Potential predictability of Arabian peninsula summer surface air temperature in the North American multimodel ensemble

Muhammad Azhar Ehsan[1,2] · Fred Kucharski[1,2] · Mansour Almazroui[2] · Muhammad Ismail[2] · Michael K. Tippett[3]

Abstract

The present study evaluates the potential predictability, and prediction skill of surface air temperature (SAT) over the Arabian peninsula (AP) referred to as AP-SAT hereafter, during boreal summer (June–July–August: JJA) from 1982 through 2017. The study was made by considering the single model, and a multimodel ensemble (MME) approach. The seasonal prediction data for JJA SAT initialized at May (Lead-1), and April (Lead-2) observed initial conditions, from six coupled atmosphere–ocean global circulation models included in the North American Multimodel Ensemble, is utilized. The potential predictability (PP) is estimated through the estimation of the signal-to-noise ratio (S/N ratio) and perfect model correlation (PMC), while prediction skill is computed by the temporal anomaly correlation coefficient (TCC). All models show a decrease in potential predictability and prediction skill with an increase in lead time. The CFSv2 and NASA models show higher PP, which indicates a high potential predictive skill for summer AP-SAT in these models. However, both models show quite low values of TCC all over the AP domain, which is an indication of overconfident predictions. The three geophysical fluid dynamics laboratory models show high prediction skill at both leads. An essential finding of the predictive analysis (PMC and TCC) is that the MME does outperform the individual model at both leads for summer AP-SAT. Each model captures the observed relationship between spatially averaged AP-SAT with sea surface temperature (SST) and 200 hPa geopotential height (Z200) during JJA, with varying details. Persistent model biases impact negatively model predictability and skill, and better AP-SAT and SST teleconnection pattern in models lead to higher predictability. Improvements in initial conditions, model physics, and larger ensemble size are necessary elements to enhance the summer AP-SAT potential predictability and skill.

Keywords Arabian peninsula · Saudi Arabia · Hajj · Air temperature · JJA · NMME · Potential predictability · Atlantic Ocean

1 Introduction

Skillful prediction of surface air temperature (SAT) over the Arabian peninsula (See supplementary material, Figure and Text S1), is of great importance due to many societal endeavors. SAT predictability at seasonal time scale has not been examined before over the Arabian peninsula (AP). This topic has a massive societal relevance now because during the next decade, "Hajj," the largest annual gathering of Muslims (several million), will occur during boreal summer months (June–July–August: JJA). Millions of Muslims from across the world will travel to Saudi Arabia to perform the Hajj pilgrimage in the holy city of Makkah and spend part of that time outdoors. Therefore, skillful seasonal prediction of summer AP-SAT could help government agencies to adopt timely appropriate adaptation strategies that may

✉ Muhammad Azhar Ehsan
mehsan@ictp.it; azhar.ehsan82@gmail.com

1 Earth System Physics Section, International Centre for Theoretical Physics (ICTP), Trieste, Italy

2 Center of Excellence for Climate Change Research (CECCR), King Abdulaziz University, Jidda, Saudi Arabia

3 Department of Applied Physics and Applied Mathematics, Columbia University, New York, USA

provide ease and comfort to pilgrims performing the ritual while outdoor.

The temperature over the AP regularly reaches 40 °C during the summer months. Recently, however, the temperature in Jeddah broke all records and reached 52 °C, on 22 June 2010. Based on the ensemble of high-resolution regional climate model simulations, it has been claimed that the future SAT in some regions of the Arabian peninsula may likely approach and exceed the critical threshold limit of human habitability (Pal and Eltahir 2015). Studies have shown a substantial increase in annual as well as seasonal mean temperature over Saudi Arabia, which represents 80% of the AP (e.g., Almazroui et al. 2012a, b). These changes indicate a warming trend, and an increase of temperature-related extremes, both in magnitude and frequency over the region (Lelieveld et al. 2012; Almazroui et al. 2014; Mustafa and Rahman 2018; Abutaleb et al. 2018). Previous studies have suggested relationships between large-scale forcings and variability of the temperature of Saudi Arabia (Almazroui 2012a, b) and United Arab Emirates (Chandran et al. 2015) using correlation analysis. AP-SAT tends to increase (decrease) during the negative (positive) phase of North Atlantic Oscillation (NAO) and Arctic Oscillation (AO) in almost all seasons, but their impact is more robust during boreal winter. The warm phase of the El Niño-Southern Oscillation (ENSO) during boreal springtime tends to decrease AP-SAT. The primary climatic memory on the timescales of months to years is located in the tropical oceans (Ropelewski and Halpert 1986; 1987). However, prediction skill for land surface temperature or even sea surface temperature (SST) outside the Tropics is irregular and remains quite low.

Several studies have examined the temperature and precipitation variability over Northern Hemisphere during different seasons (Li et al. 2013; Coumou et al. 2018; Chen and Wu 2017; Chen et al. 2016; Chooprateep and McNeil 2016; Miralles et al. 2014) including some studies specifically focusing over the Middle East and North Africa (MENA) region (de Vries et al. 2013; Donat et al. 2014; Evan et al. 2015; Sun et al. 2017; Ehsan et al. 2017a; Almazroui et al. 2017; de Vries et al. 2018). The boreal summer climate variability over Arabian peninsula is linked to various large-scale and regional processes, such as El Niño Southern Oscillation, Indo-Pacific warm pool, tropical North and South Atlantic, Western Hemisphere warm pool, the North Atlantic and Arctic Oscillations as well as South Asian summer monsoon through well-known monsoon-desert mechanism (Rodwell and Hoskins 1996; Almazroui 2012a, b; Hasanean and Almazroui 2017; Abid et al. 2018; Attada et al. 2018a, b). A wide range of seasonal climate predictability studies have been conducted previously by using the coupled global climate model (CGCM) prediction data over different regions, seasons and climate variables

(Tang et al. 2014a, b; Jia et al. 2015; Tippett et al. 2015; Abid et al. 2016; Ehsan et al. 2017b; Osman and Vera 2017; Delsole et al. 2017; Vigaud et al. 2018; among many others). The North American Multimodel Ensemble (NMME) is a research and operational multimodel seasonal forecasting system consisting of coupled models from US and Canadian modeling centers (Kirtman et al. 2014). The project provides a real-time seasonal forecast, as well as a comprehensive set of hindcast datasets for model performance evaluation and calibration. The NMME database availability provides a unique opportunity to examine predictability and prediction skill on seasonal time scales worldwide. Temperature predictability at seasonal time scale has not been examined before over the Arabian peninsula (AP). The main aim of this work is then to document the summer AP-SAT potential predictability and prediction skill depicted by individual and multimodel ensemble (MME) approach.

The article is organized as follows. Section 2 introduces the NMME forecasting system, prediction, observational and reanalysis datasets. It also describes the methodology used to assess the potential predictability and skill of the summer AP-SAT. Observed and predicted mean and variability analysis of summer AP-SAT, potential predictability and skill assessment as well as observed and predicted teleconnection patterns are presented in Sect. 3. A summary and conclusions are given in Sect. 4.

2 Data and methodology

2.1 NMME data

Prediction data used in this work come from the North American Multimodel Ensemble (NMME) project (Kirtman et al. 2014). The NMME is an ensemble forecasting system consisting of coupled atmosphere–ocean models from US and Canadian research and operational centers. Real-time predictions have been started since August 2011, and there is hindcast available (forecast in the past) for each model that includes the period from 1982 to 2010. Throughout the NMME project, some models have been left and new models introduced, and some upgraded to new versions. Also, the initialization method, prediction length and ensemble members vary by model. From all the models available in the NMME collection, we selected the six that had the most extended forecast period in common. The models used in this study are listed in Table 1. Here we use integration with May and April start dates from the hindcast period (1982–2010) and the real-time period (2011–2017) for a total 36 years. We make no distinction between hindcast and real-time forecast in our analysis and refer to both as predictions. Predictions for the lead time of up to 2 months are analyzed in this study. "Lead-1-month or Lead-1″ prediction is

Table 1 Six NMME models used in this work

Model	Acronym used	Ensemble members	Native atm. res.	Native ocn. res.	Prediction length (months)	References
NCEP-CFSv2	CFSv2	24	T126L64	MOM4L40 0.25° Eq.	10	Saha et al. (2014)
NASA-GAMO-062012	NASA	11[a]	1.0° × 1.25° L72	MOM4L40 0.25° Eq.	9	Vernieres et al. (2012)
COLA-RSMAS-CCSM4	COLA	10	1.25° × 0.9° L26	POPL60 0.27° Eq.	12	Gent et al. (2011)
GFDL-CM2p1-aer04	GFDL-Aer	10	2.0° × 2.25° L24	MOM4L50 0.3° Eq.	12	Delworth et al. (2006)
GFDL-CM2p5-FLOR-A06	GFDL-FLOR-A	12	C18L32 (50 km)	MOM4L50 0.3° Eq.	12	Vecchi et al. (2014)
GFDL-CM2p5-FLOR-B01	GFDL-FLOR-B	12	C18L32 (50 km)	MOM4L50 0.3° Eq.	12	Vecchi et al. (2014)

[a]Only June start date have 12 members. No real-time forecast now available for this version. NASA switched to newer version

based on May initial conditions (IC), and it is the prediction for June. The prediction for May itself considered as "Lead-0-month or Lead-0″ prediction. Following this, the seasonal Lead-1 forecast is for the three months (June–August: JJA) following the initial month: in the May IC example. Hindcast and real-time prediction of monthly averages of surface air temperature, sea surface temperature and geopotential height at 200 hPa are available for download from the International Research Institute for Climate and Society (IRI) (https://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/).

2.2 Observational data

We use several monthly mean observational and reanalysis datasets, including geopotential height from NCEP-DOE (National Centers for Environmental Prediction-Department of Energy) AMIP-II Reanalysis products (2.5° × 2.5°) (Kanamitsu et al. 2002), land surface temperature from the Global Historical Climatology Network + Climate Anomaly Monitoring System (0.5° × 0.5°) (GHCN + CAMS v2: Fan and Dool 2008), and sea surface temperature (SST) from the Hadley Centre SST dataset HadISST1.1 (1.0° × 1.0°) (Rayner et al. 2003). All these observational datasets are at different resolutions and were converted to a (1.0° × 1.0°) resolution using bilinear interpolation common to NMME prediction data.

2.3 Statistical methods

This work presents an assessment of potential predictability (PP) and skill of boreal summer (June–July–August: JJA) surface air temperature over the Arabian peninsula from initial conditions observed in April and May (Lead-2 and 1). We define the Arabian peninsula (AP) domain to be all the land points included within 35°–60°E and 12°–32°N. Boreal summer seasonal anomalies for each model and observation are computed relative to the 36-year climatology period 1982–2017. Different statistical measures (mean, standard deviation, correlation, bias)

are used to validate the model predictions. The Taylor diagram (Taylor 2001) is employed to summarize the performance of individual model. The relevant statistics are the pattern correlation, the ratio of the normalized root-mean-square difference and relative bias (%) estimation to quantify the each model inaccuracy as compared to observation (Details about Taylor diagram can be found at https://www.ncl.ucar.edu/Applications/taylor.shtml). For a good model, the correlation between prediction and observation is high, root-mean-square different must be small, and the variances must be similar.

