

# *Forecast skill of the Indian monsoon and its onset in the ECMWF seasonal forecasting system 5 (SEAS5)*

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Chevuturi, A. ORCID: <https://orcid.org/0000-0003-2815-7221>,  
Turner, A. ORCID: <https://orcid.org/0000-0002-0642-6876>,  
Johnson, S., Weisheimer, A., Shonk, J., Stockdale, T. N. and  
Senan, R. (2021) Forecast skill of the Indian monsoon and its  
onset in the ECMWF seasonal forecasting system 5 (SEAS5).  
Climate Dynamics, 56 (9-10). pp. 2941-2957. ISSN 0930-7575  
doi: 10.1007/s00382-020-05624-5 Available at  
<https://centaur.reading.ac.uk/95213/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1007/s00382-020-05624-5>

Publisher: Springer

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

## **CentAUR**

Central Archive at the University of Reading

Reading's research outputs online



# Forecast skill of the Indian monsoon and its onset in the ECMWF seasonal forecasting system 5 (SEAS5)

Amulya Chevuturi<sup>1,2</sup> · Andrew G. Turner<sup>1,2</sup> · Stephanie Johnson<sup>3</sup> · Antje Weisheimer<sup>3,4</sup> · Jonathan K. P. Shonk<sup>1,2,5</sup> · Timothy N. Stockdale<sup>3</sup> · Retish Senan<sup>3</sup>

Received: 4 May 2020 / Accepted: 31 December 2020  
© Crown 2021

## Abstract

Accurate forecasting of variations in Indian monsoon precipitation and progression on seasonal time scales remains a challenge for prediction centres. We examine prediction skill for the seasonal-mean Indian summer monsoon and its onset in the European Centre for Medium-Range Weather Forecasts (ECMWF) seasonal forecasting system 5 (SEAS5). We analyse summer hindcasts initialised on 1st of May, with 51 ensemble members, for the 36-year period of 1981–2016. We evaluate the hindcasts against the Global Precipitation Climatology Project (GPCP) precipitation observations and the ECMWF reanalysis 5 (ERA5). The model has significant skill at forecasting dynamical features of the large-scale monsoon and local-scale monsoon onset tercile category one month in advance. SEAS5 shows higher skill for monsoon features calculated using large-scale indices compared to those at smaller scales. Our results also highlight possible model deficiencies in forecasting the all India monsoon rainfall.

**Keywords** Indian summer monsoon · Monsoon onset · ECMWF · SEAS5 · Forecast skill

## 1 Introduction

The Indian summer monsoon (ISM) presents an interesting challenge for seasonal prediction. ISM precipitation varies across different spatio-temporal scales—including the variability of seasonal mean rainfall and ISM onset and withdrawal dates over the Indian subcontinent over seasonal and interannual timescales. These variations have large impacts on major water resources, ecosystems, agriculture and thus the Indian population. Improved forecasting of the ISM progression and seasonal rainfall can help alleviate water stress

for agriculture and domestic needs and mitigate the impacts of hydrometeorological disasters.

Progress has been made by state-of-the-art models in simulating ISM climatology (e.g., Menon et al. 2018), its interannual variability (e.g., Wang et al. 2015a), and variations in the monsoon on subseasonal-to-interannual timescales, such as for extreme precipitation (e.g., Nanjundiah et al. 2013), large-scale circulation patterns (e.g., Saha et al. 2014), intraseasonal variability (e.g., Abhilash et al. 2014), influence of teleconnections on the ISM (e.g., Preethi et al. 2010) and the monsoon onset (e.g., Chevuturi et al. 2019). Further, monsoon prediction is sensitive to features such as atmosphere-ocean coupling (Krishna Kumar et al. 2005), model resolution (Ramu et al. 2016) and atmospheric initialization (Alessandri et al. 2015) etc. However, despite considerable progress in the prediction of dynamical features of ISM variability, models still struggle to simulate the local-scale intraseasonal variability (e.g., Sperber et al. 2001; Johnson et al. 2017; Chevuturi et al. 2019).

Similar to its contemporary coupled ensemble prediction systems, the European Centre for Medium Range Weather Forecasts (ECMWF) seasonal forecasting system 4 (SEAS4) represents the mean Asian monsoon circulation well despite systematic errors associated with monsoon precipitation

✉ Amulya Chevuturi  
a.chevuturi@reading.ac.uk

<sup>1</sup> National Centre for Atmospheric Science, University of Reading, Reading, UK

<sup>2</sup> Department of Meteorology, University of Reading, Reading, UK

<sup>3</sup> European Centre for Medium-Range Weather Forecasts, Reading, UK

<sup>4</sup> National Centre for Atmospheric Science, University of Oxford, Oxford, UK

<sup>5</sup> Present Address: Met Office @Reading, University of Reading, Reading, UK

(Kim et al. 2012). It has good skill at predicting ISM inter-annual variability (Jie et al. 2017; Pandey et al. 2015), and shows relatively low model bias for ISM rainfall compared to other models (Jain et al. 2019). The new ECMWF seasonal forecasting system 5 (SEAS5) has many improvements over SEAS4, including reduction in cold SST errors over equatorial Pacific and Indian Ocean regions (Johnson et al. 2019). These improvements have led to significantly higher prediction skill for large-scale ISM rainfall interannual variability in SEAS5 compared to its predecessor, and also put SEAS5 as a front runner in predicting extreme ISM rainfall (Köhn-Reich and Bürger 2019). However, SEAS5 like SEAS4, still shows stronger than observed teleconnections with the tropical Pacific.

It remains unclear how well SEAS5 represents the local-scale ISM and its onset, and what errors remain in its monsoon dynamics. Studies have shown that model deficiencies at seasonal timescales have more serious consequences for forecast skill compared to those arising from errors in the initial conditions (Pokhrel et al. 2016). Thus, an assessment of ISM simulation and its prediction skill in the SEAS5 system will be useful to user communities (scientific and operational), by providing information on the strengths and limitations of the model. In this study, we aim to examine the representation of the ISM and its progression in the ECMWF SEAS5 forecasting system with a focus on prediction of the monsoon onset. We quantify the prediction skill of seasonal mean ISM and monsoon onset variability in the SEAS5 hindcasts, using various objective indices.

The article is organized as follows: the SEAS5 hindcast and verification data used in the study are introduced in Sect. 2.1; methods are provided in Sect. 2.2; results for the prediction skill of seasonal mean ISM are discussed in Sect. 3 and for the monsoon onset in Sect. 4; and Sect. 5 provides the summary and concludes this study.

## 2 Data and methods

### 2.1 Data

The SEAS5 coupled model features the Integrated Forecast System (IFS; cycle 43r1) atmospheric model coupled to the HTESSEL land-surface model and the Nucleus for European Modelling of the Ocean (NEMO; version 3.4.1) ocean model. SEAS5 seasonal forecasts have O320 ( $\approx 36$  km) horizontal resolution and 91 vertical levels for the atmosphere, and ORCA 0.25 ( $\approx 27$  km) horizontal resolution and 75 vertical levels for the ocean. The SEAS5 forecasts are integrated for 7 months with a 51-member ensemble initialised on the first of every month. Johnson et al. (2019) and ECMWF (2017) give a detailed description of the SEAS5 forecasting system.

We use 36 years (1981–2016) of the retrospective seasonal forecasts (also referred to as reforecasts or hindcasts) to estimate the forecast skill of the system. For our study, the focus is on analysing the 1st of May initialised hindcast set. Studies have shown that dynamical coupled models can skilfully predict ISM rainfall initialized in May (e.g., Wang et al. 2015b; DelSole and Shukla 2010). We use 124 days (01 May–01 Sep) of SEAS5 seasonal hindcasts at  $1^\circ \times 1^\circ$  horizontal resolution, with 51 ensemble members for each year.

ECMWF's new reanalysis (ERA5; Hersbach et al. 2020) at  $0.25^\circ \times 0.25^\circ$  horizontal resolution, has been used for comparison with model output for dynamic and thermodynamic fields. For rainfall observations we used the GPCP (Global Precipitation Climatology Project) dataset at  $2.5^\circ \times 2.5^\circ$  horizontal resolution, with monthly version 2.3 (Adler et al. 2018) and pentad version 2.2 (Xie et al. 2003). We use GPCP due to its better representation of the Indian monsoon in comparison to other merged rainfall datasets (Prakash et al. 2015).

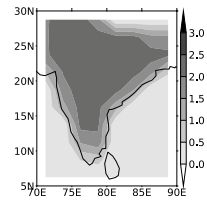
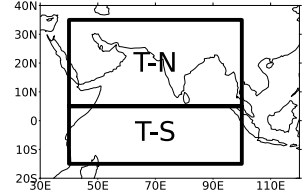
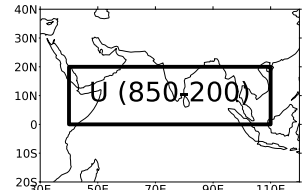
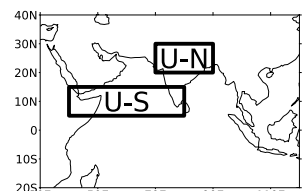
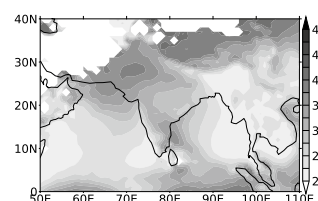
### 2.2 Methods

#### 2.2.1 Objective indices

To evaluate prediction skill we use objective indices that reflect the key physical mechanisms associated with the Indian monsoon and define the monsoon onset date based on these mechanisms. The different indices used and their respective domains are summarized in Table 1.

- **AIRI** (All India Rainfall Index) is defined as the weighted average of JJA rainfall anomalies over the Indian region and is an index used by ECMWF. The weights are scaled by the fraction of low-altitude land and normalized to a unit-area average at the GPCP native resolution. The region covered and the weights at each grid-point are shown in Table 1. This mask allows us to only calculate rainfall over India and ignore the surrounding regions.
- **TTGI** (Tropospheric Temperature Gradient Index) is defined as the difference between vertically integrated tropospheric temperature over 600–200 hPa between northern and southern regions of the South Asian domain. Pre-monsoon warming over the Asian region along with the elevated heat pump due to the Tibetan Plateau establishes a meridional temperature gradient and forms a heat low over South Asia, which is a precursor to the monsoon onset. The TTGI monsoon index is a seasonal average of the TTGI time series for the months of JJA. The TTGI monsoon onset date is specified when the northern box (depicted in the TTGI panel of Table 1) becomes warmer than the southern box (i.e. date on which the TTGI becomes positive). As our results were

