



Sub-seasonal prediction skill: is the mean state a good model evaluation metric?

Jihun Ryu¹ · Shih-Yu (Simon) Wang² · Jee-Hoon Jeong³ · Jin-Ho Yoon¹

Received: 29 November 2023 / Accepted: 14 June 2024
© The Author(s) 2024

Abstract

Evaluating forecast models encompasses assessing their ability to accurately depict observed climate states and predict future climate variables. Various evaluation methods, from computationally efficient measures like the anomaly correlation coefficient to more intricate approaches, have been formulated. While simpler methods provide limited information, climatology, due to its simplicity and immediate linkage to model performance, is a commonly utilized primary evaluation metric. In this study focusing on temperature and precipitation, we propose a novel metric based on the model's mean state, integrating both climatology and the seasonal cycle for a more accurate assessment of the relationship between mean state performance and prediction skill on weather and sub-seasonal time scales compared to relying solely on climatology. This integrated metric reveals a robust correlation between temperature and precipitation across diverse geographical locations, with a more pronounced effect in tropical areas when considering the seasonal cycle. Additionally, we find that temperature exhibits higher prediction skill compared to precipitation. The discovered relationship serves as a potential early indicator for predicting the efficacy of Seasonal to Sub-seasonal (S2S) models and offers valuable insights for model development, emphasizing the significance of this integrated metric in enhancing S2S model performance and advancing climate prediction capabilities.

Keywords Sub-seasonal to Seasonal(S2S) · Prediction skill · Mean state · S2S project

1 Introduction

Despite decades of scientific achievements, sub-seasonal prediction skill experiences a substantial decline beyond 3–4 weeks, particularly in contrast to the 1–2 week range (de Andrade et al. 2021). Such achievements efforts have been dedicated to enhancing sub-seasonal prediction skill, exemplified by initiatives like the World Weather Research Program (WWRP)/World Climate Research Program (WCRP) Sub-seasonal to Seasonal prediction (S2S) project, which aims to advance forecast skill (Robertson et al. 2015; Vitart et al. 2012). Over the past decade, this project has amassed

hindcasts and real-time forecasts from 12 models, resulting in notable advancements in predicting the Madden-Julian Oscillation (MJO), considered one of the most significant variabilities on sub-seasonal timescales (Kim et al. 2018). Particularly noteworthy is its success in forecasting extreme events, such as the 2010 Russian heatwave and the July 2015 West-European heatwave, with lead times of up to three to four weeks (Ardilouze et al. 2017; Vitart and Robertson 2018).

The assessment of a forecasting model is integral to its development, encompassing the evaluation of both the mean state and prediction skill. The mean state assessment, rooted in climatology, utilizes straightforward metrics such as root mean square error (RMSE) and correlation coefficient to measure the model's proficiency in replicating long-term averages. In evaluating anomaly-based prediction skill, metrics range from the simplicity of RMSE and the anomaly correlation coefficient (ACC) to more complex approaches, such as a six-step framework (Coelho et al. 2018).

Improving the simulation of the mean state in climate models is recognized as a key factor in enhancing their forecast skills, extending to the accurate representation of

✉ Jin-Ho Yoon
yjinho@gist.ac.kr

¹ School of Environment and Energy Engineering, Gwangju Institute of Science and Technology, Gwangju, South Korea

² Department of Plants, Soils and Climate, Utah State University, Logan, UT, USA

³ Faculty of Earth and Environmental Sciences, Chonnam National University, Gwangju, South Korea

interannual variability in regions such as the equatorial Atlantic (Ding et al. 2015). Further, climate models that effectively simulate the tropical Pacific's cold tongue exhibit reduced mean state bias in the equatorial Pacific (Ding et al. 2020). Similarly, within the realm of seasonal forecasting, the 1-month lead seasonal forecasting skill has been found to be associated with both the annual mean and the annual cycle (Lee et al. 2010). Overall, a more profound comprehension of the relationship between prediction skill and the mean state on sub-seasonal timescales is essential. If such a relationship exists, the mean state could potentially serve as both an evaluation metric and an indicator of prediction skill.

In summary, there are two main points. First, straightforward methods for evaluating prediction skill are time-efficient but yield rather limited information. In contrast, more elaborate methods offer an in-depth assessment but likely require a considerable investment of time for analysis. Second, the relationship between prediction skill and the mean state on sub-seasonal timescales has not been thoroughly investigated. Bearing these in mind, this study endeavors to explore the relationship between prediction skill in the sub-seasonal timescales and mean state.

The goal of this study is to answer the following two primary questions: (1) Is there a noticeable relationship between performance of the mean state and prediction skill in sub-seasonal timescales, and (2) Does this relationship differ based on regions, seasons, or forecast lead times? Section 2 provides the data and methodology employed to evaluate prediction skill and mean state simulation performance including improved metrics tailored for our research goals. Section 3 presents the findings regarding the evaluation of prediction skill and mean state simulation performance, along with the relationship between the two. Finally, Section 4 concludes the paper and offers further discussion.

2 Data and method

2.1 Data

In this study, temperature and precipitation hindcasts from 11 models within the S2S project were analyzed, with detailed information about each model provided in Table 1. It is important to note that (i) the Japan Meteorological Agency (JMA) result was excluded due to its hindcast frequency being twice monthly, deemed insufficient for a robust sample size (Vitart et al. 2017), and (ii) the UK Met Office (UKMO) has two versions, GloSea5(GS5) and GloSea6(GS6), which were both analyzed. To address differences in ensemble sizes across the models, only the control forecast was analyzed. The study focused on a ten-year dataset spanning from 2001

to 2010, utilizing 32 forecast days and hindcasts interpolated on a $1.5^\circ \times 1.5^\circ$ latitude-longitude grid.