The potential predictability of summer AP-SAT is assessed as the interannual variability of the ensemble mean, and the variability of individual members from the ensemble mean (Rowell 1998), also known as "Signal or external variance," and "Noise or internal variance" respectively. Following Rowell et al. (1995), internal variance or noise can be expressed as;

$$\text{Noise} = \frac{1}{N(n-1)} \sum_{k=1}^{N} \sum_{m=1}^{n} \left(T_{km} - \overline{T}_k \right)^2,$$

where T is the surface air temperature, k indicates the individual year, m denotes ensemble members, and $\overline{T}_k$ is the ensemble mean. The external variance or signal can be expressed as;

$$\text{Signal} = \frac{1}{N-1} \sum_{k=1}^{N} \left(\overline{T}_k - \overline{T} \right)^2,$$

where $\overline{T}$ is the climatological mean of the ensemble mean given by $\overline{T} = \frac{1}{Nn} \sum_{k=1}^{N} \sum_{m=1}^{n} T_{km}$. The ratio of signal and noise variances (S/N ratio) defines potential predictability (Kang and Shukla 2006). The potential predictability for summer AP-SAT is measured here in terms of classical "perfect model correlation: PMC". It is the correlation calculated by considering one-member prediction as observation and ensemble average of the rest of the members (Ehsan et al.

2013) and can provide predictive information available in that ensemble member. Following the number of ensemble members, correlations are averaged, after applying Fisher Z transformation (Faller 1981). Finally, the prediction skill of summer AP-SAT is assessed here is through the computation of the temporal anomaly correlation coefficient (TCC)
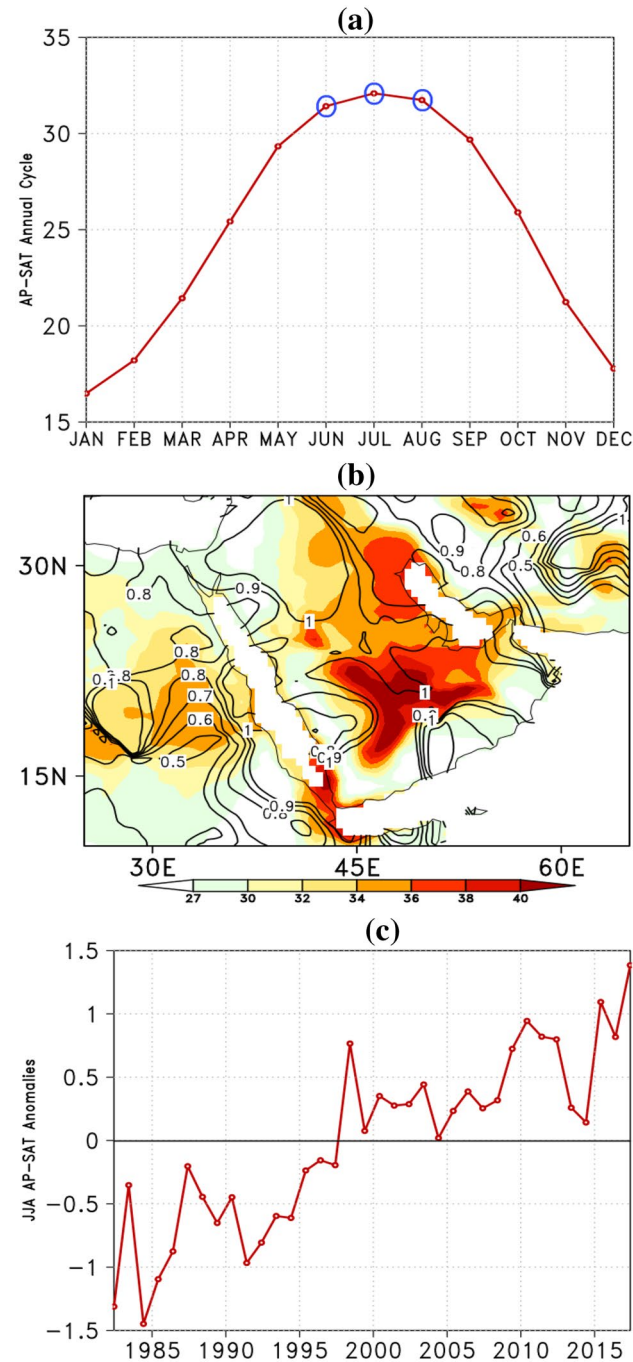

(a)

(b)

(c)

Fig. 1 **a** Annual cycle of the surface air temperature (SAT) over the AP domain. **b** Summer SAT climatology (shaded) and standard deviation (contour) during the period 1982–2017. **c** The area averaged summer mean SAT anomaly time series over the AP. Unit of SAT is °C

maps between the ensemble mean and observed time series. The statistical significance of the correlation is estimated by using the t test (Wilks 2006). Our threshold for significance is 0.05.

Spatially average observed (Fig. 1c) and predicted (Fig. 5a) summer SAT over AP exhibits a strong positive trend, which is undoubtedly related to global warming or climate change. We remove the long-term linear trend in all variables (observation and predictions) for all 36 summers by using the least squares method. The idea behind removing the long-term trend is to remove the global warming signal from the original time series, which has a stronger influence in the surface air temperature. Consequently, calculation of summer AP-SAT potential predictability (Signal, Noise, S/N ratio, PMC), skill (TCC) and teleconnection analysis are all based on the detrended data as shown in Fig. 5b.

3 Results

3.1 Observed and predicted summer AP-SAT mean and variability

The annual cycle of the AP-SAT presented in Fig. 1a shows temperature (> 30 °C) from June through August, with July as the hottest month and boreal summer (June–August: JJA) is the warmest season of the AP. JJA climatology of SAT over the AP for the period 1982–2017 is shown in Fig. 1b (shaded). The summer SAT along the Red and Arabian seas coastlines remain moderate (< 30 °C). The strong
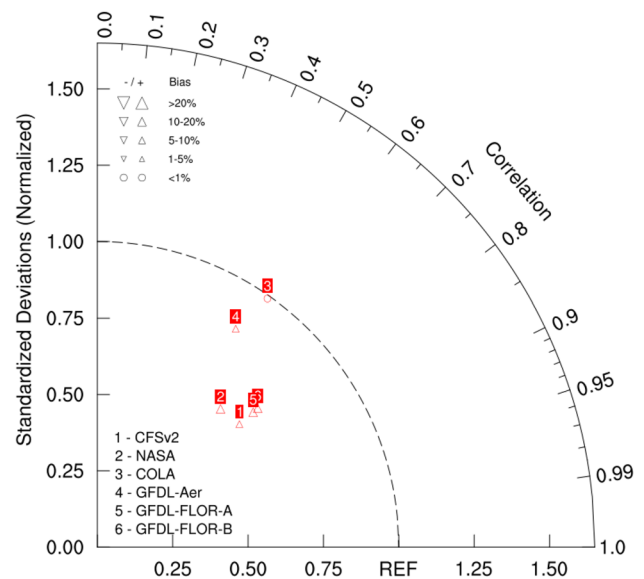

Fig. 2 Taylor diagram describing the six model's performance in predicting summer AP-SAT (initialized in May/Lead-1) in comparison to observation

cross-equatorial flow along the eastern side of the AP during summer induces upwelling over the western Arabian Sea, which cools the SST and consequently decreases temperatures over the eastern side of the AP particularly over Yemen and Oman. However, the Arabian Gulf side depicts high summer SAT (> 35 °C), which may be due to its shallow depth (See Supplementary Material, Figure and Text S1), low albedo and relatively shallower boundary layer depth that retain heat and moisture closer to the surface. Away from coastlines, the summer SAT distribution is more uniform, with highest temperature (> 40 °C) observed over the

central-eastern region. The summer AP-SAT standard deviation varies between 0.8 °C and 1.5 °C (Fig. 1b: contour). A high spatial variation in summer SAT is evident over the northern and southwestern regions of the AP, which shows a noticeable regional temperature variability over the AP. The time series of spatially averaged summer SAT over the AP domain shows increasing trend (Fig. 1c) during the period of 1982–2017, which is in agreement with earlier findings (Almazroui et al. 2012b).

Before proceeding to investigate the potential predictability and skill of summer SAT, we examine the fidelity of six
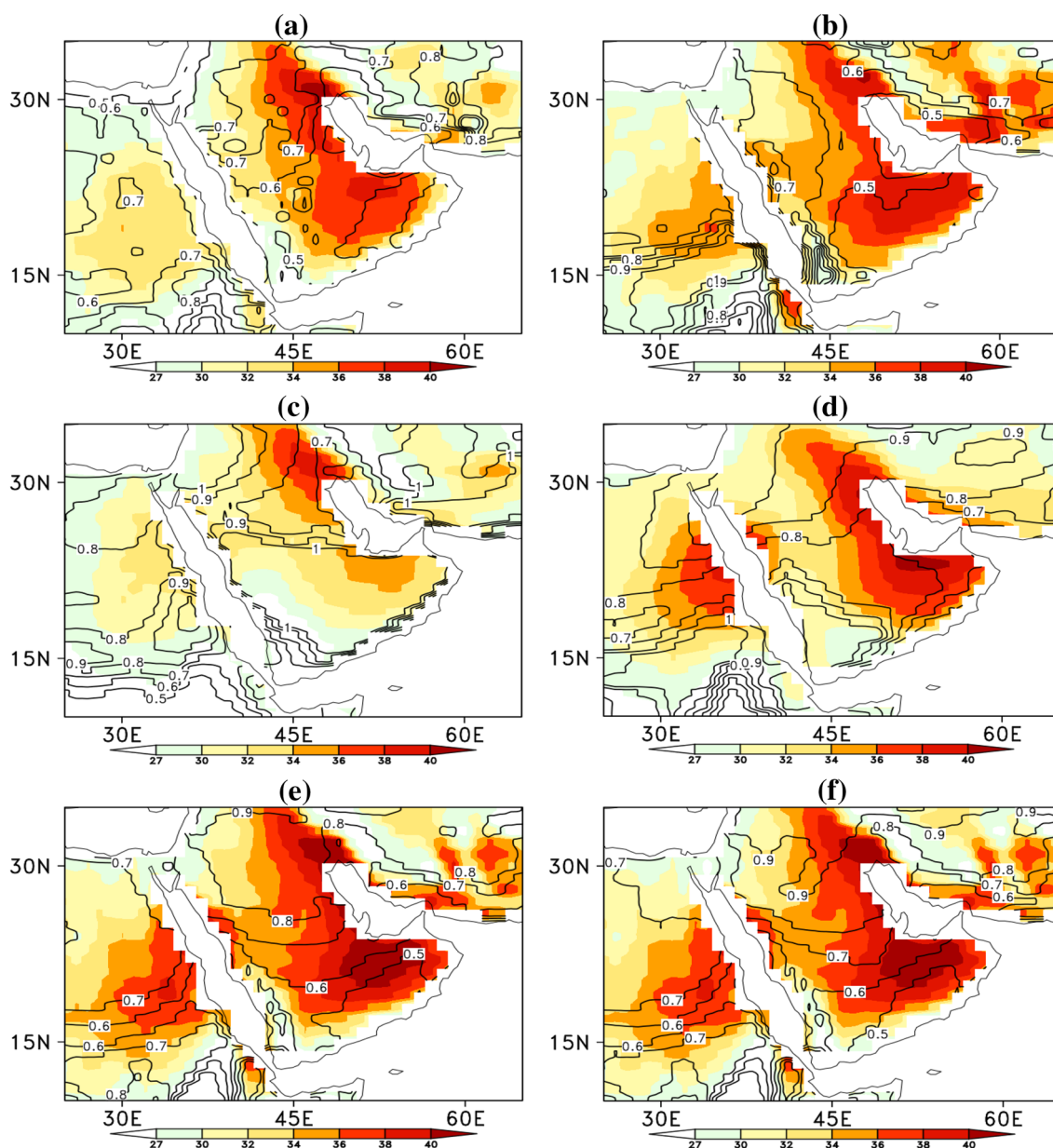

