM) and for precipitation (JFM). For forecasting near surface temperature, the time series of the leading CCA modes (Figures 4 and 5) suggest that a large part of the skill may be associated with the long-term warming trend in SST. The spatial pattern of the forecast skill between empirical and hybrid individual models are relatively similar for precipitation which shows higher skills with a southwest to northeast spatial pattern in OND, and in the western Zagros in JFM (see Figures 6 and 8). The similarity of results between the two approaches implies that: (a) there is SST-related predictability of Iran rainfall (largely associated with ENSO) that the

empirical approach is able to capture. These results agree with previous studies which found the high influence of ENSO in precipitation forecast using a statistical-based approach (e.g., Gerlitz *et al*., 2016), one similar to the empirical approach developed in this study. (b) Dynamical models are able to capture the corresponding link. Recent studies have shown statistically significant correlations in NMME associated with ENSO for all seasons and leads. For example, L'Heureux *et al*. (2017) found a high correlation (≥ 0.8) between NMME and Niño-3.4 index in 0 and 1-month lead times. Not surprisingly, two individual NMME models (CFSv2 and CMC2) which have the highest skill in forecasting Iran OND precipitation (see Figure 11), have also been found in a previous study to be the top two performing models for ENSO predictions based on deterministic skill assessment (Barnston *et al*., 2019). For temperature, empirical approach shows higher skill in the southeast part (Figure 6) in OND, similar to hybrid MME-2 (Figure 6 and 12). For forecasting JFM temperature, hybrid approach has resulted in very higher skill in many parts of the country compared to empirical (Figures 9 and 12). Since the spatial patterns of forecasting skill are found to be quite similar between empirical and hybrid approaches, at least for precipitation in both target seasons, it is argued that similarity of the result is because of inherent climate predictability in the region which changes from one season to the other. One central underlying assumption in the current research is that dynamic models are able to capture the main mechanisms responsible for regional variability of precipitation and temperature linked to the large scale atmospheric-ocean patterns. Therefore, the skill patterns and its similarity between empirical and hybrid approaches are rooted in the underlying assumption.

Except for southeastern part of study area, the skill of temperature forecast is limited with a similar pattern in empirical approach for both OND and JFM (Figure 6g–l). On the other hand, high skillful forecasts have been found by using a hybrid approach in forecasting JFM temperature (Figure 9g–l and 12). The physical mechanism between interannual variability of near surface air temperature in Southwest Asia associated to eastern Pacific El Niño, central Pacific El Niño and La Niña events has been discussed by Alizadeh-Choobari *et al*. (2018). Since the statistical method in both empirical and hybrid approaches are considered the same, perhaps, higher skill found in forecasting JFM temperature by hybrid approach is correspondence to the influence of other teleconnections such as North Atlantic Oscillation (NAO), the IDO, the Quasi-Biennial Oscillation (QBO), MJO and their combined effect.

The broad patterns of skill are similar in empirical and hybrid approach in OND, but the use of hybrid approach enables improved skill compared to the empirical one. The use of hybrid-MME enables improved skill which suggests improved skill associated with benefits of using different dynamical model structures within MME. However, as eight dynamic models are used in this study and with respect to structural model uncertainty, the spatial pattern of the skill between different dynamic models are not the same. The hybrid scheme can improve the dynamical forecast, but still poor for forecasting JFM precipitation and OND temperature. The improvement after MOS-correction is mainly associated with handling systematic biases in raw model outputs. Nevertheless, forecast skill in different seasons is still limited by large inherent predictability and representation of teleconnection patterns corresponding to variability of precipitation and temperature. Skillful prediction of the winter North Atlantic Oscillation systems in the midlatitudes is still under debate based on the recent literature and the skill level differs between state-of-the-art seasonal forecasting systems (Dunstone *et al*., 2016; Baker *et al*., 2018; Smith *et al*. 2019; Weisheimer *et al*. 2019).

Performance assessment between eight individual models and seven multi-model combinations after CCA is applied has shown that higher correlation skill is associated with increasing ensemble size. Both simple models averaging and model selection based on goodness index calculated by hybrid performance are considered as a baseline for developing MME in this study. Results shown in Figures 10 and 12 suggest that assigning equal weights to all individual seasonal forecasting systems is not necessarily end up with higher skill as including models with low skill could aggravate the MME overall skill. It is possible that the performance of an individual model is higher than the MME-8. On the other hand, specific combinations of individual models could lead to the highest performance among any of individual models and MME-8. In future studies, one can assess the forecasting skill of MMEs which are based on weighting the models according to goodness index (e.g., Bayesian Model Averaging). A meaningful baseline could also be a random selection of models. For real-time forecasting tailored for operation of water resources, hybrid models can provide useful information months in advance for areas with high climate predictability whereas to construct an empirical model, near-real-time observation of large-scale atmospheric-oceanic/general circulation patterns are required. In this study, cross-validation is used for MME model selection. It is suggested that the final optimized MME, be tested on an independent dataset (e.g., for 2014 onwards). Then, it will be recommended to end users for operational applications. Moreover, in an operational

framework, stakeholders require monthly updated hydroclimate forecast information for rolling 3-month periods starting from the first month of water year (e.g., OND, NDJ,...,SON) which could be studied in future research over southwest Asia.