For the validation of the S2S models, temperature data from ERA5 (Hersbach et al. 2020) and precipitation data from the Global Precipitation Climatology Project (GPCP) version 1.3 were employed (Huffman et al. 2001). Both datasets were re-gridded to match the resolution of the S2S data.

2.2 Method

The objective of this study was to assess the performance of the S2S models on a global scale. The evaluation considered two primary aspects: (i) Mean state and (ii) Prediction skill. To discern regional traits, the world was divided into 36 areas using a $30^\circ \times 60^\circ$ latitude-longitude grid. Both the mean state and prediction skill were analyzed within these areas, which were further categorized into global, mid-latitude, and tropical regions. The results were the same for a smaller region based on a $30^\circ \times 30^\circ$ latitude-longitude grid.

2.2.1 Evaluation metric for the mean state

Here, the mean state is characterized as a state that incorporates both climatology and the seasonal cycle (Lee et al. 2010). The first step is to compute the climatology of each S2S model. A challenge arises because the forecast frequency of the S2S model is not sufficient to uniformly compute daily climatology across different models. Therefore, this study opts to calculate a monthly climatology with a 0-month lead forecast. Initially, the forecast data for each month was averaged if it contained more than 15 days in the forecast's start month. Subsequently, the averaged forecast data for the identical months was averaged again to obtain the 0-month lead forecast. This process yielded the monthly climatology.

By employing the monthly climatology derived from a 0-month lead forecast, an evaluation metric for climatology, E_{cli} , was calculated according to Eq. 1, indicating the performance of climatology. This metric uses both the correlation (r) and normalized RMSE (nRMSE) between each S2S model and the reanalysis data. Specifically, the nRMSE across regions was considered to prevent the biasing of results towards high RMSE values in particular areas. This study utilizes the average of the two metrics because they are the most widely used metrics for evaluating models, with the correlation coefficient indicating the linear relationship between the model and reality, and the nRMSE indicating how well the model is simulating quantitatively. To give equal weight to both indicators, the r and nRMSE were subjected to min-max normalization for S2S models. For the r , we utilized the normalized values directly, assigning the highest-performing model a score of 1 and the lowest a score of 0. For nRMSE, the model with the smallest nRMSE is

Table 1 Details of the S2S models used in this study

Model (Center)	Model version	Time range	Spatial resolution	Hindcast frequency	Hindcast period
BoM (Australian Bureau of Meteorology)	POAMAP24	Days 0–62	T47, L17	6/month	1981–2013
CMA (China Meteorological Administration)	BCC-CPS-S2Sv1	Days 0–60	T106, L40	Daily	1994–2014
CNRM (Météo-France/Centre National de Recherche Météorologiques)	NCRM-CM6.1	Days 0–47	T359, L91	Every 7 days	1993–2017
ECMWF (European Centre for Medium-Range Weather Forecasts)	GEPS6	Days 0–32	Yin-Yang grid at 0.35° uniform resolution, L45	Weekly	1998–2017
ECMWF (European Centre for Medium-Range Weather Forecasts)	CY47R1	Days 0–46	Tco639/Tco319, L91	2/week	Past 20 years
HMCR (Hydrometeorological Centre of Russia)	RUMS (08/06/2017)	Days 0–61	1.1° × 1.4°, L28	Weekly	1985–2010
IAP (Institute of Atmospheric Physics and Chinese Academy of Sciences)	CAS-FGOALS-f2-V1.3	Days 0–65	C96, L32	Daily	1999–2018
ISAC (Institute of Atmospheric Sciences and Climate of the National Research Council)	GLOBO (08/06/2017)	Days 0–32	0.75° × 0.56°, L54	Every 5 days	1981–2010
KMA (Korea Meteorological Administration)	GloSea5-GC2	Days 0–60	M216, L85	4/month	1991–2016
NCEP (National Centers for Environmental Prediction)	CFSv2	Days 0–44	T126, L64	Daily	1999–2010
UKMO (United Kingdom's Met Office)	GloSea5-GC2-LI	Days 0–60	N216, L85	4/month	1993–2016
UKMO (United Kingdom's Met Office)	GloSea6	Days 0–60	N216, L85	4/month	1993–2016

given a 0, and the model with the largest nRMSE is a 1. We then adjusted the nRMSE by subtracting 1 from its normalized value and taking the absolute, so that the model with the smallest nRMSE earns a 1, aligning with the model with the best r . Consequently, a model that tops both metrics receives a composite score of 1, denoting optimal performance.

$$E_{cli} = \frac{1}{2} (Norm_{min-max}(r) + Norm_{min-max}(nRMSE)) \quad (1)$$

Following that, the seasonal cycle evaluation metric was computed. Empirical Orthogonal Function (EOF) analysis was used to derive two annual and two semi-annual cycles from the monthly climatology computed from the 0-month lead forecast. A previous study has considered up to two semi-annual cycles for precipitation (Meyer et al. 2021). However, in the case of temperature, the fraction of variance of the semi-annual cycle was generally small in most regions. For example, the variance of the first through fourth

EOF modes of the monthly climatology of global temperature is 92.3%, 5.7%, 1.1%, and 0.6%, respectively. On the other hand, precipitation showed a variance of 62.8%, 16.5%, 8.0%, and 4.8%, respectively. As a result, only two annual cycles were included in temperature, while two annual and two semi-annual cycles were used for precipitation.