Fig. 3 Predicted (initialized in May/Lead-1) summer SAT climatology (shaded) and standard deviation (contour) for **a** CFSv2, **b** NASA, **c** COLA, **d** GFDL-Aer, **e** GFDL-FLOR-A, and **f** GFDL-FLOR-B during the period 1982–2017. Unit of SAT is °C

NMME models in predicting the summer SAT climatology and variability at Lead-1 over the AP. Immediate visualization of differences in the performance of each prediction model is obtained by using the Taylor diagram (Taylor 2001) as shown in Fig. 2. The pattern correlation between different model predictions and observation is found ranging from 0.54 to 0.76. The normalized standard deviation also diverges from the reference value for all models except COLA. Also, all models show positive relative bias as compared to the observation, except COLA model that shows negative SAT bias. As the Taylor diagram show summary statistics, it is desirable that spatial maps may be presented

to highlight further details. The spatial distribution of predicted summer SAT climatology (shaded) and interannual variability represented by the standard deviation (contour) over the AP is shown in Fig. 3. Relative to the observation, all models strongly overestimate summer SAT along the periphery of the AP and underestimated in the central parts of the AP. Except for COLA, all models underestimated the summer SAT variability over the AP. Overall CFSv2 shows less mean bias (Fig. 4a). COLA (Fig. 4c) and GFDL-Aero (Fig. 4d) show a strong negative bias over the central-eastern AP. The significant error in predicted SAT may be related to the way how the models handle the surface-atmosphere
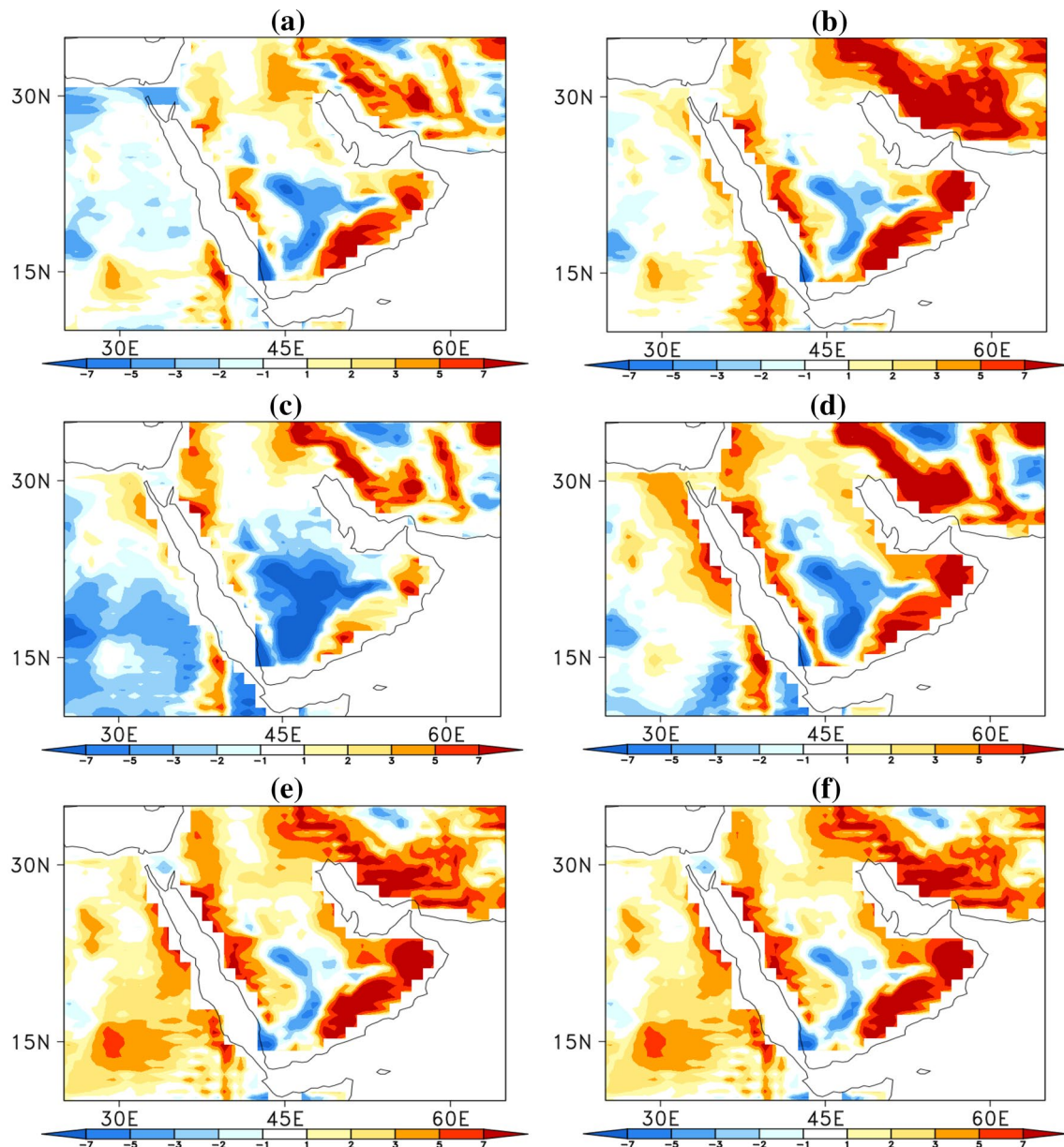

Fig. 4 Mean summer SAT bias (Model–OBS) estimated based on 36-year prediction data (initialized in May/Lead-1) for **a** CFSv2, **b** NASA, **c** COLA, **d** GFDL-Aer, **e** GFDL-FLOR-A, and **f** GFDL-FLOR-B. Unit of SAT is °C

coupling processes, including the input land-surface forcing data (Bonan 2008; Rahman et al. 2018).

Spatially averaged predictions of summer SAT over AP depict gradually increasing trend (Fig. 5a) and statistically significant correlation coefficients between observed and predicted summer AP-SAT time series. However, the COLA model shows low correlation coefficient (0.6), which shows the modest performance of this particular model. The higher correlation between predicted and observed anomalies is indeed related to global warming or climate change that has a stronger influence in the surface air temperature. The detrended summer AP-SAT time series is shown in Fig. 5b. After removing long-term trend, correlation coefficient in predicted and observed summer AP-SAT time series is reduced (about 40%), which shows the substantial impact of global warming signal over summer AP-SAT. In our gradually warming climate, recent years are warmer, so a summer forecast for above average temperature relative to its climatology can be skillful on its own. Thus, the potential predictability and skill assessment of summer AP-SAT are based on the detrended data.

3.2 Predictability assessment of individual and multi-model ensemble

The spatial distributions of signal, noise, and signal-to-noise ratio are illustrated in Figs. 6, 7, and 8, respectively, after removing trend from prediction data (initialized in May/Lead-1) for each model. Spatial distribution and magnitude of signal and noise variances over the AP domain vary by model. Signal variance is quite weak (Fig. 6) as compared to noise variance, which shows quite high values in all models (Fig. 7). Northern and south-western AP regions show high signal variance, while noise variance is almost uniformly distributed over the whole AP domain. Models like COLA and GFDL-Aer that show large signal variance also tend to have even larger noise variance. The CFSv2, NASA, GFDL-FLOR-A, and GFDL-FLOR-B show higher S/N ratio over the AP (Fig. 8). Figure 9 shows the geographical distribution of perfect model correlation (ensemble member against ensemble mean) as estimated with six models individually. PMC values show the upper limit of the dynamical seasonal prediction skill that can be reached with a perfect model (a model that forecasts perfectly its own climate) and perfect boundary conditions. The figure indicates that the potential predictability of the summer AP-SAT is relatively low (PMC ranging between 0.3 and 0.6). The CFSv2 and NASA models show higher PP, which indicates a high potential predictive skill for summer AP-SAT in CFSv2 (Fig. 9a). Figure 10 shows the spatial distribution of prediction skill
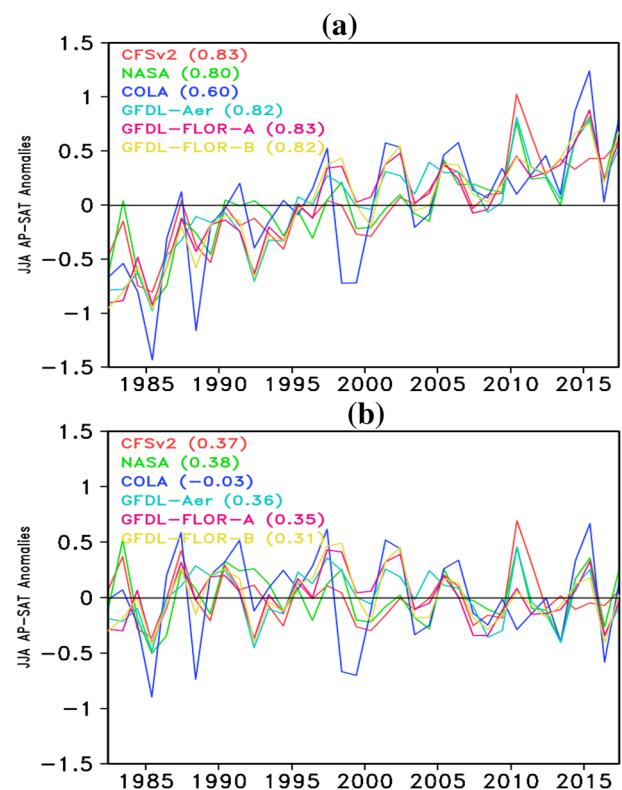