ACKNOWLEDGEMENTS

The authors thank the NMME program partners and acknowledge the help of NCEP, IRI, and NCAR personnel in creating, updating and maintaining the NMME archive. NMME project is supported by NOAA, NSF, NASA, and DOE. The first author appreciates the University of Tehran's international affair department to provide support for conducting some parts of this study at the IRI and Princeton University. The authors are also grateful for the support provided by Ángel Muñoz, Micheal Bell, and Rémi Cousin for their help in data acquisition from the IRI data library.

ORCID

Husain Najafi https://orcid.org/0000-0002-0412-3572
Ali R. Massah Bavani https://orcid.org/0000-0002-5433-1791
Niko Wanders https://orcid.org/0000-0002-7102-5454
Eric F. Wood https://orcid.org/0000-0001-7037-9675

REFERENCES

Ahmadi, M., Kashki, A. and Dadashi Roudbari, A. (2018) Spatial modeling of seasonal precipitation–elevation in Iran based on Aphrodite database. *Modeling Earth System and Environment*, 4, 619–633. https://doi.org/10.1007/s40808-018-0444-y.

Ahmadi, M., Salimi, S., Hosseini, S.A., Poorantiyosh, H. & Bayat, A. (2019) Iran's precipitation analysis using synoptic modeling of major teleconnection forces (MTF). *Dynamics of Atmospheres and Oceans*, 85, 41–56.

Alijani, B. (2008) Effect of the Zagros Mountains on the spatial distribution of precipitation. *Journal of Mountain Science*, 5(3), 218–231.

Alizadeh-Choobari, O., Adibi, P. and Irannejad, P. (2018) Impact of the El Niño–Southern Oscillation on the climate of Iran using ERA-Interim data. *Climate Dynamics*, 51(7–8), 2897–2911.

Ashouri, H., Hsu, K., Sorooshian, S., Braithwaite, D.K., Knapp, K. R., Cecil, L.D., Nelson, B.R. and Prat, O.P. (2015) PERSIANN-CDR: daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bulletin of the American Meteorological Society*, 96, 69–83. https://doi.org/10.1175/BAMS-D-13-00068.1.

Baker, L.H., Shaffrey, L.C., Sutton, R.T., Weisheimer, A & Scaife, A. A.(2018) An intercomparison of skill and overconfidence/underconfidence of the wintertime North Atlantic Oscillation in multimodel seasonal forecasts. *Geophysical Research Letters* 45, no. 15 (2018), 45(15), 7817–7817. https://doi.org/10.1029/2018GL078838.

Barbero, R., Abatzoglou, J.T. and Hegewisch, K.C. (2017) Evaluation of statistical downscaling of northamerican multimodel ensemble forecasts over the western united states. *Weather Forecasting*, 32, 327–341. https://doi.org/10.1175/WAF-D-16-0117.1.

Barlow, M., Cullen, H. and Lyon, B. (2002) Drought in central and southwest Asia: La Niña, the warm pool, and Indian Ocean precipitation. *Journal of Climate*, 15(7), 697–700.

Barlow, M., Zaitchik, B., Paz, S., Black, E., Evans, J. and Hoell, A. (2016) A Review of Drought in the Middle East and Southwest Asia. *Journal of Climate*, 29, 8547–8574. https://doi.org/10.1175/JCLI-D-13-00692.1.

Barnett, T.P. and 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. https://doi.org/10.1175/1520-0493(1987)115<1825:OALOMA>2.0.CO;2.

Barnston, A.G. and Tippett, M.K. (2017) Do statistical pattern corrections improve seasonal climate predictions in the North American Multimodel Ensemble models? *Journal of Climate*, 30, 8335–8355. https://doi.org/10.1175/JCLI-D-17-0054.1.

Barnston, A.G., Tippett, M.K., L'Heureux, M., Li, S. and DeWitt, D. (2012) Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing? *Bulletin of the American Meteorological Society*, 93, 631–651. https://doi.org/10.1175/BAMS-D-11-00111.1.

Barnston, A.G., Tippett, M.K., Ranganathan, M., Michelle, L. and L'Heureux, M.L. (2019) Deterministic skill of ENSO predictions from the North American Multimodel Ensemble. *Climate Dynamics*, 53, 7215–7234. https://doi.org/10.1007/s00382-017-3603-3.

Barnston, A.G. & van den Dool, H.M. (1993) A degeneracy in cross-validated skill in regression-based forecasts. *Journal of Climate*, 6(5), 963–977.

Batté, L. and Déqué, M. (2011) Seasonal predictions of precipitation over Africa using coupled ocean-atmosphere general circulation models: skill of the ENSEMBLES project multimodel ensemble forecasts. *Tellus A*, 63, 283–299. https://doi.org/10.1111/j.1600-0870.2010.00493.x.

Bichet, A., Kushner, P.J. and Mudryk, L. (2016) Estimating the continental response to global warming using pattern-scaled sea surface temperatures and sea ice. *Journal of Climate*, 29, 9125–9139. https://doi.org/10.1175/JCLI-D-16-0032.1.

Block, P. (2011) Tailoring seasonal climate forecasts for hydropower operations. *Meteorology and Energy Security: Simulations, Projections, and Management*, 179, 1355–1368. https://doi.org/10.5194/hess-15-1355-2011.