**Table 1** Details of seasonal mean and onset indices used in the study

Index	Methodology	Onset	Region
All India Rainfall Index (AIRI; ECMWF 2017)	Precipitation weighted average (scaled as per the figure) Area = 70°–90° E, 5°–30° N	–	
Tropospheric Temperature Gradient Index (TGI; Xavier et al. 2007)	$TTGI = T_N - T_S$ T = Vertically averaged temperature (600–200 hPa) $T_N = T(40^{\circ}\text{--}100^{\circ}\text{ E}, 5^{\circ}\text{--}35^{\circ}\text{ N})$ $T_S = T(40^{\circ}\text{--}100^{\circ}\text{ E}, 15^{\circ}\text{ S--}5^{\circ}\text{ N})$	Date when TTGI time series for each year becomes positive	
Webster and Yang Index (WYI; Webster and Yang 1992)	$WYI = U_{850} - U_{200}$ U = Zonal Wind at 850 or 200 hPa $U = U(40^{\circ}\text{--}110^{\circ}\text{ E}, 0^{\circ}\text{--}20^{\circ}\text{ N})$	Date when seven-day moving-average of WYI time series exceeds climatological WYI value of ERA5 on 2nd June	
Wang and Fan Index (WFI; Wang and Fan 1999)	$WFI = U_S - U_N$ U = Zonal Wind at 850 hPa $U_S = U(40^{\circ}\text{--}80^{\circ}\text{ E}, 5^{\circ}\text{--}15^{\circ}\text{ N})$ $U_N = U(70^{\circ}\text{--}90^{\circ}\text{ E}, 20^{\circ}\text{--}30^{\circ}\text{ N})$	Date when seven-day moving-average of WFI time series exceeds climatological WFI value of ERA5 on 2nd June	
Wang and LinHo Index (WLI; Wang and LinHo 2002)	$WLI = P - R_{jan}$ P = Pentad rainfall for different years $R_{jan}$ = Mean rainfall for January of respective years	Pentad when five-pentad moving-average of WLI time series for each grid-point for each year crosses 5 mm day <sup>-1</sup>	

Climatological monsoon onset date over Kerala from 1981–2016 is 2nd June

- found not to be sensitive to any smoothing of the TTGI, no smoothing is applied here.
- **WYI** (Webster and Yang Index) describes the strengthening of the vertical wind shear (lower tropospheric westerlies and upper tropospheric easterlies) during the monsoon period over South Asia. WYI is calculated as the JJA seasonal average of the contrast between zonal winds at lower levels (850 hPa) and in the upper troposphere (200 hPa) over the large-scale South Asian region demarcated by the box in Table 1. The onset date calculation method is the same for both circulation indices (WYI and WFI) and is described below.
  - **WFI** (Wang and Fan Index) describes the strengthening of the zonal horizontal wind shear in the lower troposphere (tropical monsoon westerlies over India and the easterly flow over the north of India). WFI is calculated as the JJA seasonal average of the contrast between

lower-level zonal winds in the southern and northern boxes (Table 1). For onset date calculations for the circulation indices (WYI and WFI), we smooth the index time series using a seven-day moving-average to prevent assigning bogus onsets. For calculating onset with the circulation indices, we use the climatological value of ERA5 on 2nd June as our threshold for the respective indices. 2nd June is the mean climatological monsoon onset date over Kerala from 1981–2016, as listed by the India Meteorological Department (IMD).

- **WLI** (Wang and LinHo Index) defines monsoon onset at each grid-point using rainfall, providing spatial variability at the local-scale. To avoid noise inherent in time series of grid-point rainfall, the WLI is calculated using a smoothed pentad rainfall time series and calculated relative to the minimum in the annual cycle (January mean rainfall). Then the timing of the monsoon onset pentad at

each grid-point is determined when the five-pentad running average of relative rainfall exceeds  $5 \text{ mm day}^{-1}$ . The spatial variability of climatological (1981–2016) WLI monsoon onset pentads for GPCP is shown in Table 1.

### 2.2.2 Verification methods

In this study, we analyse the deterministic skill of the SEAS5 ensemble mean, as well as the probabilistic skill of the SEAS5 ensemble forecasts. Comparison of interannual variability between the hindcast ensemble mean and observations uses the Pearson correlation coefficient (CC). We use a one-sided Student's *t*-test to examine the existence of skill ( $CC > 0$ ), and test for significant differences at the 5% level. To quantify the relationship between real-world skill and potential predictability of the ensemble forecast system, we use the 'ratio of predictable components' (RPC; Eade et al. 2014). RPC is calculated as the CC (actual skill) divided by the ratio between standard deviation of the model ensemble mean and the average standard deviation of individual members (potential predictability). A forecasting system having RPC equal to 1 indicates that the predictable component of the real world is the same as in the model world.

$$RPC = \frac{CC}{\sqrt{\sigma_{sig}^2 / \sigma_{tot}^2}} \quad (1)$$

where CC is the correlation coefficient,  $\sigma_{sig}^2$  is the signal variance of the model ensemble mean and  $\sigma_{tot}^2$  is the variance of all ensemble members. The actual predictability of any forecast system is usually different to its potential predictability, and RPC represents this difference. RPC greater (lesser) than 1 denotes underconfident (overconfident) forecasts.

Reliability of the ensemble forecast system is measured based on the relationship between the intraensemble spread and the error of the ensemble mean forecast, as the 'spread-error ratio' (SER; Ho et al. 2013). For a large and perfect ensemble, the RMSE (root mean square error) of the ensemble mean would be equal to the ensemble spread about the ensemble mean (Weisheimer et al. 2011). An ensemble system with SER larger (smaller) than one is considered overdispersed (underdispersed), and the probabilistic forecasts are expected to be unreliable.

$$SER = \sqrt{\frac{m+1}{m}} \frac{\sqrt{\sigma_{tot}^2}}{RMSE} \quad (2)$$

where *m* is the number of ensemble members, RMSE is the root mean square error of the ensemble mean and  $\sigma_{tot}^2$  is the variance of all ensemble members. To estimate the

sampling uncertainty of RPC and SER, we use the bootstrapping approach to generate a distribution of RPC and SER values, by randomly generating 1000 samples, from the ensemble and the hindcast years, with replacement. From the randomly generated distribution, using a two-sided test we estimate RPC and SER values which are statistically indistinguishable from 1 at the 95% confidence interval.

The skill of SEAS5 at forecasting the monsoon onset is also quantified in terms of tercile categories: (a) early, (b) normal and (c) late onset. Model skill for onset tercile categories is estimated in terms of deterministic (Accuracy; ACC and Heidke skill score; HSS) and probabilistic forecasts (Brier skill score; BSS and Ranked probability skill score; RPSS) (WCRP 2015). ACC is a score that defines the accuracy of the model performance, whereas HSS is the accuracy of forecasts at predicting the observed category, relative to that of random chance. ACC is the ratio of how many times the model forecasts the correct onset category. Negative values for HSS indicate that the model forecast is worse than a randomly generated forecast set. BSS is calculated separately for all three onset categories and measures the mean-squared forecast probability error. RPSS measures the sum of squared probability errors, which is cumulative across the three forecast categories, in order from early to normal to late onsets. Negative values for BSS and RPSS indicate a forecast which is worse than a climatological forecast (with a probability of 1/3 in each onset category). Please refer to Table 2 for detailed description of the verification skill scores used to quantify skill of an ensemble forecast system.

## 3 Monsoon dynamics and thermodynamics

In this section, we examine the model performance in terms of the dynamics and thermodynamics of the ISM in order to understand the monsoon rainfall errors. We will first assess the climatological mean state of the model, to diagnose systematic biases in the model. Next we will analyse the interannual prediction skill for the different monsoon features. For an objective analysis of monsoon characteristics, we use seasonal (JJA) mean indices: AIRI, TTGI, WYI and WFI, as described in Table 1.

### 3.1 Climatological mean and interannual variability

SEAS5 generally simulates the pattern of ISM features well (Fig. 1). It clearly shows the enhanced meridional gradient in tropospheric temperature (Li and Wang 2016), which establishes the lower level cross-equatorial flow (Findlater 1969) and consequently the upper level tropical easterly jet (Koteswaram 1958). SEAS5 shows a warmer than observed temperature gradient (Fig. 1j) which strengthens



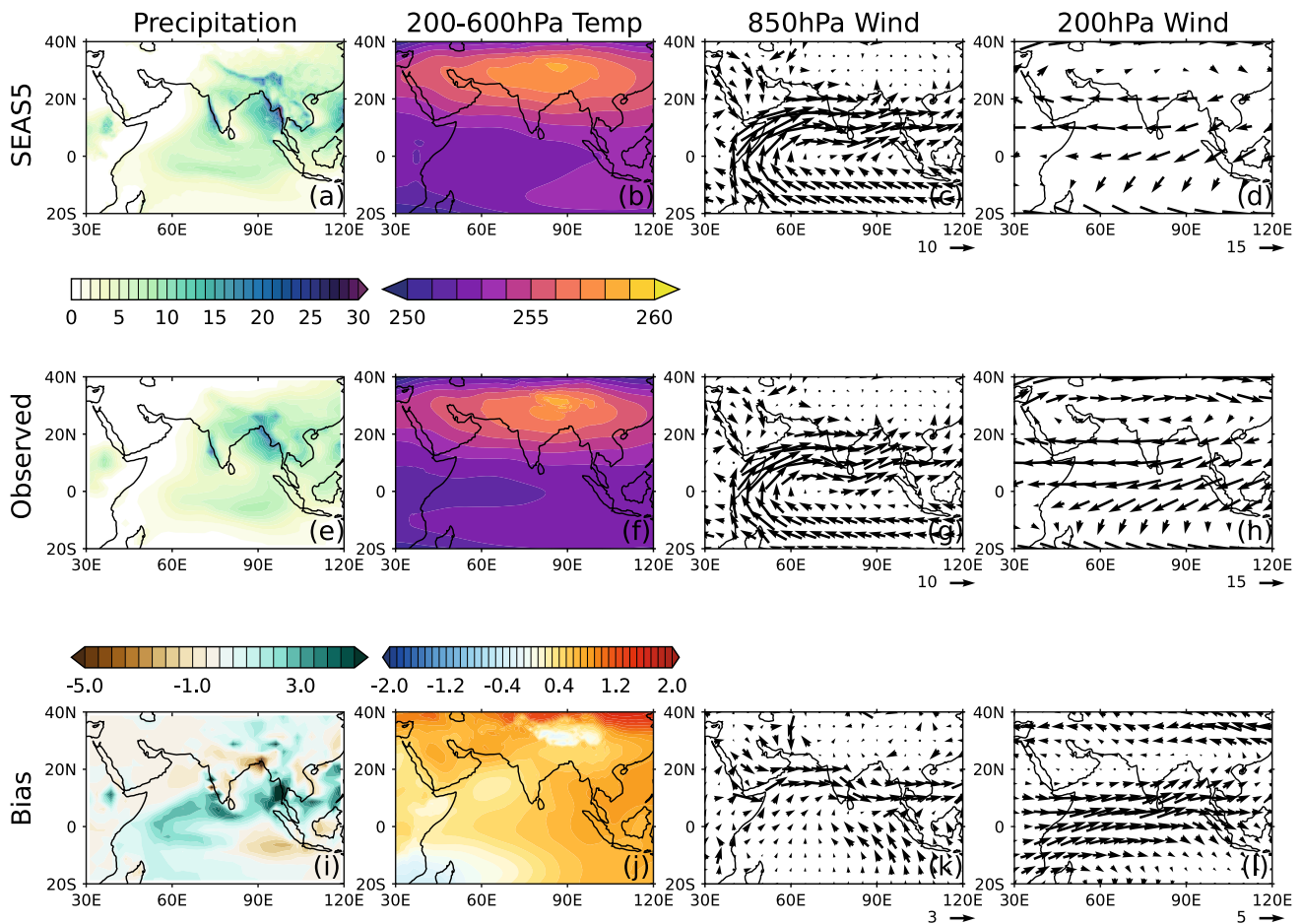
the monsoon flow, as indicated by the anomalous westerly bias at 850 hPa (Fig. 1k). An upper level easterly wind bias north of India and a westerly bias to the south (Fig. 1l) are due to a slight northward shift in the positions of the tropical easterly jet and the sub-tropical westerly jet in SEAS5 (not shown).