Fig. 5 Predicted (initialized in May/Lead-1) summer AP-SAT anomalies (spatially averaged over AP domain) for **a** undetrended, and **b** detrended data for the period 1982–2017. Each model is described in different color. For undetrended data all NMME models show statistically significant correlation (number show in bracket). For detrended data CFSv2, NASA, GFDL-Aer, GFDL-FLOR-A show statistically significant, while COLA and GFDL-FLOR-B show nonsignificant correlation at 5% confidence level, using a *t* test. Unit of SAT is °C

for summer SAT in terms of TCC (ensemble mean against observation). In contrast to PMC, CFSv2 and NASA models (Fig. 10a, b) show quite low values of TCC all over the AP domain, which is an indication of overconfident forecasts (Eade et al. 2014) in which ensemble members agree well with each other (high PMC) but do not capture the observed variations (low TCC). The two versions of GFDL (FLOR-A and FLOR-B) show PP and skill of the summer AP-SAT quite close to each other, because two versions have identical atmospheric, land, and sea ice configurations but have slightly different ocean setups (Vecchi et al. 2014). The spatial distributions of PMC and TCC for Lead-2 prediction data (initialized in April) are elucidated (See Supplementary Material, Figures S2 and S3). All models show a decrease in potential predictability and prediction skill with an increase in lead time.
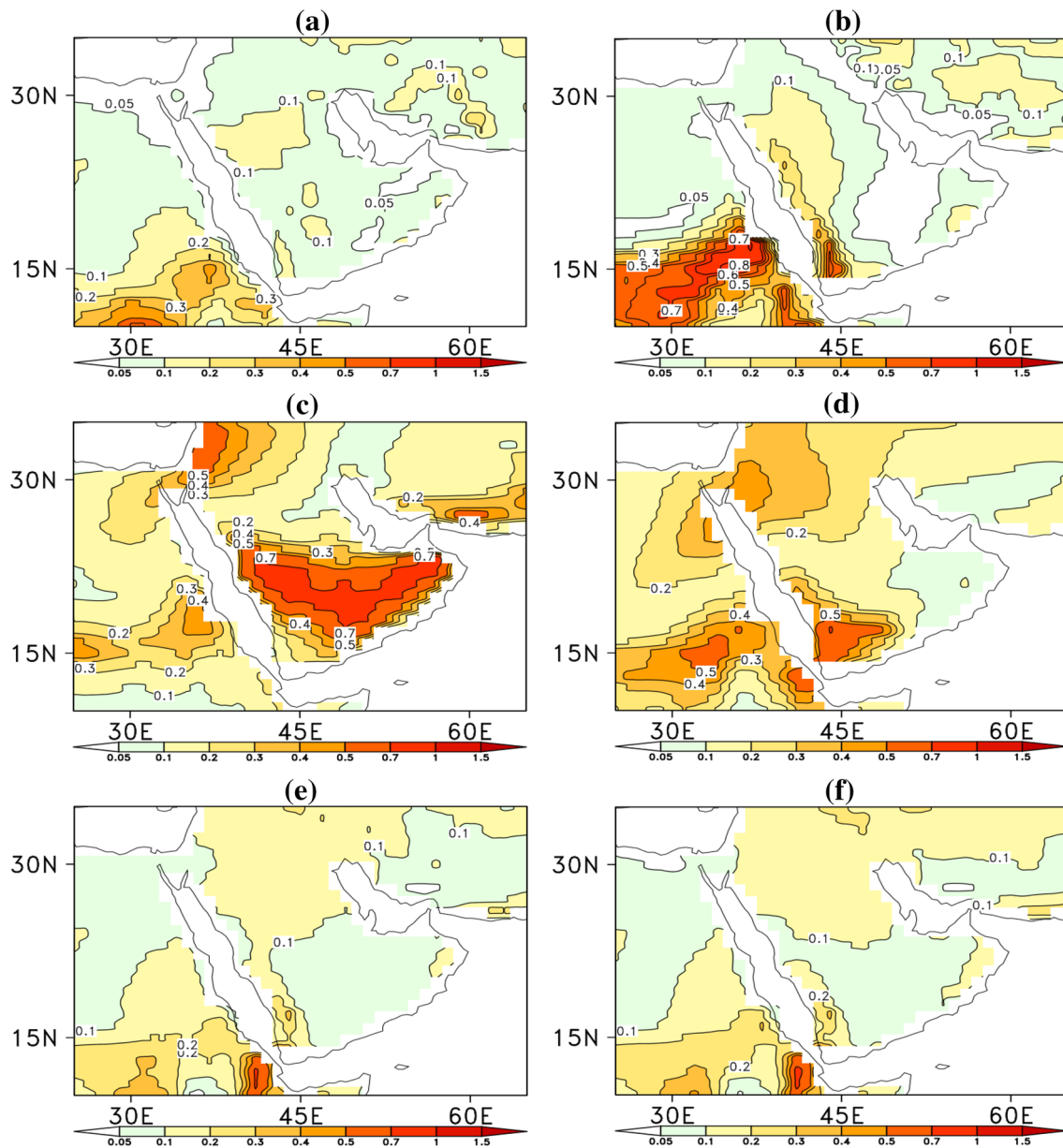
Fig. 6 Signal variance estimated based on 36-year prediction data (initialized in May/Lead-1) for **a** CFSv2, **b** NASA, **c** COLA, **d** GFDL-Aer, **e** GFDL-FLOR-A, and **f** GFDL-FLOR-B. Unit of signal variance is °C^2

Finally, we now demonstrate the potential predictability and skill of a multi-model ensemble (MME: a combination of six models and 79 ensemble members). Several previous studies have confirmed the superiority of multi-model predictions over that of a single-model (e.g., Palmer et al. 2004; Hagedorn et al. 2005; Kang and Yoo 2006; DelSole and Tippett 2014; DelSole et al. 2014). The potential predictability and skill of MME at Lead-1 and Lead-2 is shown in Fig. 11. The figure immediately reveals the advantage of the MME

(higher PMC and TCC) compared to the individual models (Figs. 9, 10) in predicting summer AP-SAT.

3.3 Teleconnections pattern in observation and prediction data

The goal of this discussion is to identify relationships between summer AP-SAT and global JJA SST and other atmospheric parameters in observation, and their
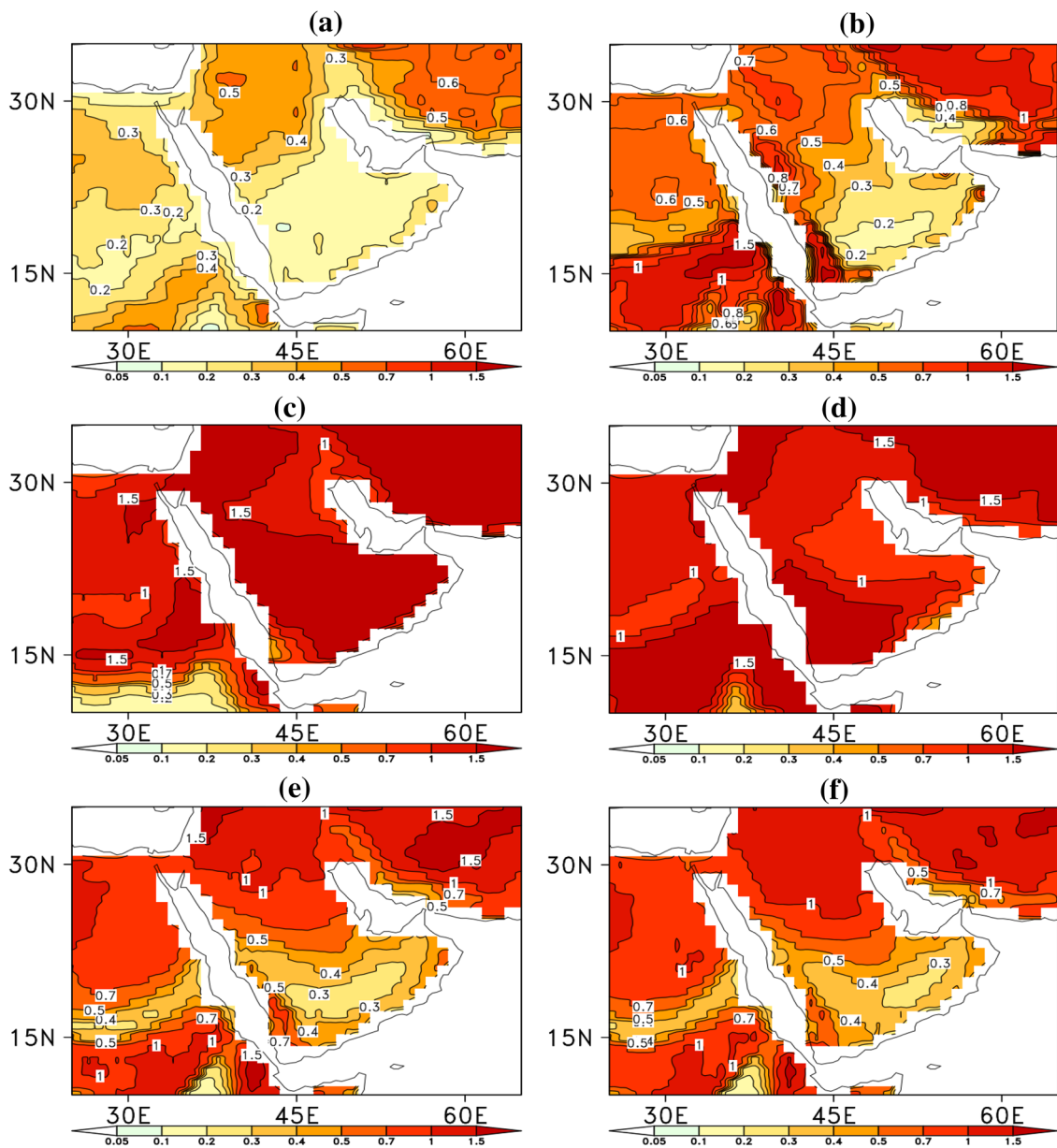
Fig. 7 Noise variance estimated based on 36-year prediction data (initialized in May/Lead-1) for **a** CFSv2, **b** NASA, **c** COLA, **d** GFDL-Aer, **e** GFDL-FLOR-A, and **f** GFDL-FLOR-B. Unit of noise variance is °C^2

reproducibility in prediction data with an argument that better teleconnection leads to higher potential and actual predictability (Ehsan et al. 2017b). Figure 12a shows the correlation between area-averaged AP-SAT and SST anomalies during JJA for the period 1982–2017. Summer AP-SAT is highly positively correlated with the equatorial Indian Ocean and tropical North Atlantic SSTs along with local SSTs (located in Mediterranean and Red seas and Arabian Gulf) and a nonexistence of signal over Pacific region