Block, P. and Goddard, L. (2011) Statistical and dynamical climate predictions to guide water resources in Ethiopia. *Journal of Water Resources Planning and Management*, 138(3), 287–298. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000181.

Bolson, J., Martinez, C., Breuer, N., Srivastava, P. and Knox, P. (2013) Climate information use among southeast US water managers: beyond barriers and toward opportunities. *Regional Environmental Change*, 13, 141–151. https://doi.org/10.1007/s10113-013-0463-1.

Brönnimann, S. (2007) Impact of El Niño–southern oscillation on European climate. *Reviews of Geophysics*, 45(3). RG3003.https://doi.org/10.1029/2006RG000199.

Ciancarelli, B., Castro, C.L., Woodhouse, C., Dominguez, F., Chang, H.-I., Carrillo, C. and Griffin, D. (2014) Dominant

patterns of US warm season precipitation variability in a fine resolution observational record, with focus on the southwest. *International Journal of Climatology*, 34, 687–707. https://doi.org/10.1002/joc.3716.

Conway, D., Van Garderen, E.A., Deryng, D., Dorling, S., Krueger, T., Landman, W., Lankford, B., Lebek, K., Osborn, T., Ringler, C. and Thurlow, J. (2015) Climate and southern Africa's water–energy–food nexus. *Nature Climate Change*, 5 (9), 837–846. https://doi.org/10.1038/nclimate2735.

Crane, T.A., Roncoli, C. and Hoogenboom, G. (2011) Adaptation to climate change and climate variability: the importance of understanding agriculture as performance. *Wageningen Journal of Life Sciences*, 57, 179–185. https://doi.org/10.1016/j.njas.2010.11.002.

Crane, T.A., Roncoli, C., Paz, J., Breuer, N., Broad, K., Ingram, K.T. and Hoogenboom, G. (2010) Forecast skill and farmers' skills: Seasonal climate forecasts and agricultural risk management in the Southeastern United States. *Weather, Climate, and Society*, 2, 44–59. https://doi.org/10.1175/2009WCAS1006.1.

Crochemore, L., Ramos, M.H. and Pappenberger, F. (2016) Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts. *Hydrology and Earth System Sciences*, 20, 3601–3618. https://doi.org/10.5194/hess-20-3601-2016.

Cullen, H.M., Kaplan, A., Arkin, P.A. and Demenocal, P.B. (2002) Impact of the north Atlantic oscillation on the Middle Eastern climate and streamflow. *Climatic Change*, 55, 315–338.

DelSole, T., Yang, X. and Tippett, M.K. (2013) Is unequal weighting significantly better than equal weighting for multi-model forecasting? *Quarterly Journal of Royal Meteorological Society*, 139, 176–183. https://doi.org/10.1002/qj.1961.

Dezfuli, A.K., Karamouz, M. and Araghinejad, S. (2010) On the relationship of regional meteorological drought with SOI and NAO over southwest Iran. *Theoretical and Applied Climatology*, 100, 57–66.

Doblas-Reyes, F.J, Déqué, Michel and Piedelievre, J.-P. (2000) Multi-model spread and probabilistic seasonal forecasts in PROVOST. *Quarterly Journal of the Royal Meteorological Society*, 126(567), 2069–2087.

Donat, M.G., Peterson, T.C., Brunet, M., King, A.D., Almazroui, M., Kolli, R.K., Boucherf, D., Al-Mulla, A.Y., Nour, A.Y., Aly, A.A. and Nada, T.A.A. (2014) Changes in extreme temperature and precipitation in the Arab region: long-term trends and variability related to ENSO and NAO. *International Journal of Climatology*, 34(3), 581–592. https://doi.org/10.1002/joc.3707.

Dunstone, N., Smith, D., Scaife, A., Hermanson, L., Eade, R., Robinson, N., Andrews, M. and Knight, J. (2016) Skilful predictions of the winter North Atlantic Oscillation one year ahead. *Nature Geoscience*, 9, 809–814. https://doi.org/10.1038/ngeo2824.

ECMWF Working Group 1 Report at Model Uncertainty workshop. (2016) *What are the sources of model error and how can we improve the physical basis of model uncertainty representation? ECMWF, Reading, United Kingdom*. Available at: http://www.ecmwf.int/en/learning/workshops-and-seminars/wwrp-workshop-model-uncertainty.

Ehsan, M.A., Tippett, M.K., Almazroui, M., Ismail, M., Yousef, A., Kucharski, F., Omar, M., Hussein, M. and Alkhalaf, A.A. (2017) Skill and predictability in multimodel ensemble forecasts for Northern Hemisphere regions with dominant winter precipitation. *Climate Dynamics*, 48, 3309–3324. https://doi.org/10.1007/s00382-016-3267-4.

Emerton, R., Cloke, H.L., Stephens, E.M., Zsoter, E., Woolnough, S. J. and Pappenberger, F. (2017) Complex picture for likelihood of ENSO-driven flood hazard. *Nature Communications*, 8, 14796. https://doi.org/10.1038/ncomms14796.

Fraedrich, K. and Müller, K. (1992) Climate anomalies in Europe associated with ENSO extremes. *International Journal of Climatology*, 12(1), 25–31.