SEAS5 represents the climatological pattern of the seasonal mean monsoon precipitation well: i.e. enhanced precipitation over the Western Ghats, monsoon core region, Gangetic Delta and the Himalayas (Fig. 1a). SEAS5 overestimates seasonal mean precipitation over the Western Ghats and Himalayas and underestimates it over the Gangetic Plains and Delta (Fig. 1i), as with other seasonal forecast models, such as the North American Multi-Model Ensemble (NMME) seasonal systems (Singh et al. 2019). The underestimation of ISM mean rainfall is not a unique problem for SEAS5 alone; other contemporary models also show a drier ISM with the same forecast lead times (Jain et al. 2019).

The verification of SEAS5 seasonal mean (JJA) monsoon indices is summarized in Table 3. SEAS5 seasonal mean monsoon indices differ significantly from those in observations/reanalyses, due to biases discussed in the above paragraph. Tropospheric temperature for SEAS5 is generally warmer than in reanalysis and the local-scale rainfall biases lead to a seasonal mean dry bias in AIRI. Due to the westerly wind biases, at lower and upper levels, SEAS5 underestimates the zonal vertical wind shear (WYI) and overestimates the zonal horizontal wind shear (WFI). For almost all the monsoon indices, the interannual spread in SEAS5 is usually larger than observed, except for WYI (Table 3). For the detailed interannual ensemble member spread of monsoon indices in SEAS5 compared to observations, please refer to Fig. 2. Compared to the ensemble spread of AIRI in SEAS5, ensemble spread is higher for WYI and lower for TTGI. SEAS5 ensemble spread generally encompasses the seasonal mean observations except in years such as 1994 for TTGI, WYI and WFI, and 2002 for AIRI and TTGI. Further, as all

**Table 2** Description and calculation of the verification skill scores used in the study

Verification Skill Score	Calculation
<i>Accuracy (ACC)</i> quantifies the fraction of forecasts predicting the correct tercile category amongst all forecasts and ranges from 0 (no skill) to 1 (perfect score)	$ACC = \frac{1}{N} \sum_{i=1}^C n(F_i, O_i) \quad (1)$ <p>where C is the forecast category (early/normal/late), N is the total number of forecasts (years) and <math>n(F_i, O_i)</math> is the number of accurate forecasts for all different categories for each year</p>
<i>Heidke Skill Score (HSS)</i> represents the accurate forecasts after eliminating those which are correct due to random chance. This score ranges from $-\infty$ to 1, with 0 meaning no skill and 1 meaning a perfect forecast score. Negative values for HSS indicate that the model forecast is worse than a randomly generated forecast set	$HSS = \frac{\frac{1}{N} \sum_{i=1}^C n(F_i, O_i) - \frac{1}{N^2} \sum_{i=1}^C n(F_i) n(O_i)}{1 - \frac{1}{N^2} \sum_{i=1}^C n(F_i) n(O_i)} \quad (2)$ <p>where C is the number of forecast categories, N is the total number of forecasts (years), <math>n(F_i, O_i)</math> represents the accurate forecasts and <math>n(F_i) n(O_i)</math> is all combinations of expected forecast and observed category combinations</p>
<i>Ranked Probability Skill Score (RPSS)</i> measures how well the multi-category probabilistic forecast predicts the actual observed category in cumulative sense. This score ranges from $-\infty$ (highest possible error) to 1 (perfect score), with 0 indicating no skill when compared to reference climatology	$RPSS = 1 - \frac{RPS}{RPS_{clim}} \quad (3)$ $RPS = \frac{1}{N} \sum_{i=1}^N \left[ \frac{1}{(C-1)} \left( \sum_{j=1}^C (F_j - O_j)^2 \right) \right] \quad (4)$ <p>where RPS is the Ranked Probability Score and <math>RPS_{clim}</math> is the reference RPS climatology, calculated with the same formula as RPS, but with climatological probability of 1/3 for the value of F in all the cases. F is the forecast probability, O is the observed category, N is the number of forecasts and C is the number of forecast categories: (1) early, (2) normal, (3) late; quantified as cumulative categorical forecast probability in the given order</p>
<i>Brier Skill Score (BSS)</i> defines the skill of the probabilistic forecast for a category and is calculated separately for each category and reflects the mean-squared probability error. This score ranges from $-\infty$ (highest possible error) to 1 (perfect score), with 0 indicating no skill when compared to the reference climatology	$BSS = 1 - \frac{BS_c}{BS_{clim}} \quad (5)$ $BS_c = \frac{1}{N} \sum_{i=1}^N (F_i - O_i)^2 \quad (6)$ <p>where <math>BS_{clim}</math> is the reference Brier Score climatology, calculated with the same formula as <math>BS_c</math> but with climatological probability of 1/3 for the value of F in all the cases and <math>BS_c</math> is the Brier Score of a particular category. F is the forecast probability of that category, O is the observed category and <math>N_c</math> is the number of forecasts (years) in the same category. O is 1 for the observed category and 0 for other categories</p>



**Fig. 1** Seasonal (JJA) mean climatology for **a** precipitation ( $\text{mm day}^{-1}$ ), **b** tropospheric temperature (K; averaged over 600–200 hPa), **c** lower-level winds ( $\text{m s}^{-1}$ ; 850 hPa) and **d** upper-level winds ( $\text{m s}^{-1}$ ; 200 hPa) for SEAS5 at native resolution. **e–h** show the same as **a–d** but for verification datasets (GPCP and ERA5) at their respective

horizontal resolution. **i–l** is the bias between the model and observed calculated at coarser resolution. **i** Precipitation bias is shown at GPCP resolution, and **j–l** biases of the other three variables are shown at SEAS5 resolution

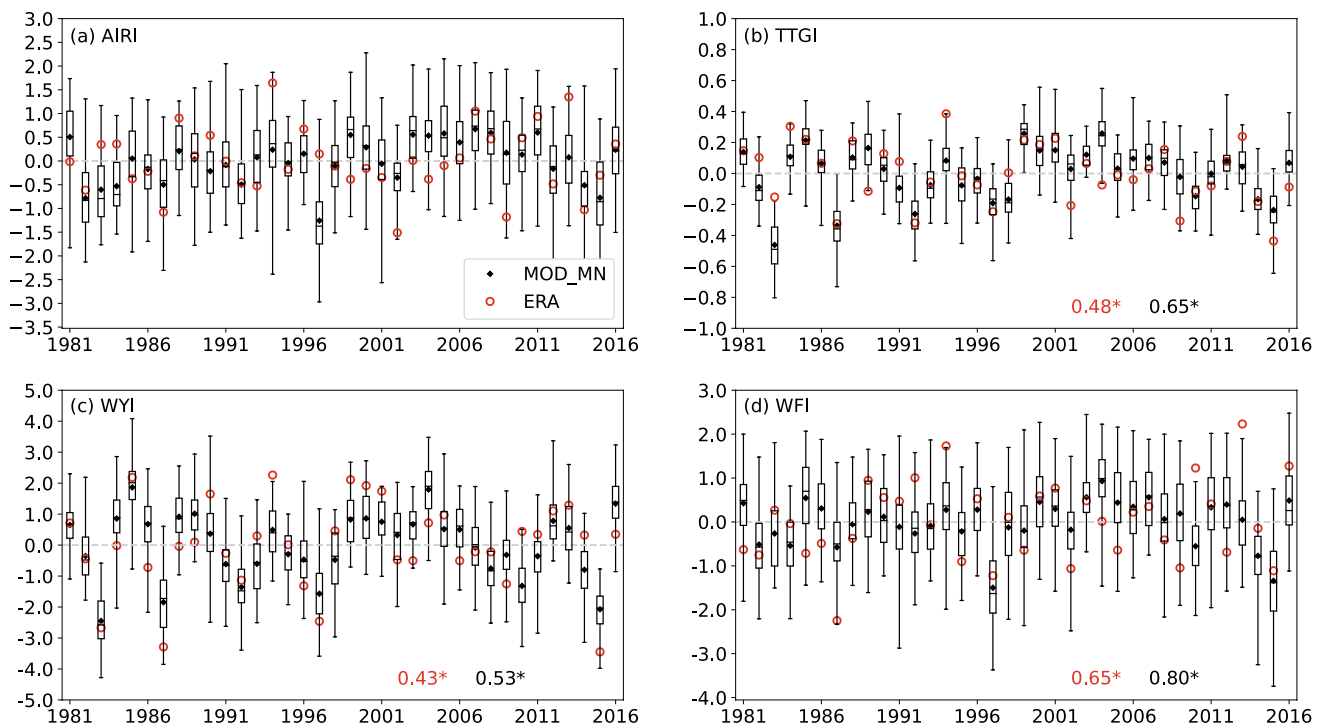
of these indices are indicators of Indian monsoon strength, they are known to not be entirely independent (Moron and Robertson 2014). Our results indicate that these indices show similar patterns of interannual variability and there are significant correlations between the monsoon rainfall index (AIRI) and the three other indices, for both SEAS5 and observations (Fig. 2b–d).

Analysing the monthly mean AIRI (Fig. 3), shows that SEAS5 also represents the monsoon (May to September) seasonal cycle well. Monthly AIRI strongly increases from May to July, as the monsoon peaks, and then shows a gradual reduction from July to September. SEAS5 has a wet bias in the months before the monsoon peak (May–June) and a dry bias from July to September. Thus, the seasonal mean dry bias for AIRI in SEAS5 stems from insufficient rainfall from July onwards. The interannual spread of the monthly mean AIRI is larger in SEAS5 than observed, similar to the interannual spread of the seasonal mean AIRI (Table 3).

### 3.2 Skill of interannual prediction

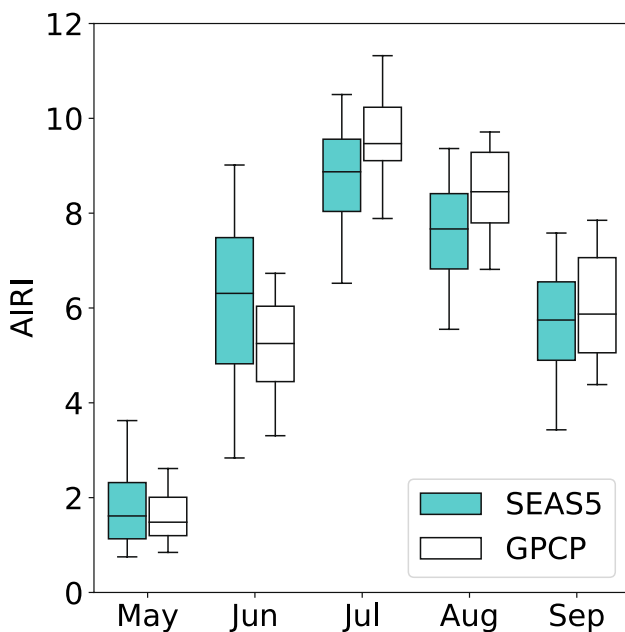
To evaluate how well SEAS5 hindcasts represent predictable modes of ISM variability, we compare the principal component (PC) time series for the first two modes (PC1 and PC2) of ISM variability between SEAS5 and GPCP (Fig. 4). For the PC time series calculation we use the first two leading modes of empirical orthogonal function (EOF) calculated for GPCP monthly-mean summer rainfall anomalies over the South Asian domain ( $15^{\circ}\text{S}$ – $30^{\circ}\text{N}$ ,  $60^{\circ}\text{E}$ – $120^{\circ}\text{E}$ ), as EOF1 and EOF2 (ECMWF 2017). The first EOF (EOF1; Fig. 4a) resembles precipitation patterns associated with summer La Niña events, with enhanced precipitation over the eastern equatorial Indian Ocean and southeast Asia and reduced precipitation over the northern Bay of Bengal and East Asia. The second EOF (EOF2; Fig. 4c) pattern has enhanced precipitation over the Indian subcontinent, surrounding oceans and western equatorial Indian Ocean; which resembles the