The seasonal cycle metric, E_{SC} , as defined in Eq. 2, was computed as the sum of the product of the time series correlation of the Principal Component (PC) time series ($r(PCt_{obs,mod}, PCt_{pre,mod})$) and the pattern correlation of the PC spatial pattern ($r(PCs_{obs,mod}, PCs_{pre,mod})$) for each mode. This metric takes into account both the direction and intensity of each mode in the seasonal cycle. To accentuate the distinctions between the modes, min-max normalization was applied. Essentially, E_{SC} considers the alignment and strength of each seasonal cycle mode and is a combination of the correlations of the PC time series and the PC spatial patterns for every mode, incorporating a constant derived through integration. Min-max normalization was utilized to emphasize the differences among them.

$$E_{SC} = \frac{1}{M} \sum_{mod=1}^M Norm_{min-max} \{ r(PCs_{obs,mod}, PCs_{pre,mod}) \times r(PCt_{obs,mod}, PCt_{pre,mod}) \} \quad (2)$$

Finally, a metric representing the mean state, E_{MS} , was computed according to Eq. 3, by simply adding both climatology and the seasonal cycle. In this computation, equal weights were assigned to climatology and the seasonal cycle, and the average of values obtained from Eqs. 1 and 2 was calculated. The theoretical maximum value of E_{MS} is 1, which signifies superior performance.

$$E_{MS} = \frac{E_{cli} + E_{SC}}{2} \quad (3)$$

2.2.2 Evaluation metric for the prediction skill

Various approaches are employed for the evaluation of model prediction skill. Direct assessment of variables often employs simple and widely-used metrics, such as RMSE and the ACC (Li and Robertson 2015; Zhu et al. 2014). The Taylor diagram offers a slightly more advanced method, providing a graphical representation that combines variance and correlation coefficient metrics (Taylor 2001). This graphical representation can be further condensed into a single metric (Yang et al. 2013). Additionally, more intricate approaches, like the six-step framework, have been utilized to generate a comprehensive spectrum of quality assessment data (Coelho et al. 2018; de Andrade et al. 2021). Alongside the

overall prediction skill assessment, some previous studies have focused on specific and critical climate variabilities, such as the MJO, quasi-biennial oscillation (QBO), and El Niño–Southern Oscillation (ENSO), to evaluate sub-seasonal prediction skill (de Andrade et al. 2019; Kim et al. 2019, 2020; Li and Robertson 2015; Lim et al. 2018). Additionally, assessments of specific phenomena, such as the onset dates of rainfall, have been explored (Kumi et al. 2020).

To assess prediction skill, precipitation, and temperature anomalies were calculated for each model's forecast lead time in the daily time scale. Due to the limited daily hindcast frequency available in just three models, a 7-day moving average was applied to obtain the daily climatology required for anomaly calculations. Following the computation of anomalies, three metrics were utilized: the ACC, RMSE, and Eq. 4. Notably, Eq. 4, a formula employed in previous studies (Wang et al. 2021; Yang et al. 2013), was predominantly used for our assessment. This particular metric is built upon Taylor diagrams and employs the standard deviation (σ) and the pattern correlation coefficient (r). r_0 is computed as 1, meaning the idealized correlation coefficient, which i denotes the time step. So, E_{pre} has a value of 0 when the prediction matches the observation, and it increases as the difference grows.

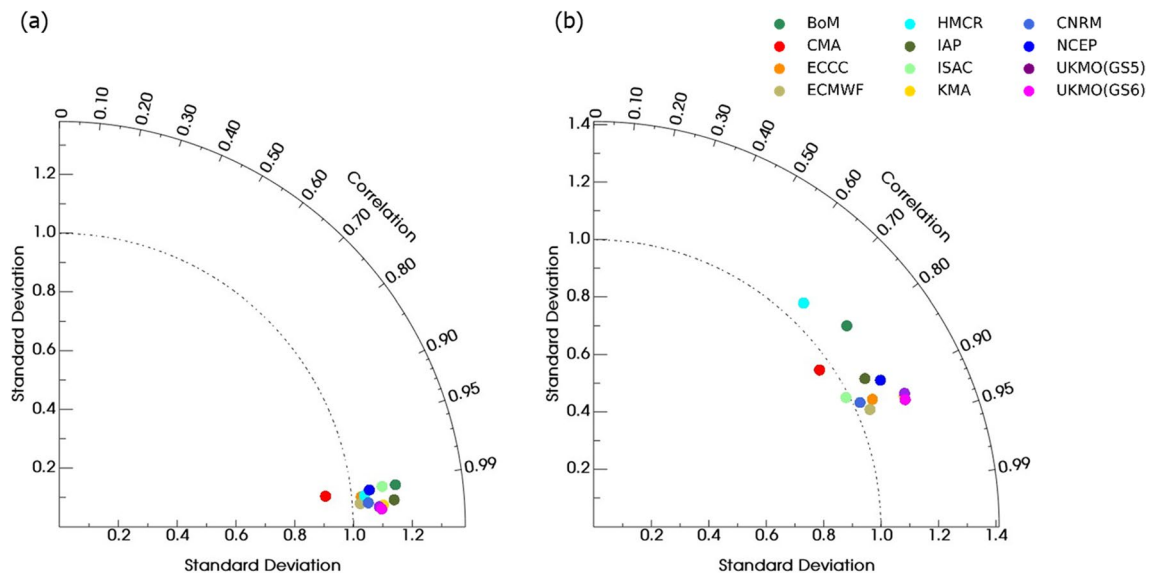


Fig. 1 Taylor diagram of the climatology for the global average 0-month forecast for temperature **a**, and precipitation **b**

$$E_{pre} = \frac{1}{N} \sum_{i=1}^N \log \left[\frac{\left(\frac{\sigma_{obs,i}}{\sigma_{pre,i}} + \frac{\sigma_{pre,i}}{\sigma_{obs,i}} \right)^2 (1 + r_0)^4}{(1 + r_i)^4} \right] \quad (4)$$