Frías, M.D., Herrera, S., Cofiño, A.S. and Gutiérrez, J.M. (2010) Assessing the skill of precipitation and temperature seasonal forecasts in Spain: Windows of opportunity related to ENSO events. *Journal of Climate*, 23, 209–220. https://doi.org/10.1175/2009JCLI2824.1.

Gent, P.R., Yeager, S.G., Neale, R.B., Levis, S. and Bailey, D.A. (2010) Improvements in a half degree atmosphere/land version of the CCSM. *Climate Dynamics*, 34(6), 819–833. https://doi.org/10.1007/s00382-009-0614-8.

Gerlitz, L., Vorogushyn, S., Apel, H., Gafurov, A., Unger-Shayesteh, K. and Merz, B. (2016) A statistically based seasonal precipitation forecast model with automatic predictor selection and its application to central and south Asia. *Hydrology and Earth System Sciences*, 20, 4605–4623. https://doi.org/10.5194/hess-20-4605-2016.

Ghasemi, A. and Khalili, D. (2006) The influence of the Arctic Oscillation on winter temperatures in Iran. *Theoretical and Applied Climatology*, 85(3–4), 149–164. https://doi.org/10.1007/s00704-005-0186-4.

Gilpin, A.R. (1993) Table for conversion of Kendall's Tau to Spearman's Rho within the context of measures of magnitude of effect for meta-analysis. *Educational and Psychological Measurement*, 53(1), 87–92.

Gobena, A.K. and Gan, T.Y. (2010) Incorporation of seasonal climate forecasts in the ensemble streamflow prediction system. *Journal of Hydrology*, 385(1–4), 336–352. https://doi.org/10.1016/j.jhydrol.2010.03.002.

Golembesky, K., Sankarasubramanian, A. and Devineni, N. (2009) Improved drought management of falls lake reservoir: role of multimodel streamflow forecasts in setting up restrictions. *Journal of Water Resources Planning and Management*, 3(188), 188–197. https://doi.org/10.1061/(ASCE)0733-9496.

Hansen, J.W., Mason, S., Sun, L. and Tall, A. (2011) Review of seasonal climate forecasting for agriculture in sub-Saharan Africa. *Experimental Agriculture*, 47(2), 205–240. https://doi.org/10.1017/S0014479710000876.

Harris, I., Jones, P.D., Osborn, T.J. and Lister, D.H. (2014) Updated high-resolution grids of monthly climatic observations-the CRU TS3.10 Dataset. *International Journal of Climatology*, 34, 623–642. https://doi.org/10.1002/joc.3711.

Hellmuth, M.E., Mason, S.J., Vaughan, C., Van Aalst, M.K. and Choularton, R. (2011) *A Better Climate for Disaster Risk Management*. New York, NY: International Research Institute for Climate and Society (IRI), Columbia University.

Hoell, A., Barlow, M., Cannon, F. and Xu, T. (2017a) Oceanic Origins of Historical Southwest Asia Precipitation During the Boreal Cold Season. *Journal of Climate*, 30, 2885–2903.

Hoell, A., Barlow, M., Xu, T. and Zhang, T. (2018) Cold Season Southwest Asia Precipitation Sensitivity to El Niño–Southern Oscillation Events. *Journal of Climate*, 31(11), 4463–4482. https://doi.org/10.1175/JCLI-D-17-0456.1.

Hoell, A., Funk, C., Barlow, M., and Cannon, F. (2017b). *A Physical Model for Extreme Drought Over Southwest Asia. Climate Extremes*. John Wiley & Sons, Inc., pp. 283–298.

Huang, J., van den Dool, J.H. and Georgarakos, K.P. (1996) *Analysis of model-calculated soil moisture over the United States (1931–1993) and applications to long-range temperature forecasts*. Washington, DC: Climate Prediction Center, National Centers for Environmental Prediction. https://doi.org/10.1175/1520-0442(1996)009<1350:AOMCSM>2.0.CO;2.

Hurrell, J.W. (1996) Influence of variation in extratropical wintertime teleconnections on Northern Hemisphere temperature. *Geophysical Research Letters*, 23, 665–668.

Irannejad, P., Ahmadi-Givi, F. and Nikouei, N. (2016) A study of winter temperature anomalies in Iran by using the NCEP/NCAR reanalysis dataset. *Iranian Geophysical Society*, 10(4), 12–27.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R. and Joseph, D. (1996) The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society*, 77, 437–472. https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Katiraie-Boroujerdy, P.S., Akbari Asanjan, A., Hsu, K.L. and Sorooshian, S. (2017) Intercomparison of PERSIANN-CDR and -TRMM-3B42V7 precipitation estimates at monthly and daily time scales. *Atmospheric Research*, 193, 36–49. https://doi.org/10.1016/j.atmosres.2017.04.005.

Katiraie-Boroujerdy, P.S., Ashouri, H., Hsu, K. and Sorooshian, S. (2016) Trends of precipitation extreme indices over a subtropical semi-arid area using PERSIANN-CDR. *Theoretical and Applied Climatology*, 139, 249–460. https://doi.org/10.1007/s00704-016-1884-9.

Khajehei, S., Ahmadalipour, A. and Moradkhani, H. (2017) An effective post-processing of the North American multi-model ensemble (NMME) precipitation forecasts over the continental US. *Climate Dynamics*, 51, 1–16. https://doi.org/10.1007/s00382-017-3934-0.