**Fig. 2** Seasonal (JJA) mean monsoon indices with symbols showing monsoon index for SEAS5 ensemble mean (black rhombi) and ERA5 (red circles). The boxplots show the ensemble spread of seasonal mean monsoon index for SEAS5. For each boxplot, whiskers show the range (max–min) of the onset dates, middle dash is the median and box ends show inter-quartile range of the ensemble spread. Sea-

sonal mean monsoon indices for each sub-figure are calculated with **a** AIRI, **b** TTGI, **c** WYI and **d** WFI. The grey horizontal line demarcates the 0 threshold. The numerical values shown in panels **b–d** are CC between AIRI against each respective index for observed (red) and SEAS5 (black) and asterisk represents statistical significance at the 5% level



**Fig. 3** Monthly interannual spread of AIRI for SEAS5 (blue) and GPCP (white) in boxplots. For each boxplot, whiskers show the range (max–min) of the AIRI, middle dash is the median and box ends show the inter-quartile range

rainfall anomalies observed during the positive phase of the Indian Ocean Dipole. PC1 (Fig. 4b) and PC2 (Fig. 4d) time series are generated by spatially regressing seasonal mean rainfall anomalies each year (SEAS5 and GPCP) onto the EOF1 and EOF2 patterns respectively. Comparison of the SEAS5 PC time series against observations shows moderate but significant correlations between the models and observed for the first two modes of interannual monsoon variability (Fig. 4b, d).

When we look at skill in ISM interannual variability, in terms of CC, RPC and SER (Table 3), the results suggest that large-scale monsoon features measured by temperature gradient (TTGI) and vertical wind-shear (WYI) are better represented in SEAS5 than smaller-scale features (AIRI and WFI), as large-scale seasonal mean monsoon indices in SEAS5 are significantly correlated with the reanalysis. Other models also show moderate skill for ISM rainfall (Rajeevan et al. 2012) but have better skill for large-scale monsoon circulation (Johnson et al. 2017). Further, for smaller-scale monsoon features (AIRI and WFI), SEAS5 RPC is significantly lower than 1 (Eade et al. 2014), which indicates overconfident forecasts, where the ensemble mean resembles the ensemble members more than the observations. RPC for large-scale monsoon features (TTGI and WYI) is

**Table 3** Seasonal (JJA) mean monsoon indices and onset dates are shown with their interannual standard deviation (SD) for SEAS5 and GPCP/ERA5 along with CC, RPC and SER, using their respective indices

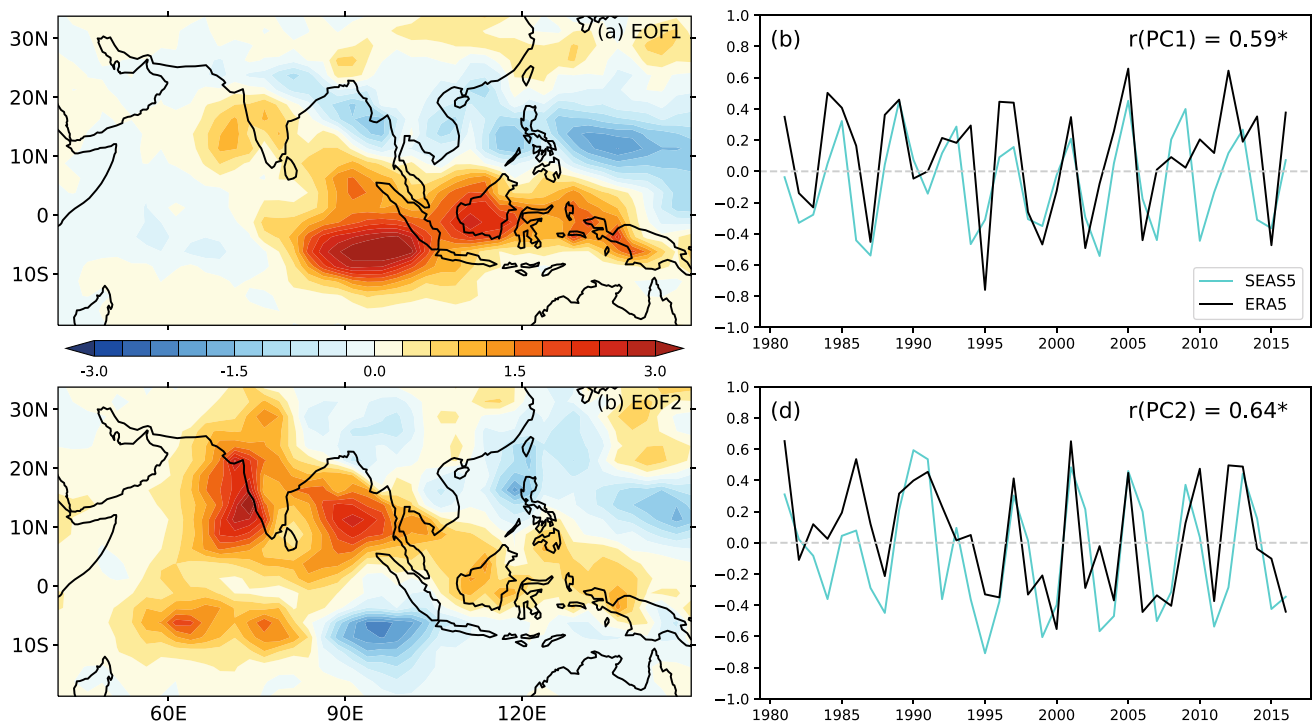
SEASONAL MEAN				
	AIRI	TTGI	WYI	WFI
MOD MEAN	7.48 <sup>+</sup>	2.63 <sup>+</sup>	23.55 <sup>+</sup>	9.64 <sup>+</sup>
OBS MEAN	7.74	2.55	26.16	8.72
MOD SD	0.84 <sup>+</sup>	0.21 <sup>+</sup>	1.35 <sup>+</sup>	0.94 <sup>+</sup>
OBS SD	0.68	0.19	1.41	0.91
CC	0.33*	0.60*	0.73*	0.35*
RPC	0.60	0.81'	0.95'	0.62
SER	0.84	1.23'	1.09'	0.75
ONSET DATES				
	TTGI	WYI	WFI	
MOD MEAN	24-May	29-May	22-May <sup>+</sup>	
OBS MEAN	26-May	28-May	28-May	
MOD SD	6.55	6.75	8.46 <sup>+</sup>	
OBS SD	7.40	8.81	9.66	
CC	0.74*	0.66*	0.57*	
RPC	1.01'	0.95'	0.93'	
SER	1.26'	1.00'	0.93'	

Model mean and standard deviation values marked with a plus sign are statistically different than observed. CC values marked with an asterisk are statistically significant at the 5% level. RPC and SER values marked with an apostrophe are statistically indistinguishable from 1 at the 95% confidence interval (more details in Sect. 3.2)

statistically indistinguishable from 1, which suggests that the forecast skill of SEAS5 is close to the potential predictability limit estimated by the model ensemble spread. Similar to RPC, SEAS5 has SER statistically indistinguishable from 1 for large-scale indices (TTGI, WYI), and SER lower than 1 for smaller-scale indices. This suggests that SEAS5 has reliable forecasts for large-scale monsoon features and the forecasts associated with smaller-scale monsoon features are underdispersed, which leads to unreliable probabilistic forecasts for the smaller-scale indices.

The model's interannual predictive skill at each grid-point for seasonal (JJA) mean rainfall, tropospheric temperature, vertical zonal wind shear and rainfall is shown in Fig. 5. SEAS5 skill, in terms of correlation, at each grid-point, indicates that the thermodynamic and dynamic features such as the vertical wind shear of zonal winds (Fig. 5c) and tropospheric temperature (Fig. 5b) are generally better represented in SEAS5 than the precipitation anomaly (Fig. 5a), which is a common feature amongst all current models (Kim et al. 2012). SEAS5 only has significant correlations with observed precipitation anomalies (p-value less than 0.05; Fig. 5a or RPC statistically indistinguishable from 1; Fig. 5d) over parts of southern,

eastern and western India, with only parts of the northern India having SER statistically indistinguishable from 1. For zonal vertical wind shear, SEAS5 over most parts of India has good prediction skill (Fig. 5b). Its actual skill matches well with potential predictability (Fig. 5e), and has reliable forecasts (Fig. 5h), over the whole Indian sub-continent and surrounding seas. Fig. 5c shows that SEAS5 skill for tropospheric temperature is high over the whole of the tropical Indian Ocean region (15°S to 15°N), and SEAS5 has reliable forecasts for tropospheric temperature over southern India (Fig. 5i). SEAS5 generally produces underdispersed and overconfident forecasts (SER < 1), for all the three variables, for most parts of the study domain (Fig. 5g–i), which is a common problem among other seasonal forecasting systems (Weisheimer et al. 2011). Nonetheless, there are some regions where SER > 1, i.e. parts of central and western Indian subcontinent for precipitation, southern India for zonal vertical wind shear, and the western equatorial Indian Ocean and parts of the Tibetan Plateau for tropospheric temperature. These regions have unreliable (underconfident) probabilistic forecasts with SEAS5, due to a larger ensemble spread than the ensemble mean error.



**Fig. 4** Modes of June–September GPCP ISM precipitation variability as **a** EOF1 and **c** EOF2 patterns over the Indian region; and principal component (PC) time series associated with the **b** first (PC1) and **d** second (PC2) leading modes of EOF of ISM precipitation

## 4 Monsoon onset and progression

In this section, we examine the model performance for the ISM onset and its progression. We will first assess the climatological mean and interannual variability of modelled monsoon onset dates and progression. Next we will analyse the model prediction skill for tercile categories of monsoon onset. To calculate the monsoon onset dates we use objective indices: TTGI, WYI, WFI and WLI, as described in Table 1.