3 Results and discussion

3.1 Performance of the mean state of the S2S models

The S2S models demonstrate commendable performance in simulating climatology of precipitation and temperature. Figure 1 presents the global performance with Taylor diagrams, illustrating correlation coefficients and variances averaged for each region (Taylor 2001). It's noteworthy that precipitation in the tropics exhibits higher variance compared to other regions, potentially biasing the results. To mitigate this, the variances were normalized for each region as the ratio of the variance between predicted and observed and then calculated as a weighted average. Similarly, we calculated the correlation coefficient for each region and subsequently determined the weighted average for the global region. For temperature climatology (Fig. 1a, all S2S models exhibit correlation coefficients surpassing 0.99, indicating that all models simulate well. As for precipitation climatology, on the other hand, most models record correlations close to 0.9, indicating that differences in precipitation among models are somewhat more discernible compared to those in temperature. These results align with an earlier

study (Lee et al. 2010) on the pattern correlation coefficients of annual precipitation in tropical regions for seasonal prediction. Overall, the assessment indicates that solely relying on climatology to distinguish between models is formidable.

To include the performance of the seasonal cycle simulated by the S2S models, EOF analysis was applied to the monthly climatology. In doing so, four components were obtained, describing the seasonal cycle. These four components were proposed in a prior study and derived by applying the Fast Fourier Transform (FFT) and EOF techniques to the monthly climatology (Meyer et al. 2021). The two annual cycle modes are characterized by a single wave per year, followed by two semi-annual cycle modes that exhibit two waves per year (Hsu and Wallace 1976). The two annual cycle modes can be subdivided into a winter-summer pattern, peaking during summer and waning during winter, and a spring-fall pattern, reaching its peak in spring and declining in fall. The two semi-annual cycle modes can be further divided into the first and second semi-annual cycles. In this study, we employed EOF analysis to derive these four components, similar to Meyer et al. (2021). When utilizing EOFs for the monthly climatology, the first and second EOF modes correspond to the winter-summer and spring-fall patterns with one wave (Lee et al. 2010). Furthermore, the third and fourth EOF modes were indicative of the semi-annual modes 1 and 2, each exhibiting two waves per year, respectively. This is shown in Fig. S1, utilizing global monthly climatology by EOF analysis.

As a result, the model simulated seasonal cycle is now decomposed into two annual and two semi-annual modes with the PC time series and the corresponding spatial patterns. This process was repeated for all S2S models and