Kiani, M., Lashkari, H. and Ghaemi, H. (2019) The effect of Zagros Mountains on rainfall changes of Sudanese low pressure system in western Iran. *Modeling Earth System and Environment*, 5, 1769–1779. https://doi.org/10.1007/s40808-019-00631-w.

Kim, H.M., Webster, P.J. and Curry, J.A. (2012) Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere Winter. *Climate Dynamics*, 39(12), 2957–2973. https://doi.org/10.1007/s00382-012-1364-6.

Kirtman, B.P., Min, D., Infanti, J.M., Kinter, J.L., Paolino, D.A., Zhang, Q., van den Dool, H., Saha, S., Mendez, M.P., Becker, E., Peng, P., Tripp, P., Huang, J., DeWitt, D.G., Tippett, M.K., Barnston, A.G., Li, S., Rosati, A., Schubert, S.D., Rienecker, M., Suarez, M., Li, Z.E., Marshak, J., Lim, Y.-K., Tribbia, J., Pegion, K., Merryfield, W.J., Denis, B., et al. (2014) The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intra-seasonal prediction. *Bulletin of American Meteorological Society*, 95, 585–601.

L'Heureux, M.L., Tippett, M.K., Kumar, A, Butler,, A.H., Ciasto,, L. M., Ding, Q, Harnos, J. & Johnson, N.C. (2017) Strong relations between ENSO and the Arctic Oscillation in the North American multimodel ensemble. *Geophysical Research Letters*, 44(11), 11-654–11-662. https://doi.org/10.1002/2017GL074854.

Li, H., Luo, L., Wood, E.F. and Schaake, J. (2009) The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting. *Journal of Geophysical Research*, 114, D04114. https://doi.org/10.1029/2008JD010969.

Li, W., Sankarasubramanian, A., Ranjithan, R.S. and Brill, E.D. (2014) Improved regional water management utilizing climate forecasts: An interbasin transfer model with a risk management framework. *Water Resources Research*, 50, 6810–6827. https://doi.org/10.1002/2013WR015248.

Lu, M., Lall, U., Robertson, A.W. and Cook, E. (2017) Optimizing multiple reliable forward contracts for reservoir allocation using multitime scale streamflow forecasts. *Water Resources Research*, 53(3), 2035–2050. https://doi.org/10.1002/2016WR019552.

Ma, F., Ye, A., Deng, X., Zhou, Z., Liu, X., Duan, Q., Xu, J., Miao, C., di, Z. and Gong, W. (2015) Evaluating the skill of NMME seasonal precipitation ensemble predictions for 17 hydroclimatic regions in continental China. *International Journal of Climatology*, 36, 132–144. https://doi.org/10.1002/joc.4333.

Madadgar, S., AghaKouchak, A., Shukla, S., Wood, A.W., Cheng, L., Hsu, K.L. and Svoboda, M. (2016) A hybrid statistical-dynamical framework for meteorological drought prediction: Application to the southwestern United States. *Water Resources Research*, 52(7), 5095–5110. https://doi.org/10.1002/2015WR018547.

Manzanas, R., Gutiérrez, J.M., Fernández, J., van Meijgaard, E., Calmanti, S., Magariño, M.E., Cofiño, A.S. and Herrera, S. (2017) Dynamical and statistical downscaling of seasonal temperature forecasts in Europe: Added value for user applications. *Climate Services*, 9, 44–56. https://doi.org/10.1016/j.cliser.2017.06.004.

Mason, S.J. and Goddard, L. (2001) Probabilistic precipitation anomalies associated with ENSO. *Bulletin of American Meteorological Society*, 82(4), 619–638. https://doi.org/10.1175/1520-0477(2001)082<0619:PPAAWE>2.3.CO;2.

Mass, C.F., Baars, J., Wedam, G., Grimit, E. and Steed, R. (2008) Removal of systematic model bias on a model grid. *Weather and Forecasting*, 23(3), 438–459. https://doi.org/10.1175/2007WAF2006117.1.

Merryfield, W.J., Lee, W., Boer, G.J., Kharin, V.V., Scinocca, J.F., Flato, G.M., Ajayamohan, R.S., Fyfe, J.C., Tang, Y. and Polavarapu, S. (2013) The Canadian seasonal to interannual prediction system. Part I: Models and initialization. *Monthly Weather Review*, 141, 2910–2945. https://doi.org/10.1175/MWR-D-12-00216.1.

Miri, M., Azizi, G., Mohamadi, H. and Pourhashemi, M. (2017) Evaluation statistically of temperature and precipitation datasets with observed data in Iran. *Iranian. Journal of Watershed Management Science*, 10(35), 39–50 Available at http://jwmsei.ir/browse.php?a_id=587&sid=1&slc_lang=en.

Mo, K.C. and Lyon, L. (2015) Global Meteorological drought prediction using the north American multi-model ensemble. *Journal of Hydrometeorology*, 16, 1409–1424. https://doi.org/10.1175/JHM-D-14-0192.1.

Molavi-Arabshahi, M., Arpe, K. and Leroy, S.A.G. (2016) Precipitation and temperature of the southwest Caspian Sea region

during the last 55 years: Their trends and teleconnections with large-scale atmospheric phenomena. *International Journal of Climatology*, 36(5), 2156–2172. https://doi.org/10.1002/joc.4483.