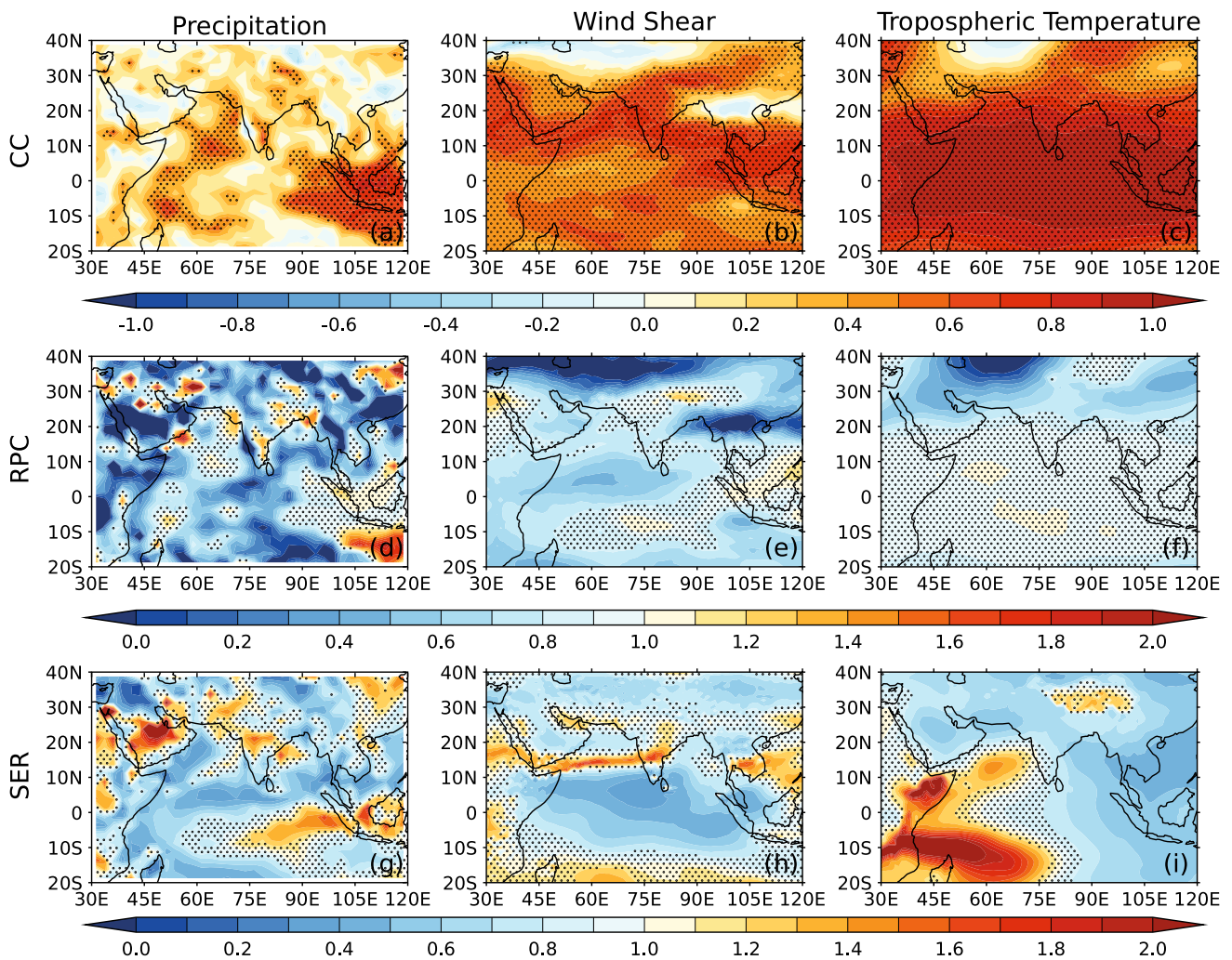
### 4.1 Climatological mean, interannual variability and skill

Before analysing SEAS5 prediction skill for the monsoon onset, we examine the climatology of the gradual progression of the monsoon over the Indian subcontinent, during different pentads (Fig. 6). Climatologically, the beginning of the monsoon onset over India is traditionally considered to occur around 1st of June (Krishnamurthy and Shukla 2000) along the coast of Kerala (Ananthakrishnan and Soman 1988). Prior to that, the onset of monsoon rains is only clearly seen over the Bay of Bengal and Arabian Sea, and a north-westerly flow dominates over the Indian

landmass, which brings dry air towards India during the pre-monsoon. In GPCP, we see monsoon onset well established over Kerala by late May, pentad 29 (Fig. 6g); however, the monsoon onset over continental India is just beginning in SEAS5 (Fig. 6a). Monsoon onset over Kerala for SEAS5 is delayed by a pentad compared to GPCP, and occurs by pentad 30 (Fig. 6b).

As the summer progresses, monsoon winds become established bringing in moist air from the south-west. Advance of the monsoon over India follows the south-east to north-west direction; that is, perpendicular to the monsoon flow. This is due to the presence of pre-monsoon mid-level north-westerly dry winds which are slowly eroded by low-level moist monsoon flow from the tropics (Parker et al. 2016). These dynamics are well represented in SEAS5. Further, although slow to start, the monsoon precipitation in SEAS5 encompasses almost the whole of India by pentad-39 (Fig. 6i), whereas in observations, it only does so by pentad-40 (Fig. 6t) i.e. mid-July (Krishnamurthy and Shukla 2000). SEAS5 shows similar south-east to north-west progression of the monsoon onset, as observed, but the onset progression is slower in SEAS5 than GPCP during May, and faster during July.

To present a holistic assessment of prediction skill for the ISM onset, we use different objective indices based on predominant physical aspects of the monsoon to calculate onset



**Fig. 5** Interannual CC for seasonal (JJA) mean **a** precipitation, **b** zonal vertical wind shear (difference between 850–200 hPa) and **c** tropospheric temperature (averaged over 600–200 hPa) between the SEAS5 ensemble mean and verification dataset at each grid-point.

Stippling indicates a CC that is significant at the 5% level. **d–f** show the same as **a–c** but for RPC and **g–i** show the same as **a–c** but for SER. Stippling shows RPC and SER values which are statistically indistinguishable from 1 at the 95% interval

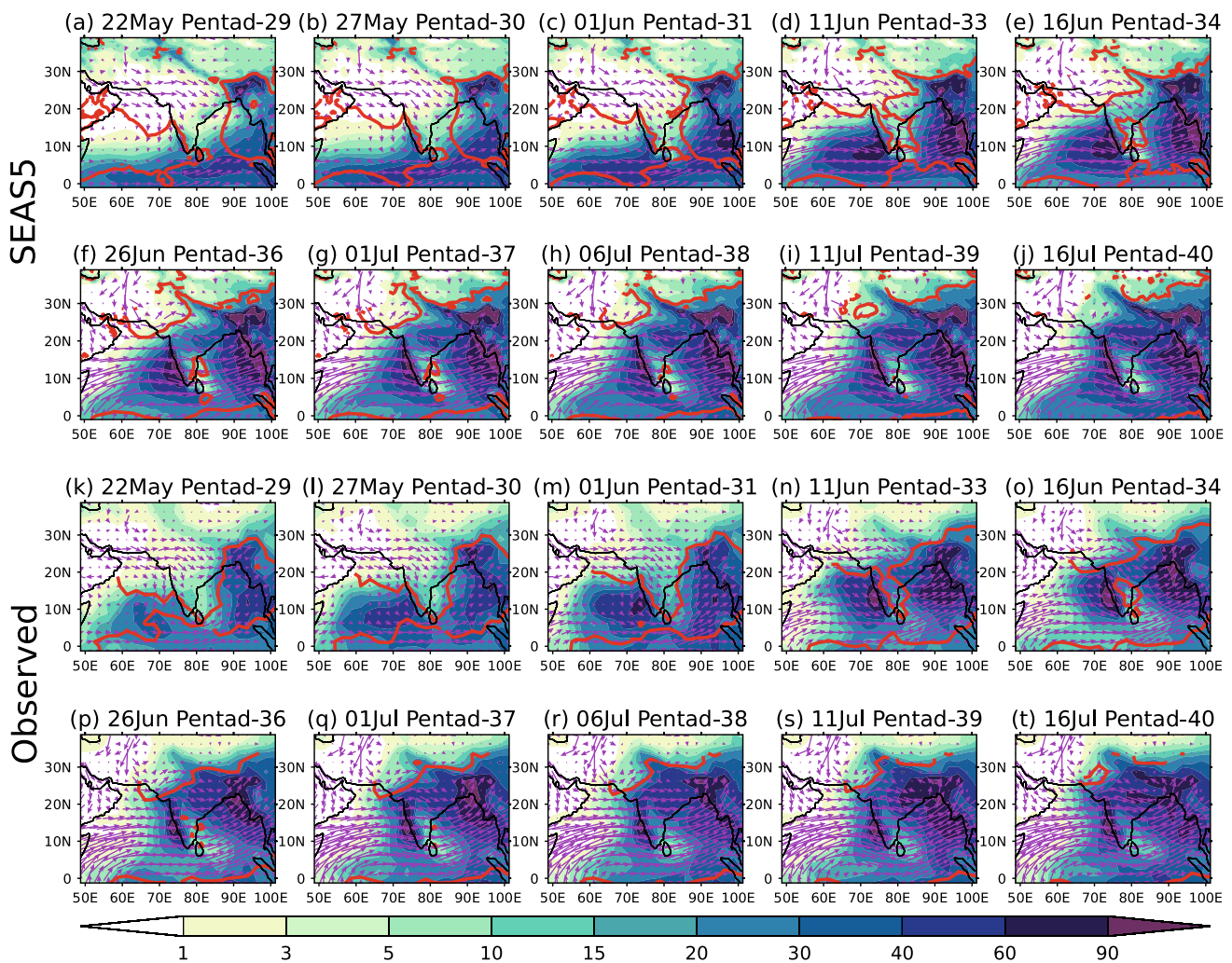
dates in the model and reanalysis, rather than verifying the model against the classical subjective criterion of increased rainfall over a small region of Kerala. SEAS5 ensemble mean climatological monsoon onset dates are statistically similar to the onset dates for reanalysis with TTGI and WYI (Table 3). For these two onset indices the interannual spread of the monsoon onset date in SEAS5 is also statistically similar to the spread in reanalysis and the interannual variability is significantly correlated between the two (Table 3). However, the seasonal mean monsoon values for the same indices (TTGI, WYI) have interannual spread that is significantly different between SEAS5 and reanalysis (Table 3).

The average onset date measured by WFI is significantly earlier in SEAS5 than ERA5 (Table 3). Even the histogram and the tercile bounds of the WFI onset dates in SEAS5 are shifted towards earlier dates than in reanalysis (Fig. 7c) and

the interannual spread is also statistically different for onset dates between SEAS5 and reanalysis (Table 3). However, SEAS5 represents the WFI onset variability moderately well, unlike for the WFI seasonal mean monsoon index. The RPC and SER of SEAS5, with all onset indices, are statistically indistinguishable from 1. RPC equal to 1 suggests that, for monsoon onset dates, SEAS5 skill is comparable to its potential predictability, i.e. the model predicts itself and the reality with similar skill. SER statistically indistinguishable from 1 suggests that the SEAS5 monsoon onset probabilistic forecasts are reliable, wherein ensemble members are statistically indistinguishable from the observations.