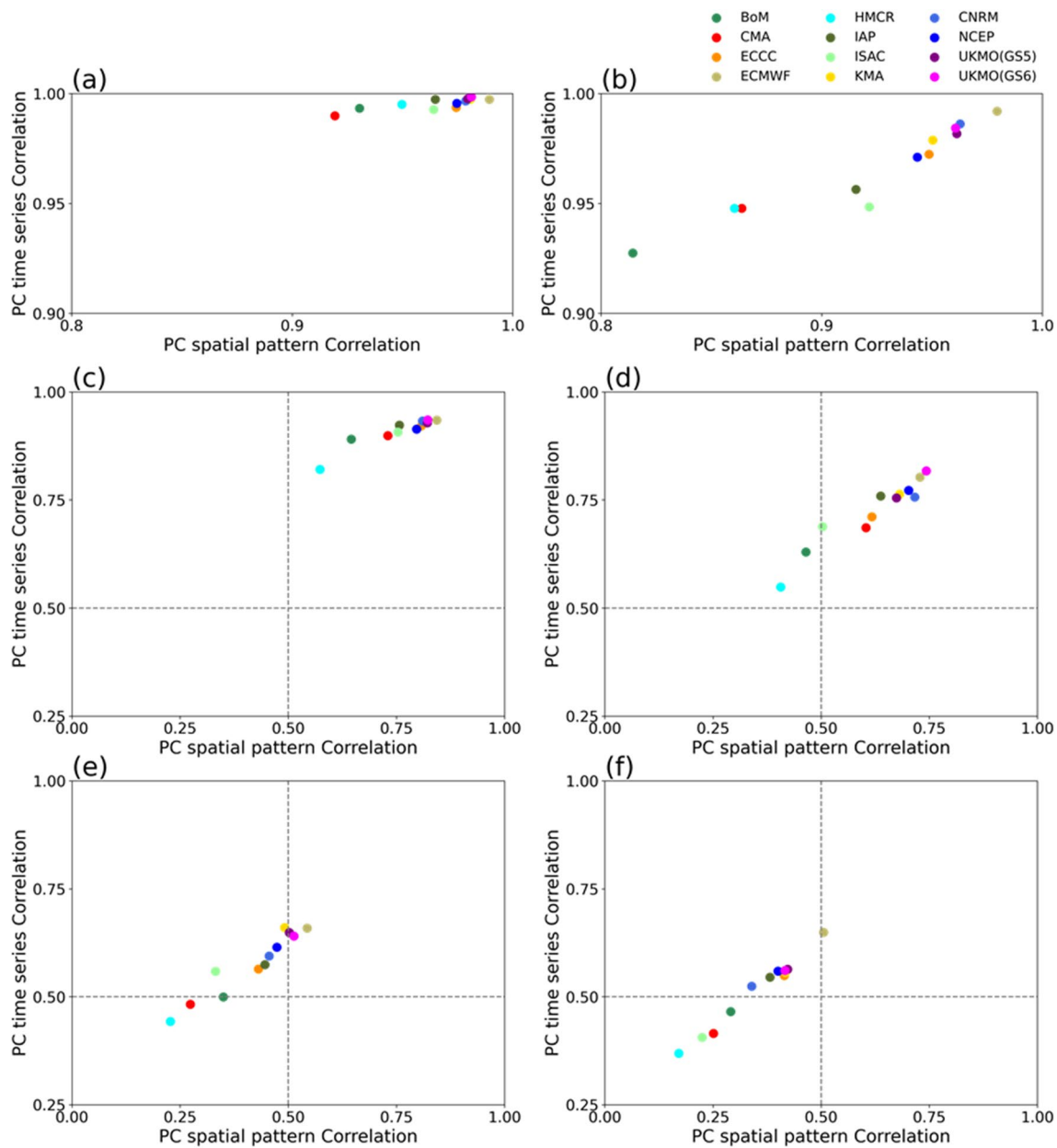


Fig. 2 Correlation between PC time series and PC spatial patterns of each mode in climatology by EOF analysis. **a** and **b** are the first and second modes of temperature, respectively, and **c** to **f** are the first to fourth modes of precipitation

reanalysis. The similarity between the model and reanalysis was evaluated with correlation coefficients. The scatter plot in Fig. 2 displays a positive relationship between the performance of spatial patterns and that of time series measured by the correlation between the reanalysis and each S2S model. It is noteworthy that the PC time series exhibits higher correlation coefficients than the PC spatial patterns. The first mode is distinguished as the most accurate depiction by almost all models, with diminishing coefficients for modes with lesser fractions. In the case

of temperature, the S2S models demonstrate high performance in simulating the seasonal cycle, with correlation coefficients surpassing 0.8, respectively, up to the second mode. Conversely, precipitation exhibits lower correlations compared to temperature, and the disparity among models is more conspicuous.

To evaluate the relationship between climatology and the seasonal cycle, two metrics, E_{cli} and E_{SC} , were computed and depicted in Fig. 3. These metrics were min-max normalized to ensure equal comparison and to scale the

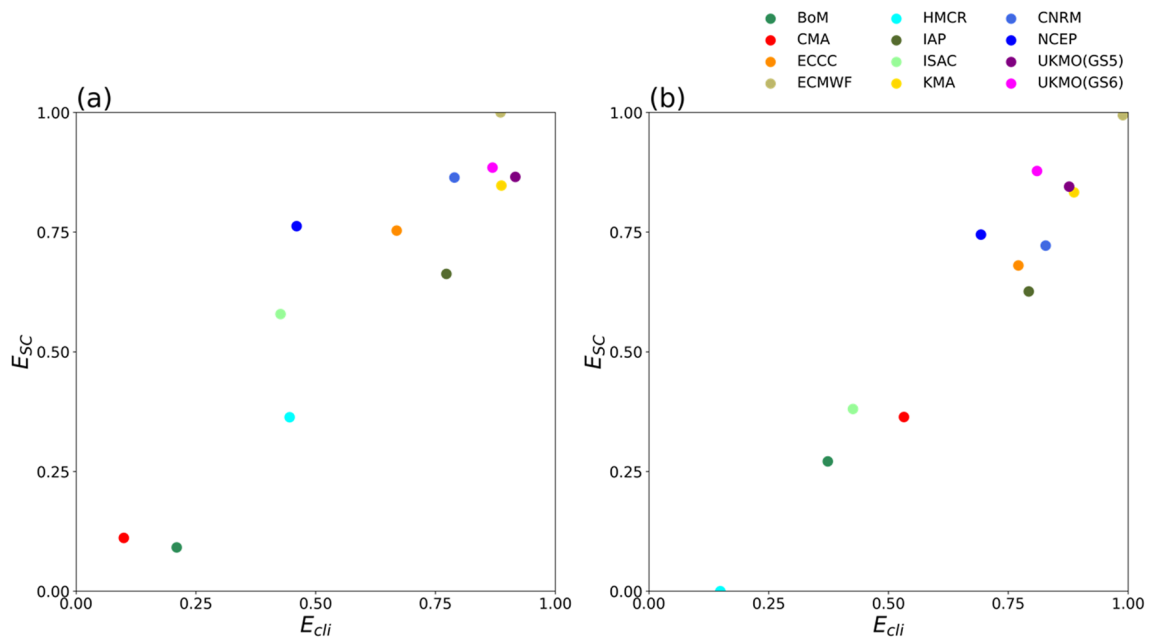


Fig. 3 Relationship between climatology and seasonal cycle. **a** is temperature, **b** is precipitation