Mason, S. J. & Tippet, M. K.(2017) Climate Predictability Tool version 15.6.1. Columbia University Academic Commons. https://doi.org/10.7916/D8SF37S0.

Moron, V. and Plaut, G. (2003) The impact of El Niño–southern oscillation upon weather regimes over Europe and the North Atlantic during boreal winter. *International Journal of Climatology*, 23, 363–379.

Mwangi, E., Wetterhall, F., Dutra, E., Di Giuseppe, F. and Pappenberger, F. (2014) Forecasting droughts in East Africa. *Hydrology and Earth System Sciences*, 18(2), 611–620. https://doi.org/10.5194/hess-18-611-2014.

National Research Council. (2010) *Assessment of intraseasonal to interannual climate prediction and predictability*. Washington, DC: National Academies Press.

Nazemosadat, M.J. and Cordery, I. (2000) On the relationships between ENSO and autumn rainfall in Iran. *International Journal of Climatology*, 20(1), 47–61. https://doi.org/10.1002/(SICI)1097-0088(200001)20:1<47::AID-JOC461>3.0.CO;2-P.

Nazemosadat, M.J. and Ghaedamini, H. (2010) On the relationships between the Madden–Julian Oscillation and precipitation variability in Southern Iran and the Arabian Peninsula: atmospheric circulation analysis. *Journal of Climate*, 23, 887–904. https://doi.org/10.1175/2009JCLI2141.1.

Nazemosadat, M.J. and Ghasemi, A.R. (2004) Quantifying the ENSO-related shifts in the intensity and probability of drought and wet periods in Iran. *Journal of Climate*, 17, 4005–4018. https://doi.org/10.1175/1520-0442(2004)017<4005:QTESIT>2.0.CO;2.

Nazemosadat, M.J., Samani, N., Barry, D.A. and Molaii, N. (2006) ENSO forcing on climate change in Iran: precipitation analyses. *Iranian Journal of Science & Technology, Transaction B, Engineering*, 30(B4), 47–61.

Nazemosadat, M.J. and Shahgholian, K. (2017) Heavy precipitation in the southwest of Iran: association with the Madden–Julian Oscillation and synoptic scale analysis. *Climate Dynamics*, 49, 3091–3109. https://doi.org/10.1007/s00382-016-3496-6.

Oludhe, C., Arumugam, S., Sinha, T., Devineni, N. and Lall, U. (2013) The role of multimodel climate forecasts in improving water and energy management over the Tana River Basin, Kenya. *Journal of Applied and Meteorolological Climatology*, 52, 2460–2475. https://doi.org/10.1175/JAMC-D-12-0300.1.

Pegion, K. (2015) *The North American Mul2-Model Ensemble: seasonal to subseasonal predictions, workshop on sub-seasonal predictability, ECMWF, Reading, United Kingdom* Available at: http://www.ecmwf.int/en/learning/workshops-and-seminars/past-workshops/workshop-sub-seasonal-predictability.

Pourasghar, F., Tozuka, T., Ghaemi, H., Oettli, P., Jahanbakhsh, S. and Yamagata, T. (2015) Influences of the MJO on intra-seasonal rainfall variability over southern Iran. *Atmospheric Science Letters*, 16, 110–118. https://doi.org/10.1002/asl2.531.

Pourasghar, F., Tozuka, T., Jahanbakhsh, S., Sarraf, B.S., Ghaemi, H. and Yamagata, T. (2012) The interannual precipitation variability in the southern part of Iran as linked to large-scale climate modes. *Climate Dynamics*, 39, 2329–2341. https://doi.org/10.1007/s00382-012-1357-5.

Pourasghar, F, Oliver, E. C.J. & Holbrook, N. J.(2019) Modulation of wet-season rainfall over Iran by the Madden–Julian Oscillation, Indian Ocean Dipole and El Niño–Southern Oscillation. *International Journal of Climatology*, 10,4029–4040.

Rana, S., McGregor, J. and Renwick, J. (2017) Wintertime precipitation climatology and ENSO sensitivity over central southwest Asia. *International Journal of Climatology*, 37, 1494–1509.

Rana, S., McGregor, J. and Renwick, J. (2019) Dominant modes of winter precipitation variability over Central Southwest Asia and inter-decadal change in the ENSO teleconnection. *Climate Dynamics*, 53, 5689–5707. https://doi.org/10.1007/s00382-019-04889-9.

Raziei, T., Bordi, I. and Pereira, L.S. (2013b) Regional drought modes in Iran using the SPI: the effect of time scale and spatial resolution. *Water Resources Management*, 27, 1661–1674. https://doi.org/10.1007/s11269-012-0120-3.

Raziei, T., Bordi, I., Santos, J.A. and Mofidi, A. (2013a) Atmospheric circulation types and winter daily precipitation in Iran. *International Journal of Climatology*, 33(9), 2232–2246.

Robertson, A.W., Baethgen, W., Block, P., Lall, U., Sankarasubramanian, A., de Assis de Souza Filho, F. and Verbist, K.M.J. (2014) Climate risk management for water in semi–arid regions. *Earth Perspectives*, 1, 12. https://doi.org/10.1186/2194-6434-1-12.