We also consider SEAS5 onset forecast skill at each grid-point based on the WLI (Fig. 8). SEAS5 represents the interannual variability of the monsoon onset well over parts of northern India, central India, coasts of the Indian peninsula





**Fig. 6** Spatial pattern of total pentad precipitation (mm; shaded), pentad-mean low-level winds ( $\text{m s}^{-1}$ ; vectors) and the WLI monsoon onset isochrone for that pentad (red contour). **a** Pentad-29 to **j**

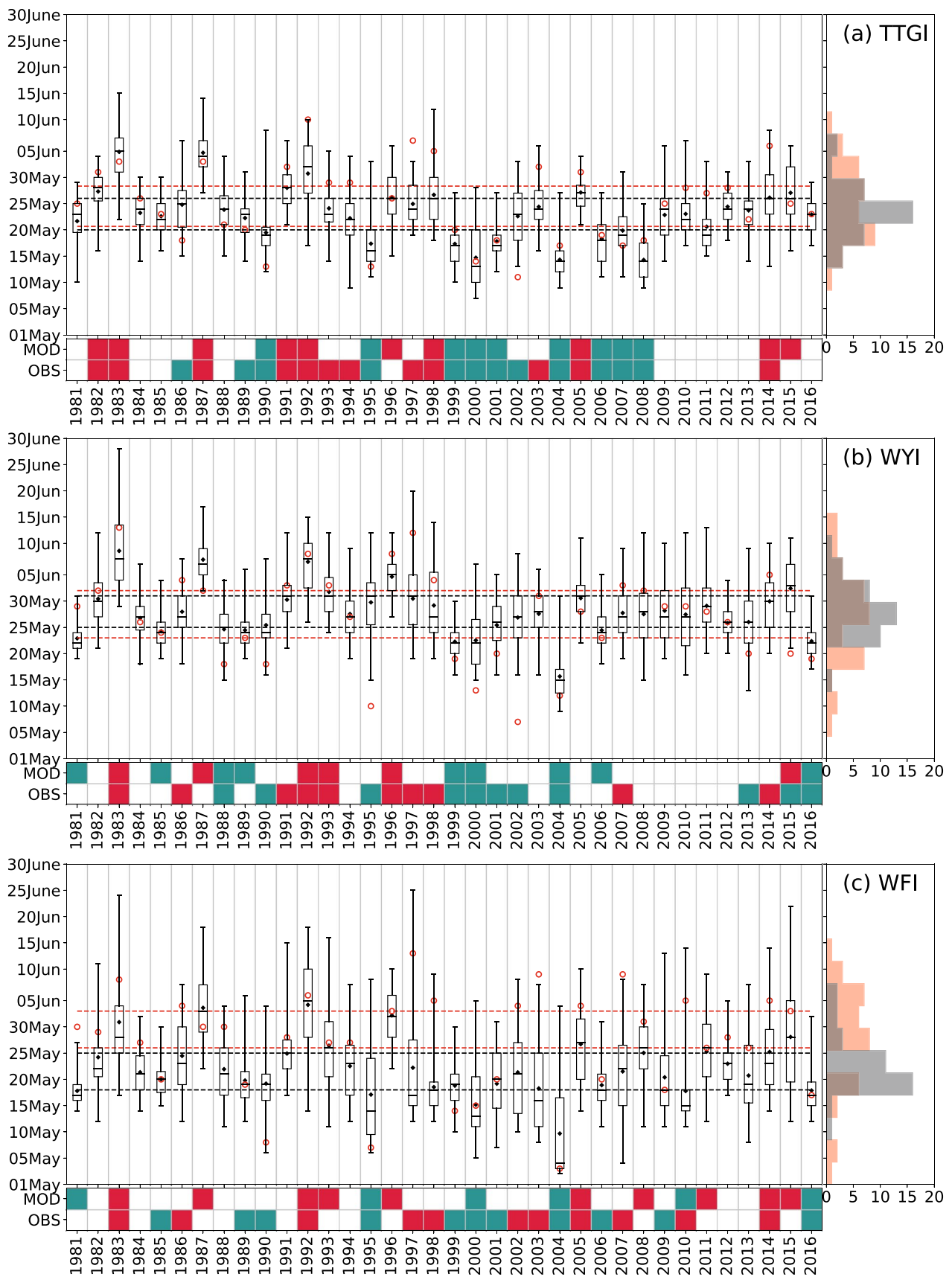
pentad-40 in SEAS5. **k–t** show the same as **a–j** but with GPCP and ERA5 datasets for verification. Dates in the subplot titles show the start dates of the respective pentads

and marginal seas of the Indian Ocean, as the hindcast is significantly correlated with observations (Fig. 8a). SEAS5 shows good forecast of onset variability for most parts of India (Fig. 8a) higher than its skill at representing seasonal mean precipitation (Fig. 5a). Over all of the Indian subcontinent and surrounding seas the forecast skill for monsoon onset is higher than that of a random chance ( $\text{HSS} > 0$ ; Fig. 8b). As the monsoon onset occurs within 1–2 months lead time, whereas seasonal mean precipitation is accumulated over months 2–4 of the forecast, this difference in skill between monsoon onset and mean precipitation is expected. However, for the local-scale onset, RPC and SER are statistically indistinguishable from 1 (Fig. 8c, d), over only some small parts of India and the marginal seas, similar to seasonal mean AIRI RPC and SER (Fig. 5d, g). SEAS5 has reliable forecasts ( $\text{SER} \approx 1$ ) and close-to-perfect predictability potential ( $\text{RPC} \approx 1$ ), over only small parts of northern,

central and western India. Over most parts of India, RPC is less than 1 (overconfident forecasts), and over some parts of eastern India and the Bay of Bengal RPC is greater than 1 (underconfident forecasts). Over most parts of India, SER is greater than 1, indicating unreliable underconfident forecasts, where the forecast spread exceeds the ensemble mean error.

## 4.2 Tercile forecast skill

Forecasting systems are generally better at categorical forecasts than absolute deterministic forecasts. SEAS5 ensemble mean ISM onset forecast skill, for the correct onset category (defined by early, normal and late terciles), for all single onset indices has been verified using ACC and HSS (Fig. 9a–b; see Sect. 2.2.2). Higher values for ACC and HSS indicate better forecast skill, with positive values of HSS





**Fig. 7** Same as Fig. 2 but showing the onset dates calculated with a TTGI, b WYI and c WFI. The dashed horizontal lines show the upper and lower terciles for monsoon onset for SEAS5 (black) and ERA5 (red). The heatmap below the boxplot shows the tercile categories of the onset dates for SEAS5 and ERA5. If the onset date of SEAS5 or ERA5 lies in between their respective upper and lower tercile lines, then it is considered a normal onset (white box in heatmap), but if it is before or after it is considered an early (blue box in heatmap) or late (red box in heatmap) onset respectively. Histogram distributions of the onset dates for SEAS5 ensemble (black) and ERA5 (red) are shown on the right of the boxplots corresponding to the onset dates on the y-axis

indicating forecast skill better than that from a randomly generated forecast. As for the deterministic forecast skill for the onset dates (Table 3), SEAS5 ensemble mean prediction skill for monsoon onset categories (Fig. 9a–b) is also higher with larger scale onset indices (TTGI and WYI) than a smaller scale index (WFI).

The overall performance of SEAS5 in delivering a probabilistic categorical forecast is indicated by BSS and RPSS (see Sect. 2.2.2), shown in Fig. 9c–f. Higher positive values of these skill scores indicate better forecast skill (RPSS and BSS range from  $-\infty$  to 1). Any value higher than 0 for BSS and RPSS indicates a forecast better than that from climatology. BSS provides skill for SEAS5 in forecasting onset for each category; early (BSS-E), normal (BSS-N) and late (BSS-L). RPSS summarizes the model performance scores over the three tercile categories. Figure 9 shows that SEAS5 categorical probabilistic forecast skill is better than that of a random forecast (positive HSS) and a climatological forecast (positive RPSS and BSS for all categories).

The skill in SEAS5 increases with the scale of the monsoon onset index, from indices computed at the local-scale (WFI) to those at the large scale (TTGI), perhaps due to the model's ability to simulate the large-scale monsoon features better. Early monsoon onsets are better predicted than later onsets in SEAS5, due to the shorter lead time of forecast. However, skill in different categories of onset, calculated using BSS, does not change linearly with the spatial scale of onset index. Forecast skill for WFI in the late onset category is worse than climatology, and the model does relatively better for the early onset category with WFI. This might stem from the fact that SEAS5 generally predicts earlier onsets with WFI than in reanalysis (Fig. 7c) due to stronger low-level westerlies (Fig. 1k) strengthening the WFI index in SEAS5 (Table 3).

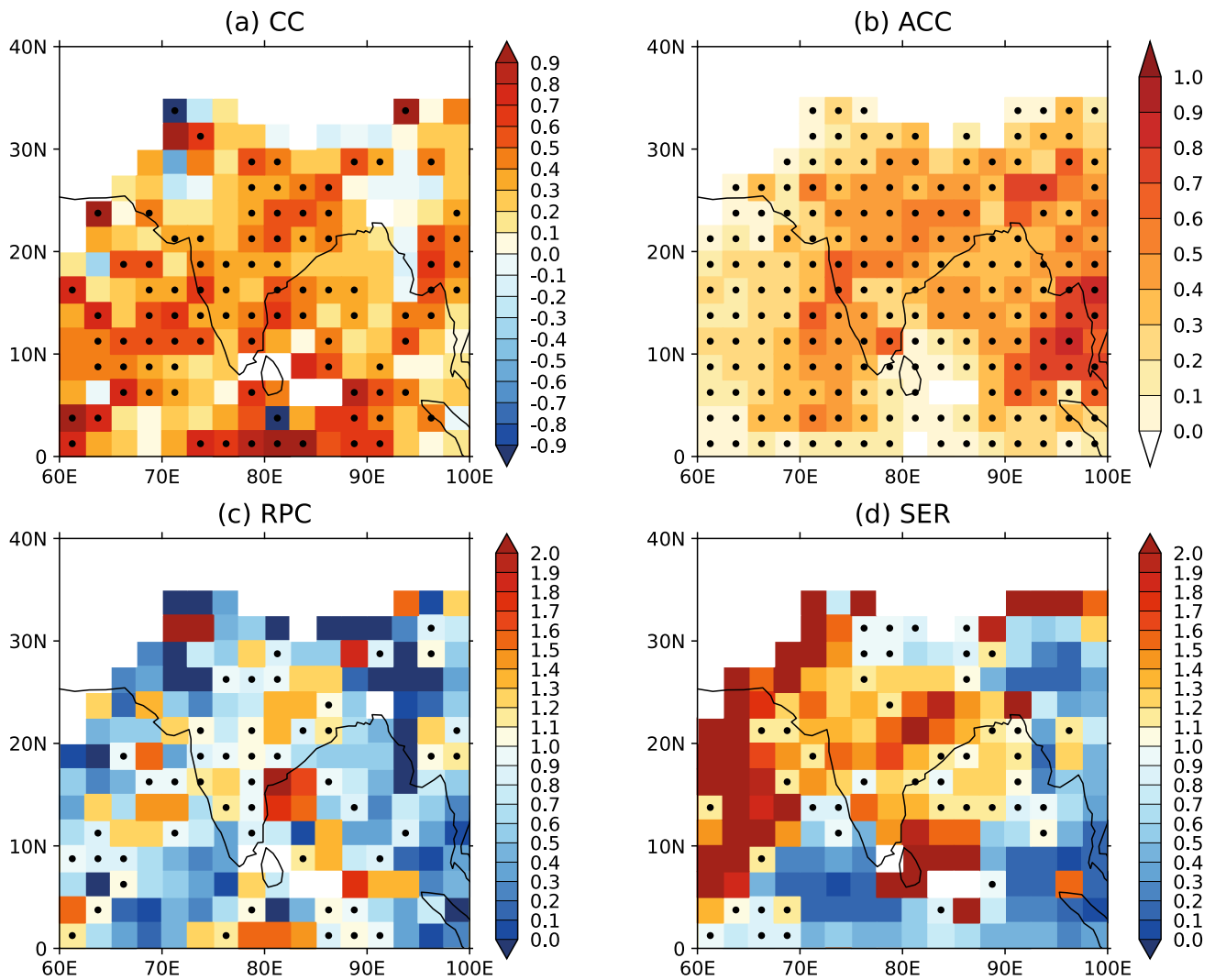
SEAS5 skill for representing monsoon onset categories at the local-scale is shown in Fig. 8b. The SEAS5 ensemble forecast has the skill to represent the local-scale monsoon onset category accurately in more than 50% of the forecasts ( $ACC > 0.5$ ) over most parts of India. The model predicts tercile categories of monsoon onset better than that of a random forecast over large parts of India ( $HSS > 0$ ). Similar to other forecasting systems like GloSea5-GC2 (Chevuturi

et al. 2019), SEAS5 has very good skill at predicting monsoon onset categories over most of the Indian subcontinent, whereas deterministic skill for the monsoon onset is limited to certain regions.