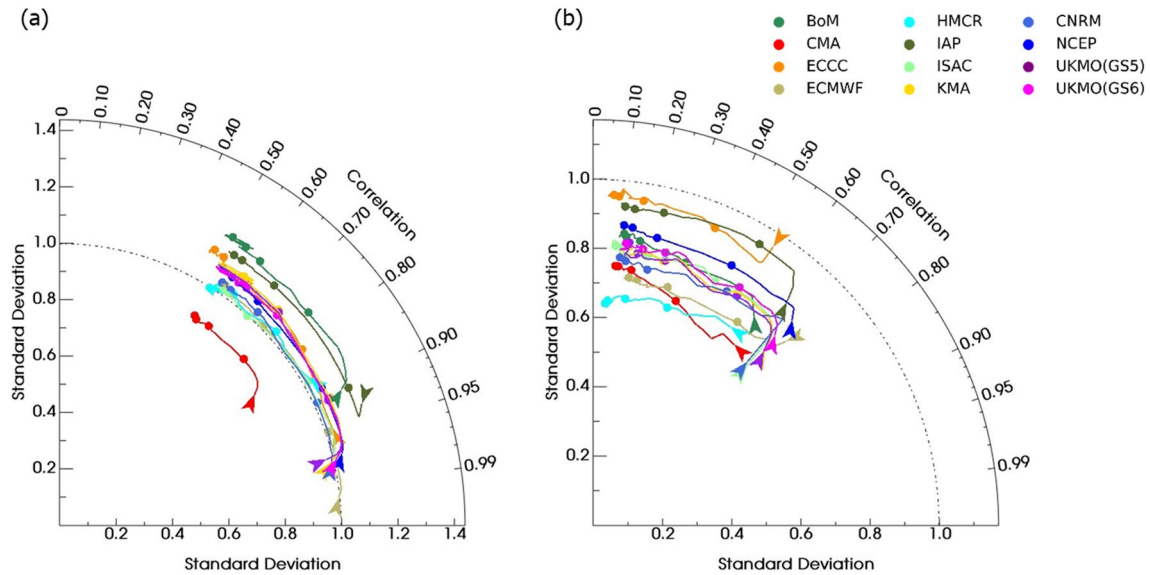


Fig. 4 Taylor diagram of the S2S models prediction for 1–32 forecast days. **a** is temperature, **b** is precipitation. Starting at the arrow, each mark represents a prediction in 7, 14, 21, and 28 days

original values within a range from 0 to 1. Although a linear relationship between the two metrics generally holds, deviation from a perfect linear relationship is found in the S2S models. This suggests that climatology and the seasonal cycle encompass distinct facets of the mean state simulation. As a result, our final metric (E_{MS}) was introduced, which incorporates both climatology and the annual cycle in assessing the performance of the mean state.

3.2 Weather and sub-seasonal prediction skill of S2S models

Figure 4 shows the prediction skill for each S2S model on daily base using Taylor diagrams, which include an arrow and four circles representing forecasts at a lead-time of 7, 14, 21, and 28 days. As the forecast lead time increases, the correlation coefficient decreases. However, after approximately

14 days, there is a tendency of convergence, as indicated by the close proximity or merging of the circles representing forecasts at lead time of 21 and 28 days. The variance is fairly stable as the lead time increases, except for the initial lead day. Generally, the prediction skill of temperature is robust, albeit some models have a propensity to overstate the variance. Conversely, the prediction skill of precipitation is consistently underestimated across the models.

Further analysis on the prediction skill of temperature was carried out by using correlation, nRMSE, and E_{pre} (Fig. S2). The analysis is segmented into the whole globe, mid-latitude, and tropics. Among the 36 divided regions, the first includes all 36, mid-latitude only includes 12 regions between 30 °N(°S) and 60 °N(°S), and tropics only includes 12 regions between 30 °S and 30 °N. On a global scale, prediction skill tends to stabilize at a certain value depending on the S2S models, with stabilization generally occurring around 15 days for all three metrics. Moreover, a noteworthy level of prediction skill is exhibited by 10 out of the 12 models, with correlation coefficients consistently exceeding 0.5 for forecast lead times of up to 31 days. Such elevated prediction skill has been demonstrated through probabilistic analyses to facilitate the prediction of temperature-based heatwaves up to four weeks in advance (Ardilouze et al. 2017; Vitart and Robertson 2018). Furthermore, it was observed that the predictive capabilities of the S2S models were extensively distributed in the tropics, as evident in the correlation coefficient, while in the mid-latitudes for the nRMSE. In contrast, the E_{pre} metric reveals a substantial disparity among the S2S models prior to convergence, but post-convergence, the values tend to be similar across the models.

Similarly, the prediction skill of precipitation was further analyzed in Fig. S3. On a global scale, the outcomes are somewhat parallel to those of temperature. However, the correlation coefficient for precipitation takes a dip, falling below 0.5, between 2 and 8 days after the forecast, depending on the model. This signifies that precipitation prediction skill is inferior to that of temperature. This finding is consistent with previous studies in which the prediction skill of precipitation starts to decrease from the second week onwards (de Andrade et al. 2019; Moron and Robertson 2021). On a regional level, the majority of the models exhibit superior prediction skill in the tropics compared to the mid-latitudes across all indicators, but there is a noticeable variation among the S2S models. Although the exact cause for such diverging result is unclear, it can be related to each model's performance on MJO, which is considered a crucial contributor to sub-seasonal prediction skill and the prevailing variability during winter (Li and Robertson 2015; Zhang and Dong 2004). Furthermore, while the mid-latitudes distinctly converged to a specific

value, the tropics exhibit a persistent decline in both correlation coefficients and nRMSE.