Robertson, A.W., Lall, U, Zebiak, S.E. & Goddard, L. (2004) Improved combination of multiple atmospheric GCM ensembles for seasonal prediction. *Monthly Weather Review*, 132(12), 2732–2744.

Robertson, D.E. and Wang, Q.J. (2013) Seasonal forecasts of unregulated inflows into the Murray River, Australia. *Water Resources Management*, 27(8), 2747–2769. https://doi.org/10.1007/s11269-013-0313-4.

Rodríguez-Fonseca, B., Suárez-Moreno, R., Ayarzagüena, B., López-Parages, J., Gómara, I., Villamayor, J., Mohino, E., Losada, T. and Castaño-Tierno, A. (2016) A review of ENSO influence on the North Atlantic. A non-stationary signal. *Atmosphere*, 7(7), 87. https://doi.org/10.3390/atmos7070087.

Sabziparvar, A.A., Mirmasoudi, S.H., Tabari, H., Nazemosadat, M.J. and Maryanaji, Z. (2011) ENSO teleconnection impacts on reference evapotranspiration variability in some warm climates of Iran. *Intertnational Journal of Climatology*, 31, 1710–1723. https://doi.org/10.1002/joc.2187.

Sabziparvar, A.A., Movahedi,, S & Asakereh, H.(2015) Geographical factors affecting variability of precipitation regime in Iran. *Theoretical and Applied Climatology*, 120,367–376. https://doi.org/10.1007/s00704-014-1174-3.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M.P., van den Dool, H., Zhang, Q., Wang, W., Chen, M. and Becker, E. (2014) The NCEP climate forecast system version 2. *Journal of Climate*, 27, 2185–2208.

Sahu, N., Robertson, A.W., Boer, R., Behera, S., DeWitt, D.G., Takara, K., Kumar, M. and Singh, R.B. (2017) Probabilistic seasonal streamflow forecasts of the Citarum River, Indonesia, based on general circulation models. *Stochastic Environmental Research and Risk Assessment*, 31, 1747–1758. https://doi.org/10.1007/s00477-016-1297-4.

Sankarasubramanian, A., Lall, U., Devineni, N. and Espinueva, S. (2009b) The role of monthly updated climate forecasts in improving intraseasonal water allocation. *Journal of Applied Meteorology and Climatology*, 48, 1464–1482. https://doi.org/10.1175/2009JAMC2122.1.

Sankarasubramanian, A., Lall, U., Souza Filho, F.A. and Sharma, A. (2009a) Improved water allocation utilizing probabilistic climate forecasts: Short-term water contracts in a risk management framework. *Water Resources Research*, 45. W11409.https://doi.org/10.1029/2009WR007821.

Sheffield, J., Wood, E.F., Chaney, N., Guan, K., Sadri, S., Yuan, X., Olang, L., Amani, A., Ali, A., Demuth, S. and Ogallo, L. (2014) A drought monitoring and forecasting system for sub-Sahara African water resources and food security. *Bulletin of American Meteorological Society*, 95(6), 861–882. https://doi.org/10.1175/BAMS-D-12-00124.1.

Shukla, S., Roberts, J., Hoell, A., Funk, C.C., Robertson, F. and Kirtman, B. (2016) Assessing North American multimodel ensemble (NMME) seasonal forecast skill to assist in the early warning of anomalous hydrometeorological events over East Africa. *Climate Dynamics*, 53, 7411–7427. https://doi.org/10.1007/s00382-016-3296-z.

Slater, L.J., Villarini, G. and Bradley, A.A. (2017) Weighting of NMME temperature and precipitation forecasts across Europe. *Journal of Hydrology*, 552, 646–659. https://doi.org/10.1016/j.jhydrol.2017.07.029.

Smith, D.M., Eade, R., Scaife, A.A., Caron, L.P., Danabasoglu, G., DelSole, T.M., Delworth, T., Doblas-Reyes, F.J., Dunstone, N.J., Hermanson, L. & Kharin, V., Kimoto, M., Merryfield, W.J., Mochizuki, T., Müller, W.A., Pohlmann, H., Yeager S. and Yang X. (2019) Robust skill of decadal climate predictions. *Npj Climate and Atmospheric Science*, 2(1), 1–10. https://doi.org/10.1038/s41612-019-0071-y.

Smith, T.M., Reynolds, R.W., Peterson, T.C. and Lawrimore, J. (2008) Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880-2006). *Journal of Climate*, 21, 2283–2296.

Soltani, M., Laux, P., Kunstmann, H., Stan, K., Sohrabi, M.M., Molanejad, M., Sabziparvar, A.A., SaadatAbadi, A.R., Ranjbar, F., Rousta, I. and Zawar-Reza, P. (2016) Assessment of climate variations in temperature and precipitation extreme events over Iran. *Theoretical and Applied Climatology*, 126(3–4), 775–795. https://doi.org/10.1007/s00704-015-1609-5.

Sun, X., Renard, B., Thyer, M., Westra, S. and Lang, M. (2015) A global analysis of the asymmetric effect of ENSO on extreme precipitation. *Journal of Hydrology*, 530, 51–65. https://doi.org/10.1016/j.jhydrol.2015.09.016.

Tabari, H., Somee, B.S. and Zadeh, M.R. (2011) Testing for long-term trends in climatic variables in Iran. *Atmospheric Research*, 100, 132–140. https://doi.org/10.1016/j.atmosres.2011.01.005.