(commonly known as ENSO). We plotted the correlation between summer AP-SAT and 200 hPa geopotential height (Z200) to show the influence of large-scale circulation on the summer AP-SAT (Fig. 12b). Positive Z200 anomalies were found over the AP and whole Mediterranean and eastern Europe region and negative values are observed over the northern Pakistan and Afghanistan. We also calculated the correlation between the summer AP-SAT and atmospheric thickness (not shown) from the surface to the upper level
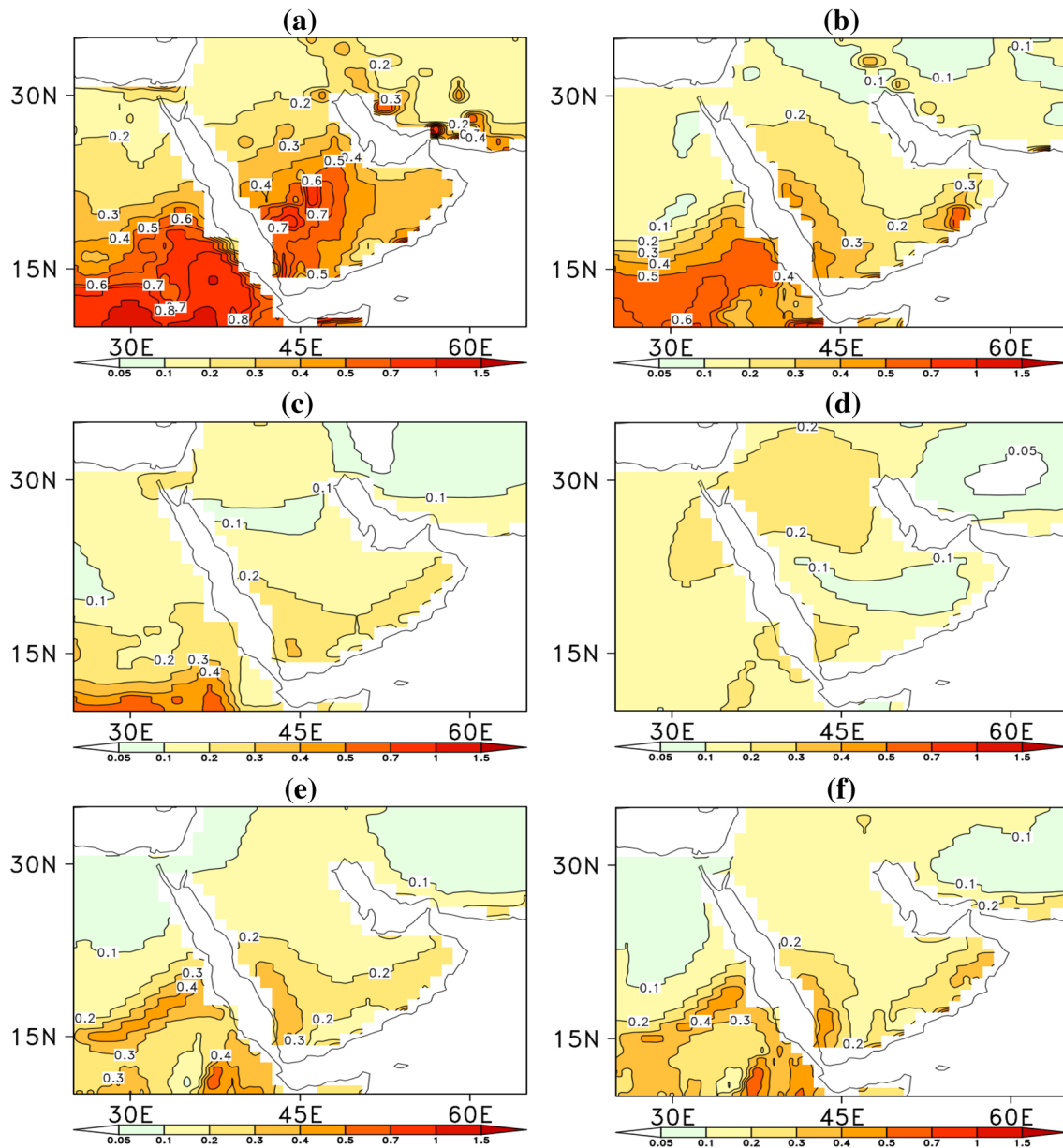
Fig. 8 Signal-to-Noise ratio (S/N ratio) estimated based on 36-year prediction data (initialized in May/Lead-1) for **a** CFSv2, **b** NASA, **c** COLA, **d** GFDL-Aer, **e** GFDL-FLOR-A, and **f** GFDL-FLOR-B. S/N ratio is unitless

(1000–300 hPa). This is in good agreement with the above normal Z200 (Fig. 12b) and closely related with the warm temperature anomalies over the AP during summer. The statistically significant correlation between summer AP-SAT and SST anomalies in the equatorial Indian (EIO: 40–80 E,

10 S–15 N), and tropical North Atlantic (ATL: 300–340 E, 0–23 N) oceanic regions is 0.63 and 0.61, respectively (Fig. 12c). These results are in agreement with the earlier findings (Hasanean and Almazroui 2017; Attada et al. 2018a, b), which documented that summer AP-SAT variability is
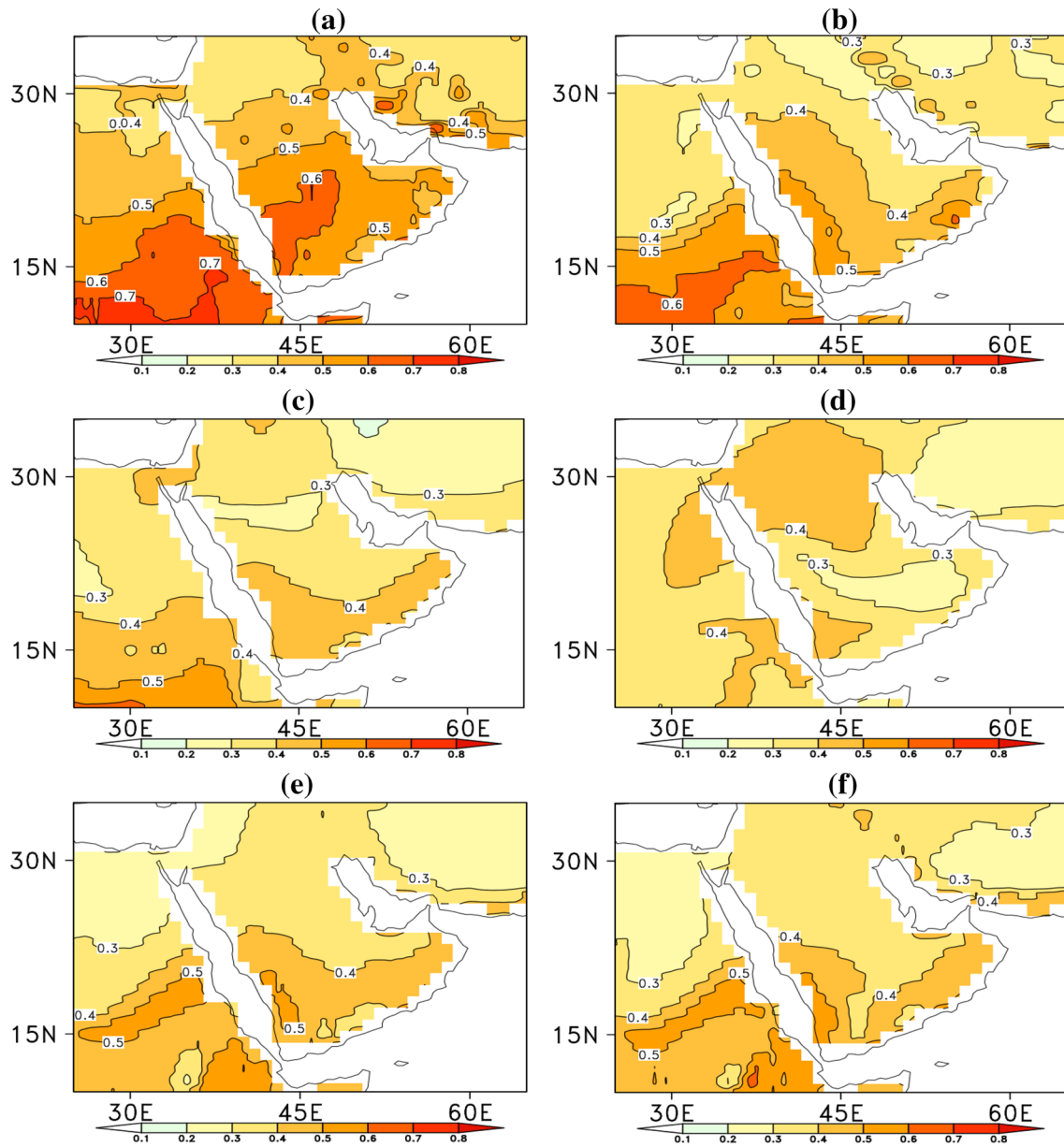
Fig. 9 Perfect model correlation estimated based on 36-year prediction data (initialized in May/Lead-1) for **a** CFSv2, **b** NASA, **c** COLA, **d** GFDL-Aer, **e** GFDL-FLOR-A, and **f** GFDL-FLOR-B. A correlation coefficient higher than 0.32 is statistically significant at 5% confidence level, using a *t* test
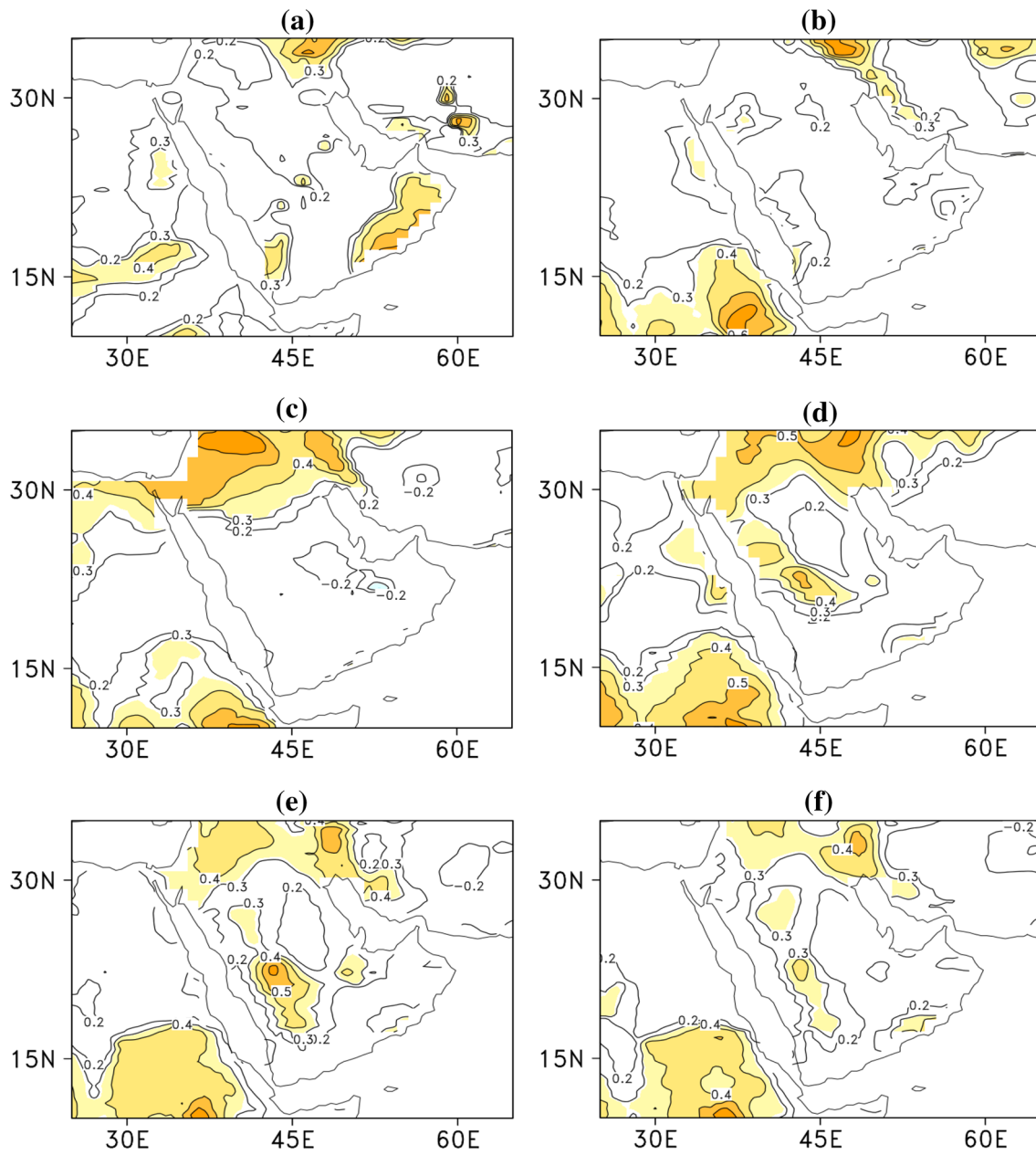
Fig. 10 Prediction skill (correlation between ensemble mean and observed detrended anomalies) for **a** CFSv2, **b** NASA, **c** COLA, **d** GFDL-Aer, **e** GFDL-FLOR-A, and **f** GFDL-FLOR-B at Lead-1. A correlation coefficient higher than 0.32 is statistically significant at 5% confidence level, using a *t* test
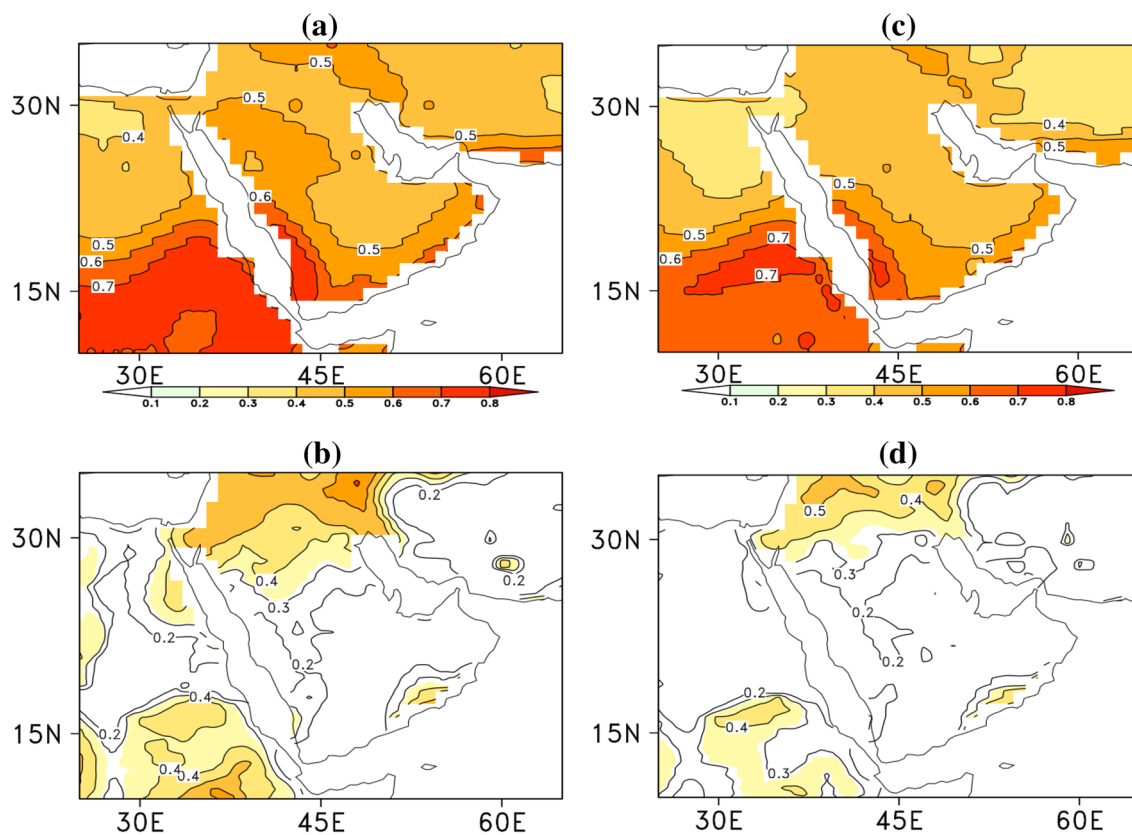
Fig. 11 a Perfect model correlation for MME (six models and 79 members) estimated based on 36-year prediction data (initialized in May/Lead-1) and **b** prediction skill (correlation between MME and observed detrended anomalies). **c, d** same as (**a** and **b**) but for predic-tion data initialized in April/Lead-2. A correlation coefficient higher than 0.32 is statistically significant at 5% confidence level, using a *t* test