In light of our result here, it was decided to divide the forecast period into two regimes – a 14-day forecast and a sub-seasonal forecast, which encompasses 15 days onward, as different from conducting an analysis on a weekly basis. The weekly analysis was introduced by Li and Robertson (2015) and widely used to evaluate S2S prediction skill (Coelho et al. 2018; de Andrade et al. 2019). There is also a study that utilized a bi-weekly approach (Moron and Robertson 2021). Nonetheless, we decided that if the prediction skill of S2S models hinges on the feature of convergence, splitting the forecast period of 0–31 days into two regimes is more appropriate than dividing it into weeks. Secondly, we opted to employ E_{pre} as the metric for assessing prediction skill. This is based on the fact E_{pre} is a more stringent and discerning metric in comparison to the correlation coefficient and nRMSE, as it more clearly delineates the convergence of prediction skill, especially for precipitation.

E_{pre} was used to evaluate each model for prediction skill at the weather and the sub-seasonal regime, respectively. First and foremost, concerning temperature as shown in Table S1, spring emerged as the most reliable prediction skill globally. It also demonstrated modest prediction skill for summer in the tropics, and for fall and winter in the mid-latitudes, regardless of the time scale. Moving on to precipitation, as depicted in Table S2, winter stood out as the most dependable prediction skill globally across both weather and sub-seasonal regimes. Conversely, summer in the tropics proved to be less predictable across all forecast periods. Moreover, fall exhibited slightly diminished prediction skill at the weather regime, and spring at the sub-seasonal regime in mid-latitude regions.

3.3 Relationship between the performance of mean state and prediction skill

So far, we have conducted individual analyses of the mean state performance and prediction skill. Our attention now turns to investigating the relationship between the two aspects. To commence this exploration, we utilize simple mean state and prediction skill indicators in Figs. 5 and 6 for temperature and precipitation correspondingly. Figures 5 and 6 include four mean state metrics: the mean absolute error (MAE) of the annual mean, the correlation coefficient of the annual pattern, the RMSE of climatology, and the correlation coefficient of climatology. Additionally, two prediction skill metrics, RMSE and the correlation coefficient, are utilized. Additionally, Table S3 presents an overview of the mean state simulation performance across models, including the mean, standard deviation, count, maximum, and minimum values.