## 5 Conclusion and discussion

We have assessed the seasonal prediction skill of the Indian monsoon and its onset in the ECMWF SEAS5 coupled ensemble seasonal forecast system. Using multiple monsoon indices we verified the deterministic, probabilistic and categorical skill of the SEAS5 forecasting system. SEAS5 shows an overall dry bias over the Indian subcontinent, as seen in other contemporary models (Jain et al. 2019). The strengthened lower-level monsoon winds (caused by a warmer than observed temperature gradient) are associated with increased rainfall over the Arabian Sea and Western Ghats. SEAS5 has notable mean-state ISM precipitation errors, including overestimation of rainfall over the high orography regions (Western Ghats and Himalayas) and underestimation of rainfall over the Gangetic plains and delta, as shown by other seasonal forecasting systems (Singh et al. 2019). Known difficulties in representing orographic precipitation (Pokhrel et al. 2016) and irrigation in surface processes (Mathur and Achuta-Rao 2020) may play a role in these errors. We should, however, be cautious of the fact that the observations over the high orography have large uncertainties due to sparse observational networks. Despite local-scale precipitation errors, SEAS5 represents the interannual variability of the precipitation patterns, associated with the first two EOFs, moderately well. SEAS5 has better skill at predicting the large-scale circulation variability than the all India rainfall, consistent with other forecasting systems (Kim et al. 2012; Johnson et al. 2017). SEAS5 also has better skill for seasonal mean monsoon indices averaged over their defined domains compared to the respective monsoon index calculated over each grid-point. This is because spatially averaging over larger domains yields improved skill due to extended spatial coherence of monsoon variability (Jain et al. 2019). The progression of the monsoon onset over the Indian subcontinent in SEAS5, compared to that observed, is slower during May and faster during July.

SEAS5 has small biases in representing the climatological strengths of JJA mean monsoon indices as well as the mean monsoon onset dates calculated with smaller scale indices. However, the interannual spread for the JJA mean ISM monsoon indices in SEAS5 is generally higher than observed, although the spread in monsoon onset dates is statistically similar to that of the observations. This difference in the skill for mean monsoon features versus monsoon onset dates is also seen in the UK Met Office



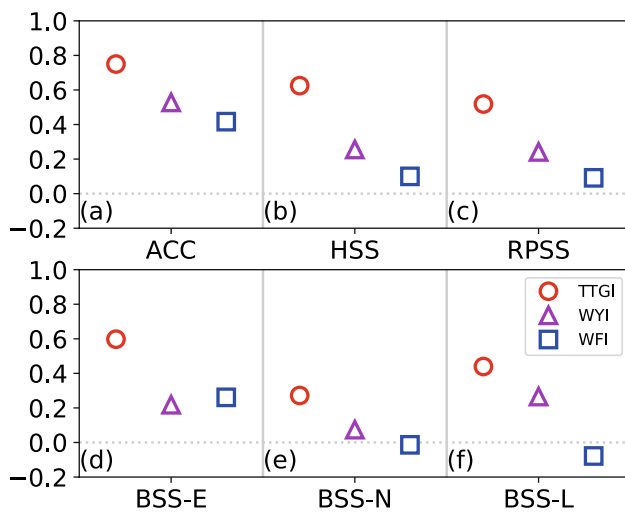
**Fig. 8** **a** Grid-point wise CC of WLI onset pentads between SEAS5 and GPCP. Stippling shows locations where the p-value is less than 0.05 (significant at the 5% level). **b** ACC calculated at each grid-point for WLI onset pentads based on tercile categories. Stippling shows

regions of positive HSS. **c** same as **a** but for RPC, and **d** same as **a** but for SER. Stippling shows RPC and SER values which are statistically indistinguishable from 1 at the 95% confidence interval

seasonal coupled forecast model, GloSea5-GC2 (Johnson et al. 2017; Chevuturi et al. 2019), which is linked to the difference in forecast lead times. SEAS5 has good skill at representing the large-scale monsoon onset date for ISM and large-scale JJA monsoon, with reliable probabilistic forecasts ( $SER \approx 1$ ) and close-to-perfect predictability potential ( $RPC \approx 1$ ). Generally, SEAS5 has SER and RPC values significantly lower than 1, at the 95% confidence interval, for smaller scale monsoon indices, such as all India rainfall, which indicates unreliable (underdispersive or overconfident) forecasts. Thus, we can conclude that the SEAS5 forecasting systems for small-scale monsoon features predicts itself better than it predicts the reality, as seen in GloSea5-GC2 (Chevuturi et al. 2019). Parts of

central India in SEAS5 have unreliable (overdispersive and underconfident) forecasts for both monsoon rainfall and onset, which indicates that the forecast ensemble spread for these features is much larger than the ensemble mean error.

Our results show a steady decrease in skill as we move from onsets calculated using large-scale indices to those at smaller scales. However, we also show that this is not true when we analyse the skill for different onset categories (defined by early, normal and late terciles). Onsets calculated from horizontal wind shear (WFI) have higher skill for early onsets and lower skill for late onsets. Due to the model's westerly wind bias at lower levels, WFI in SEAS5 is stronger than in observations. This strengthened WFI leads



**Fig. 9** Model skill-scores compared against ERA5 shown as **a** ACC, **b** HSS, **c** RPSS, and BSS of model for predicting **d** early onsets (BSS-E), **e** normal onsets (BSS-N) and **f** late onsets (BSS-L). The colored symbols show the skill calculated for model onset with the three different indices: TTGI (red circle), WYI (magenta triangle) and WFI (blue square). The grey horizontal dotted line demarcates the 0 threshold

to an increased tendency in SEAS5 for earlier WFI onsets. In SEAS5, the application of mean-state bias correction techniques to reduce the error in low-level circulation may improve the representation of precipitation biases and the associated monsoon onset date.

SEAS5 has skill which is better than either random or climatological forecasts, when giving probabilistic forecasts for the monsoon onset category. This skill for SEAS5 is not only applicable for large-scale onset prediction but also for rainfall onsets at the local scale on a pentad-by-pentad basis. Better skill for monsoon onset compared to the poor skill for precipitation forecasts in SEAS5 is also seen in GloSea5-GC2 (Johnson et al. 2017; Chevuturi et al. 2019), which is associated with the difference in forecast lead time for the mean monsoon precipitation (2–4 months) and monsoon onset (1–2 months). Over most parts of India, SEAS5 shows good skill at forecasting the onset pentad category and can predict the onset pentad accurately with moderate skill over parts of northern central, western and eastern India, despite the systematic biases in precipitation. The two seasonal forecast models, SEAS5 and GloSea5-GC2 (Chevuturi et al. 2019), both show good skill at predicting local-scale monsoon onset over the core monsoon region at the same lead time (May forecasts). Previous studies have shown multi-model ensembles to enhance forecast skill for Indian monsoon rainfall (e.g. Kumar et al. 2012). Such a multi-model ensemble should improve forecast skill for the local-scale monsoon onset, but detailed analysis is

required in future in order to identify the best approach for multi-model combinations.

Operationally, IMD issues probabilistic seasonal forecasts of ISM by mid-April, with an update by 1st of June for region-wise or all India rainfall using statistical and dynamical models (IMD 2020). IMD also issues a monsoon onset date for Kerala using a statistical forecast model by mid-May and have recently updated their methodology to identify local-scale monsoon onset date with gridded datasets rather than station information (Pai et al. 2020). Local-scale agro-advisories are currently only provided to the farmers at different timescales, through the Gramin Krishi Mausam Sewa project by the Agricultural Meteorology Division of IMD (AMD 2020). Although the current study's outcomes may not be directly beneficial to the end-user (e.g. farmers), our results show that SEAS5 has appreciable skill for operational state-level products of ISM rainfall and local-scale monsoon onset almost a month in advance over parts of India. Future investigation can help identify specific user-oriented ISM metrics and SEAS5's skill for such metrics relative to current operation forecasts (Rao et al. 2019). Good prediction for the local-scale monsoon features provided to the farmers a month in advance, over major agricultural regions may help reduce resource wastage, mitigate losses and improve crop yield though better-informed decision making in the agriculture sector.

**Acknowledgements** This study was partially supported by the National Centre for Atmospheric Science ODA national capability programme ACREW (NE/R000034/1), which is supported by NERC and the GCRF. This study has also been partially funding by the Copernicus Climate Change Service. ECMWF implements this service and the Copernicus Atmosphere Monitoring Service on behalf of the European Commission. The contribution of AT is funded through the Weather and Climate Science for Service Partnership (WCSSP) India, a collaborative initiative between the Met Office, supported by the UK Government's Newton Fund, and the Indian Ministry of Earth Sciences (MoES) as well as INCOMPASS (NERC grants NE/L01386X/1 and NE/P003117/1). The contribution of JKPS is supported by NERC grant NE/N018591/1. The hindcasts from SEAS5 (<https://cds.climate.copernicus.eu/cdsapp#!dataset/seasonal-original-single-levels?tab=overview>) and ERA5 reanalysis (<https://doi.org/10.24381/cds.6860a573>) are publicly available from ECMWF's Copernicus Climate Change Service (C3S), through its climate data store. GPCP Precipitation data provided by the NOAA/OAR/ESRL PSL, Boulder, Colorado, USA (<https://psl.noaa.gov/data/gridded/data.gpcp.html>).