associated with the SSTs located in Indo-Pacific warm pool, tropical North and South Atlantic Ocean and western hemisphere warm pool regions. The geographical distribution of correlation between predicted AP-SAT and SST anomalies in each model for JJA is shown in Fig. 13. Each model captures the observed JJA AP-SAT-SST relationship, with varying details. NCEP (Fig. 13a) and NASA (Fig. 13b) models show strong correlation over the equatorial Indian Ocean as well as North Atlantic Ocean, which is even stronger than the observation (Fig. 12c). Three GFDL models (GFDL-Aer, GFDL-FLOR-A and GFDL-FLOR-B) also captures the observed JJA AP-SAT-SST relationship as shown in Fig. 13d–f. The JJA AP-SAT-SST teleconnection pattern in COLA model (Fig. 13c) shows strong positive correlation in Pacific and weak negative values in Atlantic and Indian Ocean basin, which is quite different from the observed pattern. This diverse behavior of COLA model is also evident in the JJA AP-SAT-Z200 analysis, which shows strong negative values over the northern AP and whole Mediterranean and eastern Europe region as compared to observed correlation (Fig. 12b). This clearly indicates that the predicted teleconnection pattern in the COLA model is quite different from the observed pattern, that may impact the summer AP-SAT predictability in this model. GFDL-FLOR-A and B (Fig. 14e, f) captures the JJA AP-SAT-Z200 quite well as compared to observation (Fig. 12b), which could cause a moderate level of potential predictability and skill in these models.

4 Summary and conclusions

Forecasting the seasonal climate over a region is an important issue, both for the physical perspective of the climate drivers and for decision makers to have sufficient time to take pre-emptive actions. In this study, we assess the
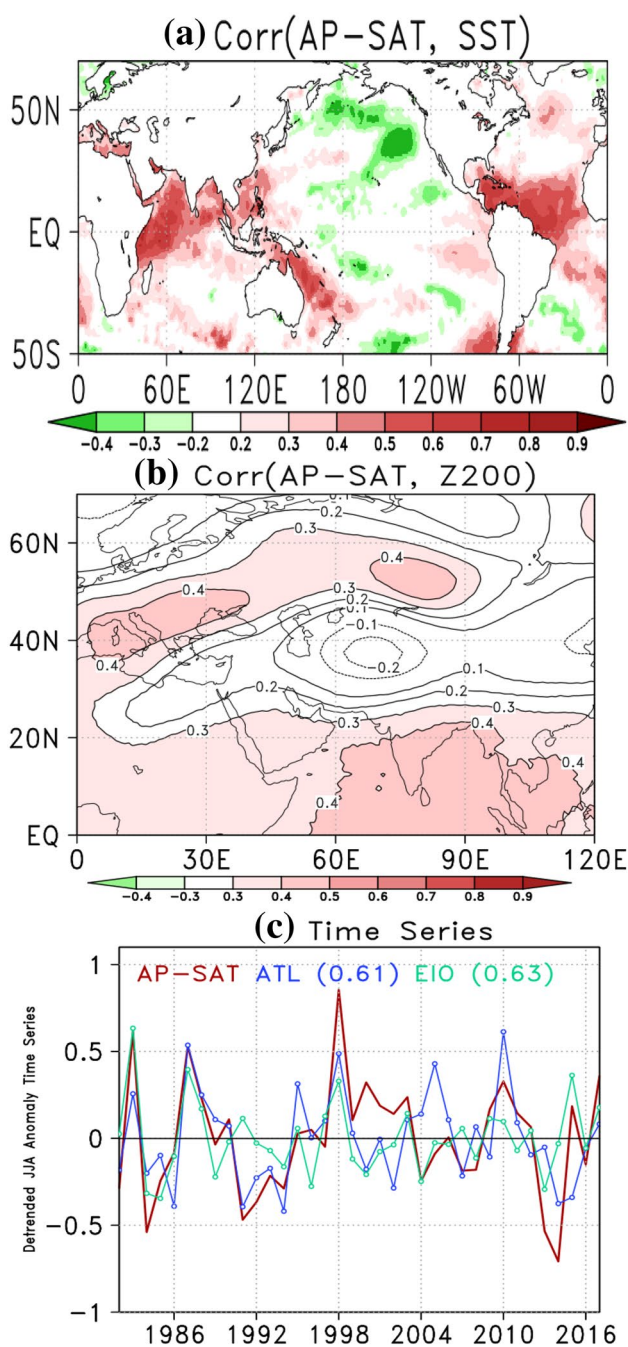
Fig. 12 The correlation of AP-SAT to **a** SSTs and, **b** Z200 detrended anomalies for JJA at each grid, during 1982–2017. **c** The detrended spatially averaged time series of summer AP-SAT and SST in EIO and ATL regions. The correlation between them is shown in bracket. A correlation coefficient higher than 0.32 is statistically significant at 5% confidence level, using a *t* test

potential predictability and prediction skill of boreal summer surface air temperature over the Arabian peninsula by utilizing seasonal prediction data of six models from the North American Multimodel Ensemble project. Predictions for the lead time of up to 2 months are analyzed in this study. "Lead-1-month or Lead-1″ prediction is based on May initial conditions (IC), and it is the seasonal prediction for the 3 months (June–August), following the initial month of May. The Lead-2-month or Lead-2 prediction is based on the observed April IC. The study was made by considering the single model, and a multimodel ensemble (MME) approach.