Thober, S., Kumar, R., Sheffield, J., Mai, J., Schäfer, D. and Samaniego, L. (2015) Seasonal soil moisture drought prediction over Europe using the North American Multi-Model Ensemble (NMME). *Journal of Hydrometeorology*, 16, 2329–2344. https://doi.org/10.1175/JHM-D-15-0053.1.

Tippett, M.K., Barlow, M. and Lyon, B. (2003) Statistical correction of central southwest Asia winter precipitation simulations. *International Journal of Climatology*, 23(12), 1421–1433.

Tippett, M.K. and DelSole, T. (2013) Constructed analogs and linear regression. *Monthly Weather Review*, 141, 2519–2525. https://doi.org/10.1175/MWR-D-12-00223.1.

Troccoli, A., Harrison, M., Anderson, D.L.T. and Mason, S.J. (Eds.). (2008) *Seasonal Climate: Forecasting and Managing Risk*, Vol. 82. Netherlands: Springer Science & Business Media.

van den Dool, H., Huang, J. and Fan, Y. (2003) Performance and analysis of the constructed analogue method applied to U.S. soil moisture over 1981–2001. *Journal of Geophysical Research*, 108(D16), 8617. https://doi.org/10.1029/2002JD003114.

Vernieres, G., Rienecker, M., Kovach, R., and Keppenne, C.H.L. (2012) *The GEOSiODAS: Description and Evaluation. Technical Report Series on Global Modeling and Data Assimilation*, TM-2012-104606 30, pp. 1–61

Vigaud, N., Robertson, A.W. and Tippett, M.K. (2017) Multimodel ensembling of subseasonal precipitation forecasts over North America. *Monthly Weather Review*, 145, 3913–3928. https://doi.org/10.1175/MWR-D-17-0092.1.

Vogel, H.C. and O'Brien, K.L. (Eds.). (2003) *Coping with climate variability: The use of seasonal climate forecasts in Southern Africa*. Aldershot: Ashgate Publishing.

Walker, G. T. and Bliss, E. W. (1932) World Weather V. Memoirs of the Royal Meteorological Society. 4, 53–84.

Wanders, N., Thober, S., Kumar, R., Pan, M., Sheffield, J., Samaniego, L. and Wood, E.F. (2019) Development and evaluation of a Pan-European multimodel seasonal hydrological forecasting system. *Journal of Hydrometeorology*, 1(20), 99–115. https://doi.org/10.1175/JHM-D-18-0040.1.

Wanders, N. and Wood, E.F. (2016) Improved sub-seasonal meteorological forecast skill using weighted multi-model ensemble simulations. *Environmental Research Letters*, 11(9), 094007. https://doi.org/10.1088/1748-9326/11/9/094007.

Wang, H. and Liu, J. (2013) Reservoir operation incorporating hedging rules and operational inflow forecasts. *Water Resources Management*, 27, 1427–1438. https://doi.org/10.1007/s11269-012-0246-3.

Weigel, A.P., Liniger, M.A. and Appenzeller, C. (2008) Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? *Quarterly Journal of Royal Meteorological Society*, 134(630), 241–260. https://doi.org/10.1002/qj.210.

Weisheimer, A. and Palmer, T. (2014) On the reliability of seasonal climate forecasts. *Journal of the Royal Society Interface*, 11, 9620131162.

Weisheimer, A, Decremer, D, MacLeod, D, Reilly, C.O., Stockdale, T. N., Johnson, S & Palmer, T.N. (2019) How confident are predictability estimates of the winter North Atlantic Oscillation?. *Quarterly Journal of the Royal Meteorological Society* 145(Suppl.1), 140–159. https://doi.org/10.1002/qj.3446.

Wood, E.F., Schubert, S.D., Wood, A.W., Peters-Lidard, P.D., Mo, K.C., Mariotti, A. and Pulwarty, R.S. (2015) Prospects for advancing drought understanding, monitoring, and prediction. *Journal of Hydrometeorology*, 16, 1636–1657. https://doi.org/10.1175/JHM-D-14-0164.1.

Xue, Y., Smith, T.M. and Reynolds, R.W. (2003) Interdecadal changes of 30-yr SST normals during 1871-2000. *Journal of Climate*, 16, 1601–1612.

Yu, P.S., Yang, T.C. and Kuo, C.M. (2014) A stochastic approach for seasonal water-shortage probability forecasting based on seasonal weather outlook. *Water Resources Management*, 28, 3905–3920. https://doi.org/10.1007/s11269-014-0717-9.

Yuan, X. and Wood, E.F. (2013) Multimodel seasonal forecasting of global drought onset. *Geophysical Research Letters*, 40(18), 4900–4905. https://doi.org/10.1002/grl.50949.

Zhang, S., Harrison, M.J., Rosati, A. and Wittenberg, A. (2007) System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Monthly Weather Review*, 135, 3541–3564. https://doi.org/10.1175/MWR3466.1.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Najafi H, Robertson AW, Massah Bavani AR, Irannejad P, Wanders N, Wood EF. Improved multi-model ensemble forecasts of Iran's precipitation and temperature using a hybrid dynamical-statistical approach during fall and winter seasons. *Int J Climatol*. 2021;41:5698–5725. https://doi.org/10.1002/joc.7148