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abhilash S, Sahai A, Pattnaik S, Goswami B, Kumar A (2014) Extended range prediction of active-break spells of Indian summer monsoon rainfall using an ensemble prediction system in NCEP Climate Forecast System. *Int J Climatol* 34(1):98–113. <https://doi.org/10.1002/joc.3668>
- Adler RF, Sapiiano MR, Huffman GJ, Wang JJ, Gu G, Bolvin D, Chiu L, Schneider U, Becker A, Nelkin E et al (2018) The Global Precipitation Climatology Project (GPCP) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere* 9(4):138. <https://doi.org/10.3390/atmos9040138>
- Alessandri A, Borrelli A, Cherchi A, Materia S, Navarra A, Lee JY, Wang B (2015) Prediction of Indian summer monsoon onset using dynamical subseasonal forecasts: effects of realistic initialization of the atmosphere. *Month Weather Rev* 143(3):778–793. <https://doi.org/10.1175/MWR-D-14-00187.1>
- AMD (2020) About Agrimet: Background. <http://imd.agrimet.gov.in/imdproject/AGBackground.php>. Accessed 20 Oct 2020
- Ananthakrishnan R, Soman M (1988) The onset of the southwest monsoon over Kerala: 1901–1980. *J Climatol* 8(3):283–296. <https://doi.org/10.1002/joc.3370080305>
- Chevuturi A, Turner AG, Woolnough SJ, Martin GM, MacLachlan C (2019) Indian summer monsoon onset forecast skill in the UK Met Office initialized coupled seasonal forecasting system (GloSea5-GC2). *Clim Dyn* 52(11):6599–6617. <https://doi.org/10.1007/s00382-018-4536-1>
- DelSole T, Shukla J (2010) Model fidelity versus skill in seasonal forecasting. *J Clim* 23(18):4794–4806. <https://doi.org/10.1175/2010JCLI3164.1>
- 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(15):5620–5628. <https://doi.org/10.1002/2014GL061146>
- ECMWF (2017) ECMWF SEAS5 user guide. [https://www.ecmwf.int/sites/default/files/medialibrary/2017-10/System5\\_guide.pdf](https://www.ecmwf.int/sites/default/files/medialibrary/2017-10/System5_guide.pdf)
- Findlater J (1969) A major low-level air current near the Indian Ocean during the northern summer. *Q J R Meteorol Soc* 95(404):362–380. <https://doi.org/10.1002/qj.49709640919>
- Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R, Schepers D et al (2020) The ERA5 global reanalysis. *Q J R Meteorol Soc*. <https://doi.org/10.1002/qj.3803>
- Ho CK, Hawkins E, Shaffrey L, Bröcker J, Hermanson L, Murphy JM, Smith DM, Eade R (2013) Examining reliability of seasonal to decadal sea surface temperature forecasts: The role of ensemble dispersion. *Geophys Res Lett* 40(21):5770–5775. <https://doi.org/10.1002/2013GL057630>
- IMD (2020) Seasonal Forecast. [https://mausam.imd.gov.in/imd\\_latest/contents/seasonal\\_forecast.php](https://mausam.imd.gov.in/imd_latest/contents/seasonal_forecast.php). Accessed 20 Oct 2020
- Jain S, Scaife AA, Mitra AK (2019) Skill of Indian summer monsoon rainfall prediction in multiple seasonal prediction systems. *Clim Dyn* 52(9–10):5291–5301. <https://doi.org/10.1007/s00382-018-4449-z>
- Jie W, Vitart F, Wu T, Liu X (2017) Simulations of the Asian summer monsoon in the sub-seasonal to seasonal prediction project (S2S) database. *Q J R Meteorol Soc* 143(706):2282–2295. <https://doi.org/10.1002/qj.3085>
- Johnson SJ, Turner A, Woolnough S, Martin G, MacLachlan C (2017) An assessment of Indian monsoon seasonal forecasts and mechanisms underlying monsoon interannual variability in the Met Office GloSea5-GC2 system. *Clim Dyn* 48(5–6):1447–1465. <https://doi.org/10.1007/s00382-016-3151-2>
- Johnson SJ, Stockdale TN, Ferranti L, Balmaseda MA, Molteni F, Magnusson L, Tietsche S, Decremier D, Weisheimer A, Balsamo G et al (2019) SEAS5: the new ECMWF seasonal forecast system. *Geosci Model Dev* 12(3):1087–1117. <https://doi.org/10.5194/gmd-12-1087-2019>
- Kim HM, Webster PJ, Curry JA, Toma VE (2012) Asian summer monsoon prediction in ECMWF System 4 and NCEP CFSv2 retrospective seasonal forecasts. *Clim Dyn* 39(12):2975–2991. <https://doi.org/10.1007/s00382-012-1470-5>
- Köhn-Reich L, Bürger G (2019) Dynamical prediction of Indian monsoon: Past and present skill. *Int J Climatol* 39(8):3574–3581. <https://doi.org/10.1002/joc.6039>
- Koteswaram P (1958) The easterly jet stream in the tropics. *Tellus* 10(1):43–57. <https://doi.org/10.1111/j.2153-3490.1958.tb01984.x>
- Krishna Kumar K, Hoerling M, Rajagopalan B (2005) Advancing dynamical prediction of Indian monsoon rainfall. *Geophys Res Lett*. <https://doi.org/10.1029/2004GL021979>
- Krishnamurthy V, Shukla J (2000) Intraseasonal and interannual variability of rainfall over India. *J Clim* 13(24):4366–4377. [https://doi.org/10.1175/1520-0442\(2000\)013<0001:IAIVOR>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<0001:IAIVOR>2.0.CO;2)
- Kumar A, Mitra A, Bohra A, Iyengar G, Durai V (2012) Multi-model ensemble (MME) prediction of rainfall using neural networks during monsoon season in India. *Meteorol Appl* 19(2):161–169. <https://doi.org/10.1002/met.254>
- Li J, Wang B (2016) How predictable is the anomaly pattern of the Indian summer rainfall? *Clim Dyn* 46(9–10):2847–2861. <https://doi.org/10.1007/s00382-015-2735-6>
- Mathur R, AchutaRao K (2020) A modelling exploration of the sensitivity of the India's climate to irrigation. *Clim Dyn* 54(3):1851–1872. <https://doi.org/10.1007/s00382-019-05090-8>
- Menon A, Turner A, Martin G, MacLachlan C (2018) Modelling the moistening of the free troposphere during the northward progression of Indian monsoon onset. *Q J R Meteorol Soc* 144(713):1152–1168. <https://doi.org/10.1002/qj.3281>
- Moron V, Robertson AW (2014) Interannual variability of Indian summer monsoon rainfall onset date at local scale. *Int J Climatol* 34(4):1050–1061. <https://doi.org/10.1002/joc.3745>
- Nanjundiah RS, Francis P, Ved M, Gadgil S (2013) Predicting the extremes of Indian summer monsoon rainfall with coupled ocean-atmosphere models. *Curr Sci* 104(10):1380–1393. <https://www.currentscience.ac.in/Volumes/104/10/1380.pdf>
- Pai D, Bandgar A, Devi S, Musale M, Badwaik M, Kundale A, Gadgil S, Mohapatra M, Rajeevan M (2020) New normal dates of onset/progress and withdrawal of southwest monsoon over India. CRS Research Report RR No. 3/2020. [https://internal.imd.gov.in/press\\_release/20200515\\_pr\\_804.pdf](https://internal.imd.gov.in/press_release/20200515_pr_804.pdf). Accessed 20 Oct 2020
- Pandey DK, Rai S, Shahi N, Mishra N (2015) Seasonal prediction of ISMR and relationship with EL-NINO and IOD in ECMWF system 4 coupled model. *Clim Change* 1(4):447–455. [http://www.discoveryjournals.org/climate\\_change/current\\_issue/v1/n4/A17.pdf](http://www.discoveryjournals.org/climate_change/current_issue/v1/n4/A17.pdf)
- Parker DJ, Willetts P, Birch C, Turner AG, Marsham JH, Taylor CM, Kollu S, Martin GM (2016) The interaction of moist convection and mid-level dry air in the advance of the onset of the Indian monsoon. *Q J R Meteorol Soc* 142(699):2256–2272. <https://doi.org/10.1002/qj.2815>
- Pokhrel S, Saha SK, Dhakate A, Rahman H, Chaudhari HS, Salunke K, Hazra A, Sujith K, Sikka D (2016) Seasonal prediction of Indian summer monsoon rainfall in NCEP CFSv2: forecast and predictability error. *Clim Dyn* 46(7–8):2305–2326. <https://doi.org/10.1007/s00382-015-2703-1>
- Prakash S, Mitra AK, Momin IM, Rajagopal E, Basu S, Collins M, Turner AG, Achuta Rao K, Ashok K (2015) Seasonal intercomparison of observational rainfall datasets over India during the southwest monsoon season. *Int J Climatol* 35(9):2326–2338. <https://doi.org/10.1002/joc.4129>



- Preethi B, Kripalani R, Kumar KK (2010) Indian summer monsoon rainfall variability in global coupled ocean-atmospheric models. *Clim Dyn* 35(7–8):1521–1539. <https://doi.org/10.1007/s00382-009-0657-x>
- Rajeevan M, Unnikrishnan C, Preethi B (2012) Evaluation of the ENSEMBLES multi-model seasonal forecasts of Indian summer monsoon variability. *Clim Dyn* 38(11–12):2257–2274. <https://doi.org/10.1007/s00382-011-1061-x>
- Ramu DA, Sabeerali C, Chattopadhyay R, Rao DN, George G, Dhakate A, Salunke K, Srivastava A, Rao SA (2016) Indian summer monsoon rainfall simulation and prediction skill in the CFSv2 coupled model: Impact of atmospheric horizontal resolution. *J Geophys Res Atmos* 121(5):2205–2221. <https://doi.org/10.1002/2015JD024629>
- Rao SA, Goswami B, Sahai AK, Rajagopal E, Mukhopadhyay P, Rajeevan M, Nayak S, Rathore L, Shenoi S, Ramesh K et al (2019) Monsoon mission: a targeted activity to improve monsoon prediction across scales. *Bull Am Meteorol Soc* 100(12):2509–2532. <https://doi.org/10.1175/BAMS-D-17-0330.1>
- Saha S, Pokhrel S, Chaudhari H, Dhakate A, Shewale S, Sabeerali C, Salunke K, Hazra A, Mahapatra S, Rao A (2014) Improved simulation of Indian summer monsoon in latest NCEP climate forecast system free run. *Int J Climatol* 34(5):1628–1641. <https://doi.org/10.1002/joc.3791>
- Singh B, Cash B, Kinter JL III (2019) Indian summer monsoon variability forecasts in the North American multimodel ensemble. *Clim Dyn* 53(12):7321–7334. <https://doi.org/10.1007/s00382-018-4203-6>
- Sperber K, Brankovic C, Deque M, Frederiksen C, Graham R, Kitoh A, Kobayashi C, Palmer T, Puri K, Tennant W et al (2001) Dynamical seasonal predictability of the Asian summer monsoon. *Month Weather Rev* 129(9):2226–2248. [https://doi.org/10.1175/1520-0493\(2001\)129%3C2226:DSPOTA%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129%3C2226:DSPOTA%3E2.0.CO;2)
- Wang B, Fan Z (1999) Choice of South Asian summer monsoon indices. *Bull Am Meteorol Soc* 80(4):629–638. [https://doi.org/10.1175/1520-0477\(1999\)080%3C0629:COSAS M%3E2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080%3C0629:COSAS M%3E2.0.CO;2)
- Wang B, LinHo L (2002) Rainy season of the Asian-Pacific summer monsoon. *J Clim* 15(4):386–398. [https://doi.org/10.1175/1520-0442\(2002\)015%3C0386:RSOTAP%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015%3C0386:RSOTAP%3E2.0.CO;2)
- Wang B, Lee JY, Xiang B (2015a) Asian summer monsoon rainfall predictability: a predictable mode analysis. *Clim Dyn* 44(1–2):61–74. <https://doi.org/10.1007/s00382-014-2218-1>
- Wang B, Xiang B, Li J, Webster PJ, Rajeevan MN, Liu J, Ha KJ (2015b) Rethinking Indian monsoon rainfall prediction in the context of recent global warming. *Nat Commun* 6:7154. <https://doi.org/10.1038/ncomms8154>
- WCRP (2015) Forecast Verification-Issues, Methods, and FAQ. <http://www.cawcr.gov.au/projects/verification/>. Accessed 1 Nov 2019
- Webster PJ, Yang S (1992) Monsoon and ENSO: selectively interactive systems. *Q J R Meteorol Soc* 118(507):877–926. <https://doi.org/10.1002/qj.49711850705>
- Weisheimer A, Palmer T, Doblas-Reyes F (2011) Assessment of representations of model uncertainty in monthly and seasonal forecast ensembles. *Geophys Res Lett*. <https://doi.org/10.1029/2011GL048123>
- Xavier PK, Marzin C, Goswami B (2007) An objective definition of the Indian summer monsoon season and a new perspective on the ENSO-monsoon relationship. *Q J R Meteorol Soc* 133(624):749–764. <https://doi.org/10.1002/qj.45>
- Xie P, Janowiak JE, Arkin PA, Adler R, Gruber A, Ferraro R, Huffman GJ, Curtis S (2003) GPCP pentad precipitation analyses: an experimental dataset based on gauge observations and satellite estimates. *J Clim* 16(13):2197–2214. <https://doi.org/10.1175/2769.1>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.