Relative to the observation, all models strongly overestimate summer SAT along the periphery of the AP and underestimated in the central parts of the AP. The observed and predicted summer SAT over AP shows a strong positive trend and high correlation coefficient (Before removal of trend) for each model. The correlation coefficient in detrended predicted and observed summer AP-SAT time series is reduced drastically, which shows the significant impact of the global warming signal over the SAT. Therefore, predictability analysis presented in this work is based on the detrended data. The CFSv2 and NASA models show higher PMC, which indicates a high potential predictive skill. However, both models show quite low values of TCC all over the AP domain, which shows an overconfident summer AP-SAT prediction in these models. The three Geophysical Fluid Dynamics Laboratory (GFDL) models show good prediction skill at both leads while the COLA model shows the lowest values. All models show a decrease in potential predictability and prediction skill with an increase in lead time. An essential finding of the predictive analysis (PMC and TCC) is that the MME, which is a combination of six models and 79 ensemble members, does outperform the individual model at both leads. Summer AP-SAT is highly positively correlated with the equatorial Indian Ocean, tropical North Atlantic SSTs along with local SSTs, located in the Mediterranean and Red seas and Arabian Gulf and the absence of signal in the Pacific region. Each model captures the observed relationship between spatially averaged AP-SAT with sea surface temperature (SST) and 200 hPa geopotential height (Z200) during JJA, with varying details. This study implies that persistent model biases impact badly model potential predictability and skill, better teleconnection pattern in prediction data accompanied by larger ensemble size lead to higher predictability of the regional climate.
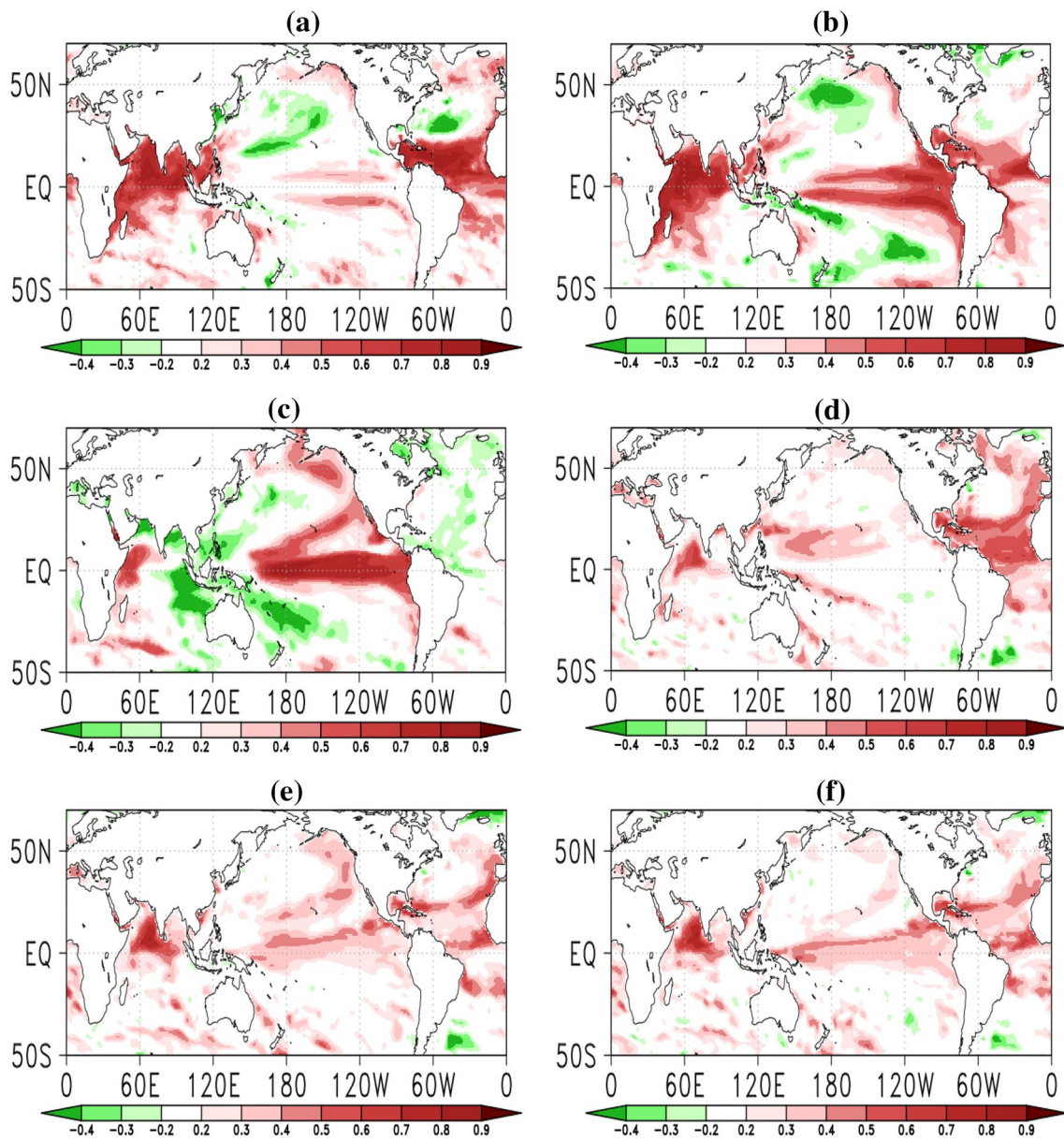
Fig. 13 The correlation AP-SAT to SSTs detrended anomalies for JJA at each grid, during 1982–2017 for **a** CFSv2, **b** NASA, **c** COLA, **d** GFDL-Aer, **e** GFDL-FLOR-A, and **f** for GFDL-FLOR-B, respectively at Lead-1. A correlation coefficient higher than 0.32 is statistically significant at 5% confidence level, using a *t* test
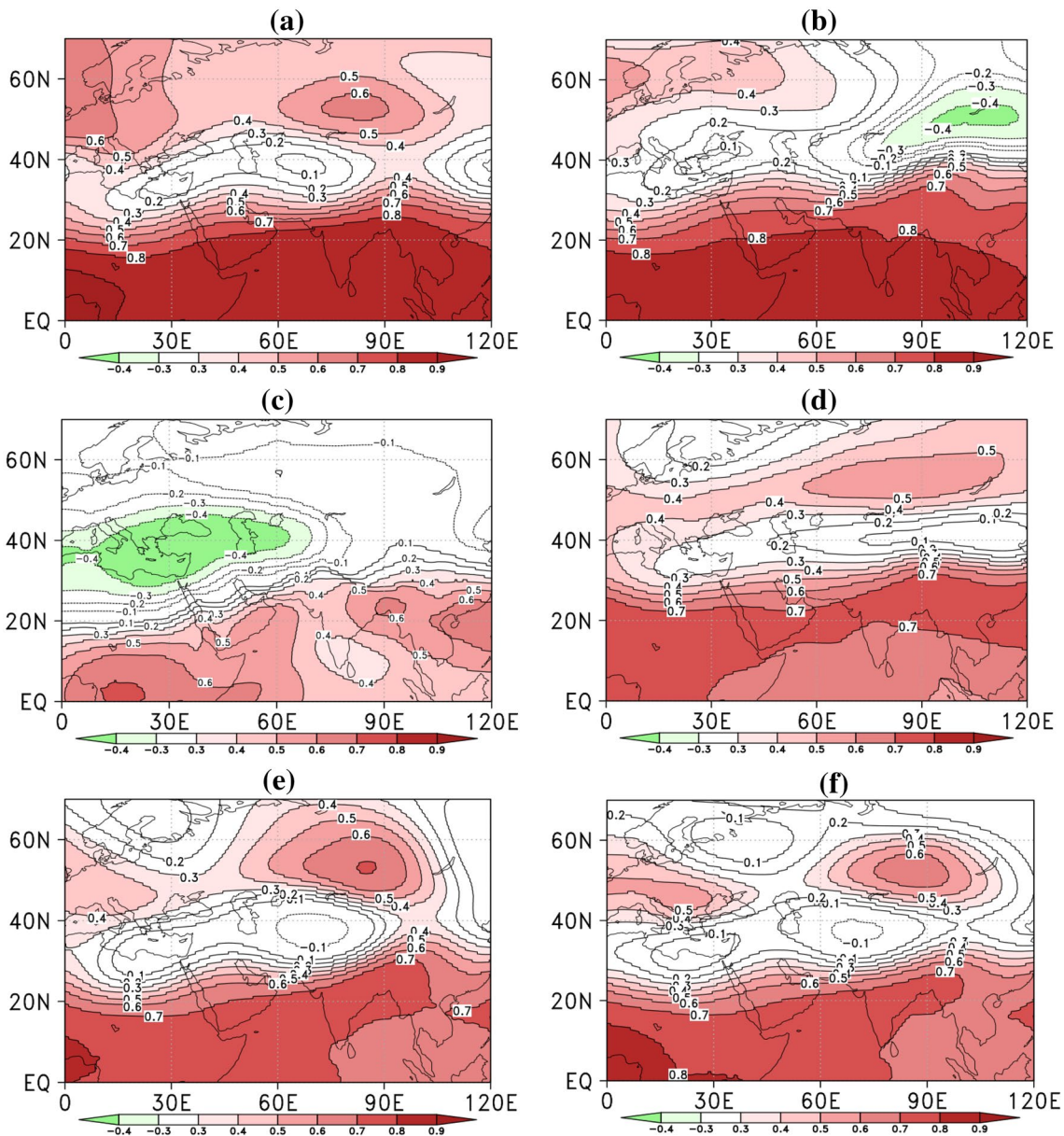
Fig. 14 The correlation AP-SAT to Z200 detrended anomalies for JJA at each grid, during 1982-2017 for **a** CFSv2, **b** NASA, **c** COLA, **d** GFDL-Aer, **e** GFDL-FLOR-A, and **f** for GFDL-FLOR-B, respectively at Lead-1. A correlation coefficient higher than 0.32 is statistically significant at 5% confidence level, using a *t* test

References

Abid MA, Kucharski F, Almazroui M, Kang IS (2016) Interannual rainfall variability and ECMWF-Sys4-based predictability over the Arabian peninsula winter monsoon region. Q J R Meteorol Soc 142:233–242. https://doi.org/10.1002/qj.2648

Abid MA, Almazroui M, Kucharski F, O'Brien E, Yousef AE (2018) ENSO relationship to summer rainfall variability and its potential predictability over Arabian peninsula region. NPJ Clim Atmos Sci 1:20171

Abutaleb KAA, Mohammed AHES, Ahmed MHM (2018) Climate change impacts, vulnerabilities and adaption measures for Egypt's

nile delta. Earth Syst Environ 2:183. https://doi.org/10.1007/s41748-018-0047-9

Almazroui M (2012a) Temperature variability over Saudi Arabia during the period 1978–2010 and its association with global climate indices. JKAU: Meteoro. Environ Arid Land Agric Sci 23:85–108. https://doi.org/10.4197/Met.23-1.6

Almazroui M (2012b) Temperature variability over Saudi Arabia during the period 1978–2010 and its association with global climate indices. JKAU Meteorol Environ Arid Land Agric Sci 1:1. https://doi.org/10.4197/met.23-1.6

Almazroui M, Islam MN, Athar H, Jones PD, Rahman MA (2012a) Recent climate change in the Arabian peninsula: annual rainfall and temperature analysis of Saudi Arabia for 1978–2009. Int J Climatol 32:953–966

Almazroui M, Islam MN, Jones PD, Athar H, Rahman MA (2012b) Recent climate change in the Arabian peninsula: seasonal rainfall and temperature climatology of Saudi Arabia for 1979–2009. Atmos Res 111:29–45

Almazroui M, Islam MN, Dambul R, Jones PD (2014) Trends of temperature extremes in Saudi Arabia. Int J Climatol 34:808–826

Almazroui M et al (2017) Saudi-KAU coupled global climate model: description and performance. Earth Syst Environ 1:7. https://doi.org/10.1007/s41748-017-0009-7

Attada R, Dasari HP, Chowdary JS, Yadav RK, Knio O, Hoteit I (2018a) Surface air temperature variability over the Arabian peninsula and its links to circulation patterns. Int J Climatol 39:445–464. https://doi.org/10.1002/joc.5821

Attada R, Yadav RK, Kunchala RK, Dasari HP, Knio O, Hoteit I (2018b) Prominent mode of summer surface air temperature variability and associated circulation anomalies over the Arabian peninsula. Atmos Sci Lett 1:1. https://doi.org/10.1002/asl.860

Bonan GB (2008) Forests and climate change: forcings, feedbacks, and the climate benefits of forests. Science 320:1444–1449

Chandran A, Basha G, Ouarda TBMJ (2015) Influence of climate oscillations on temperature and precipitation over the United Arab Emirates. Int J Climatol 36:225–235

Chen S, Wu R (2017) Interdecadal changes in the relationship between interannual variations of spring North Atlantic SST and Eurasian surface air temperature. J Clim 30:3771–3787

Chen W, Hong X, Lu R, Jin A, Jin S, Nam J, Shin J, Goo T, Kim B (2016) Variation in summer surface air temperature over northeast Asia and its associated circulation anomalies. Adv Atmos Sci 33:1–9. https://doi.org/10.1007/s00376-015-5056-0

Chooprateep S, McNeil N (2016) Surface air temperature changes from 1909 to 2008 in Southeast Asia assessed by factor analysis. Theor Appl Climatol 123:361–368

Coumou D, Capua GD, Vavrus S, Wang L, Wang S (2018) The influence of Arctic amplification on mid-latitude summer circulation. Nat Commun 9:2959

De Vries AJ, Tyrlis E, Edry D, Krichak S, Steil B, Lelieveld J (2013) Extreme precipitation events in the Middle East: dynamics of the active Red Sea trough. J Geophys Res Atmos 118:7087–7108. https://doi.org/10.1002/jgrd.50569

De Vries AJ, Ouwersloot HG, Feldstein S, Riemer M, Kenawy AE, McCabe M, Lelieveld J (2018) Identification of tropica-lextratropical interactions and extreme precipitation events in the Middle East based on potential vorticity and moisture transport. J Geophys Res Atmos 123:861–881. https://doi.org/10.1002/2017JD027587

DelSole T, Tippett MK (2014) Comparing forecast skill. Mon Weather Rev 142:4658–4678. https://doi.org/10.1175/MWR-D-14-00045.1

DelSole T, Nattala J, Tippett MK (2014) Skill improvement from increased ensemble size and model diversity. Geophys Res Lett 41:7331–7342. https://doi.org/10.1002/2014GL060133

DelSole T, Trenary L, Tippett MK, Pegion K (2017) Predictability of week-3–4 average temperature and precipitation over the contiguous United States. J Clim 30:3499–3512

Delworth et al (2006) GFDL's CM2 global coupled climate models. Part I: formulation and simulation characteristics. J Clim 19:643

Donat MG et al (2014) Changes in extreme temperature and precipitation in the Arab region: long-term trends and variability related to ENSO and NAO. Int J Climatol 34:581–592

Eade R, Smith D, Scaife A, Wallace E, Dunstone N, Hermanson L, Robinson N (2014) Do seasonal-to-decadal climate predictions underestimate the predictability of the real world? Geophys Res Lett 41:5620–5628. https://doi.org/10.1002/2014GL061146

Ehsan MA, Kang IS, Almazroui M, Abid MA, Kucharski F (2013) A quantitative assessment of changes in seasonal potential predictability for the twentieth century. Clim Dyn 41:2697–2709

Ehsan MA, Tippett MK, Almazroui M, Ismail M, Yousef A, Kucharski F, Omar M, Hussein M, Alkhalaf AK (2017a) Skill and predictability in multimodel ensemble forecasts for Northern Hemisphere regions with dominant winter precipitation. Clim Dyn 48:3309–3324

Ehsan MA, Almazroui M, Yousef A, Enda OB, Tippett MK, Kucharski F, Alkhalaf AK (2017b) Sensitivity of AGCM simulated regional summer precipitation to different convective parameterizations. Int J Climatol 37:4594–4609

Evan AT, Flamant C, Lavaysse C, Kocha C, Saci A (2015) Water vapor–forced greenhouse warming over the Sahara Desert and the recent recovery from the Sahelian drought. J Clim 28:108–123. https://doi.org/10.1175/JCLI-D-14-00039.1

Faller AJ (1981) An average correlation coefficient. J Appl Meteorol 20:203–205

Fan Y, Dool HVD (2008) A global monthly land surface air temperature analysis for 1948-present. J Geophys Res 113:D01103. https://doi.org/10.1029/2007JD008470

Gent et al (2011) The community climate system model version 4. J Clim 24:4973–4991

Hagedorn R, Doblas-Reyes FJ, Palmer TN (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting. Part I: basic concept. Tellus A 57:219–233

Hasanean HM, Almazroui M (2017) Teleconnections of the tropical sea surface temperatures to the surface air temperature over Saudi Arabia in summer season. Int J Climatol 37:1040–1049

Jia L, Yang X, Vecchi GA, Gudgel RG, Delworth TL, Rosati A, Stern WF, Wittenberg AT, Krishnamurthy L, Zhang S, Msadek R, Kapnick S, Underwood S, Zeng F, Anderson WG, Balaji V, Dixon K (2015) Improved seasonal prediction of temperature and precipitation over land in a high-resolution GFDL climate model. J Clim 28:2044–2062

Kanamitsu M et al (2002) NCEP-DEO AMIP-II reanalysis (R-2). Bull Am Meter Soc 83:1631–1643

Kang IS, Shukla J (2006) The Asian monsoon. In: Dynamic Seasonal Prediction and Predictability of the Monsoon. Springer, Chichester

Kang IS, Yoo JH (2006) Examination of multi-model ensemble seasonal prediction methods using a simple climate system. Clim Dyn 26:285–294. https://doi.org/10.1007/s00382-005-0074-8

Kirtman BP et al (2014) The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. Bull Am Meteorol Soc 95:585–601

Lelieveld J, Hadjinicolaou P, Kostopoulou E, Chenoweth J, Giannakopoulos C, Hannides C, Lange MA, El Maayar M, Tanarthe M, Tyrlis E, Xoplaki E (2012) Climate change and impacts in the eastern Mediterranean and the Middle East. Clim Change 114:667–687

Li JP, Sun C, Jin FF (2013) NAO implicated as a predictor of Northern Hemisphere mean temperature multidecadal variability. Geophys Res Lett 40:5497–5502

Miralles DG, Teuling AJ, Van Heerwaarden CC (2014) Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. Nat Geosci 7:345–349. https://doi.org/10.1038/ngeo2141

Mustafa A, Rahman G (2018) Assessing the spatio-temporal variability of meteorological drought in Jordan. Earth Syst Environ 2:247–264. https://doi.org/10.1007/s41748-018-0071-9

Osman M, Vera CS (2017) Climate predictability and prediction skill on seasonal time scales over South America from CHFP models. Clim Dyn 49:2365–2383

Pal JS, Eltahir EAB (2015) Future temperatures in southwest Asia projected to exceed a threshold for human adaptability. Nat Clim Change 6:197–200

Palmer TN et al (2004) Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). Bull Amer Meteor Soc 85:853–872

Rahman MA, Almazroui M, Islam MN, Brien EO, Yousef AE (2018) The role of land surface fluxes in Saudi-KAU AGCM: temperature climatology over the Arabian peninsula for the period 1981–2010. Atmos Res 200:139–152

Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. J Geophys Res 108:4407

Rodwell MJ, Hoskins BJ (1996) Monsoons and the dynamics of deserts. Q J Roy Meteorol Soc 122:1385–1404

Ropelewski CF, Halpert MS (1986) North American precipitation and temperature patterns associated with the El Nino-Southern Oscillation (ENSO). Mon Weather Rev 114:2352–2362

Ropelewski CF, Halpert MS (1987) Global and regional scale precipitation patterns associated with the El Nino/Southern Oscillation. Mon Weather Rev 115:1606–1626

Rowell DP (1998) Assessing potential seasonal predictability with an ensemble of multidecadal GCM simulations. J Clim 11:109–120

Rowell DP, Folland CK, Maskell K, Ward MN (1995) Variability of summer rainfall over tropical North Africa (190692): observations and modeling. Q J R Meteorol Soc 121:669–704

Saha S et al (2014) The NCEP climate forecast system version 2. J Clim 27:2185–2208

Sun C, Li JP, Ding RQ, Jin Z (2017) Cold season Africa-Asia multidecadal teleconnection pattern and its relation to the Atlantic multidecadal variability. Clim Dyn 48:3903–3918. https://doi.org/10.1007/s00382-016-3309-y

Tang Y, Chen D, Yan X (2014a) Potential predictability of Northern America surface temperature part I: information-based vs signal-to-noise based metrics. J Clim 27:1578–1599

Tang Y, Chen D, Yan X (2014b) Potential predictability of Northern America surface temperature in AGCMs and CGCMs. Clim Dyn 45:353–374

Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res 106:7183–7192

Tippett MK, Almazroui M, Kang IS (2015) Extended-range forecasts of areal-averaged Saudi Arabia rainfall. Weather Forecast 30:1090–1105. https://doi.org/10.1175/WAF-D-15-0011.1

Vecchi GA et al (2014) On the seasonal forecasting of regional tropical cyclone activity. J Clim 27:7994–8016

Vernieres G, Keppenne C, Rienecker MM, Jacob J, Kovach R (2012) The GEOS-ODAS, description and evaluation. NASA Technical report series on global modeling and data assimilation, NASA/TM 30:104606. National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland

Vigaud N, Tippett MK, Robertson AW (2018) Probabilistic skill of subseasonal precipitation forecasts for the East Africa-west Asia sector during September to May. Weather Forecast 33:1513–1532. https://doi.org/10.1175/WAF-D-18-0074.1

Wilks DS (2006) Statistical methods in the atmospheric sciences (Chapter no. 5), 2nd edn. Elsevier, New York