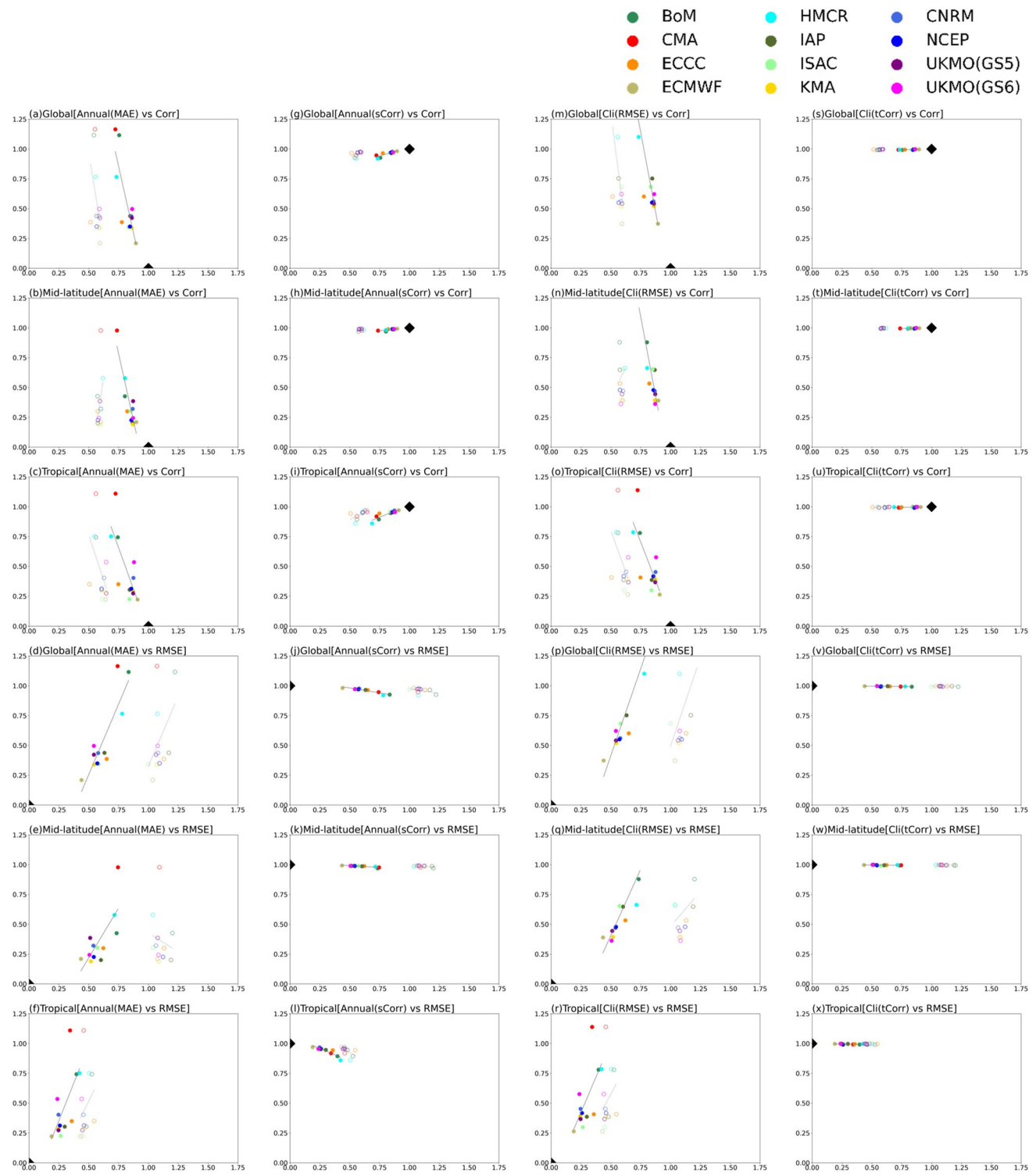


Fig. 5 Relationship between mean state performance and prediction skill in temperature. The black diamond is an ideal value. **a–f** is based on the annual mean absolute error for the mean state metric, **g–l** is based on the annual mean pattern error correlation, **m–r** is based on

RMSE of climatology, and **s–x** is based on correlation of climatology. **a–c**, **g–i**, **m–o**, and **s–u** are based on correlation correlations for the prediction skill metrics, and the rest are RMSE

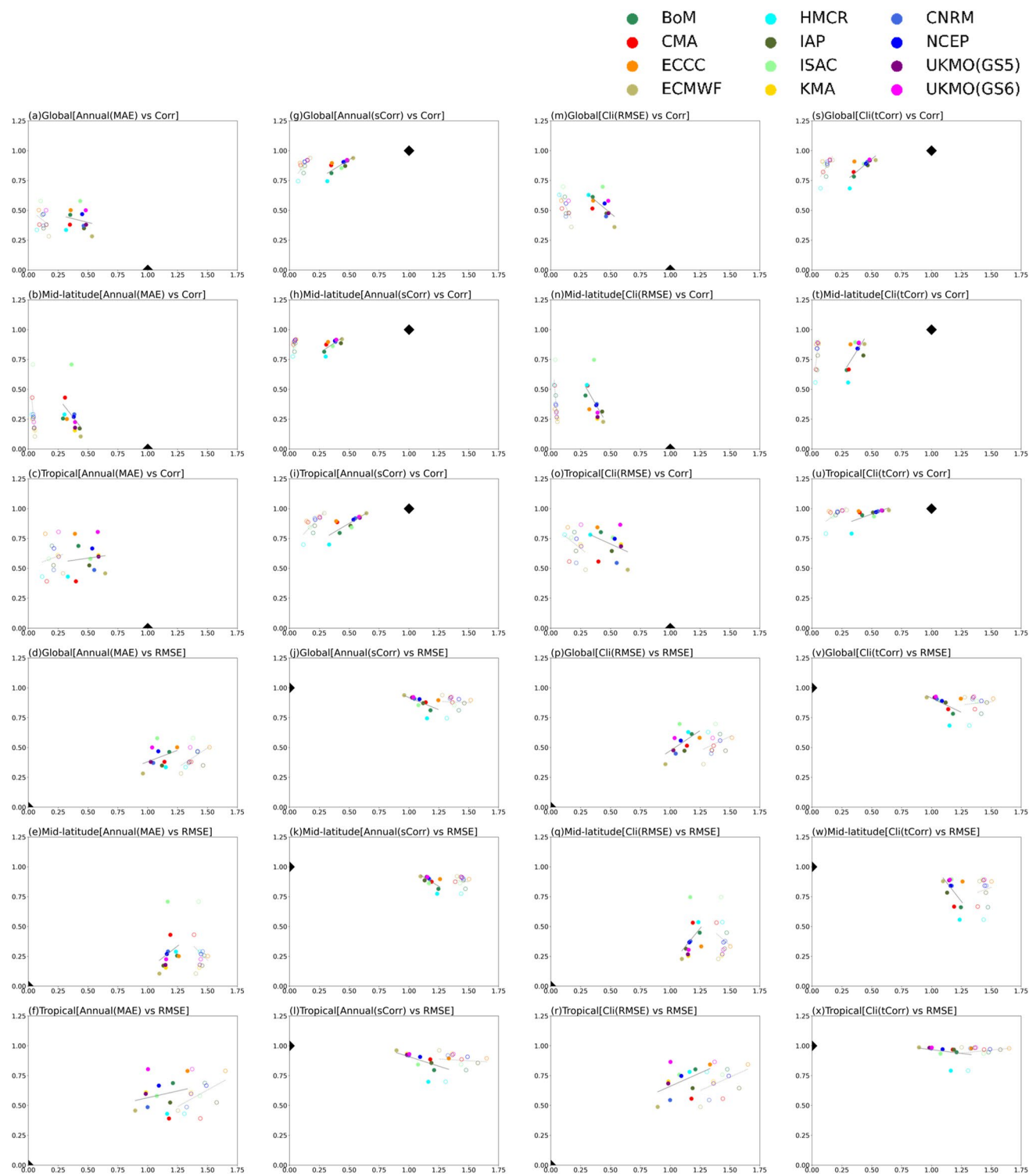


Fig. 6 Same as Fig. 5, but for the precipitation

Analyzing temperature reveals that the correlation coefficient indicates a mean state greater than 0.92, and the RMSE and MAE results demonstrate a clustering of the majority of models.

Precipitation, with a wider distribution than temperature, clusters in certain values. This implies that the S2S models have advanced to a level where differentiating their abilities to simulate the mean

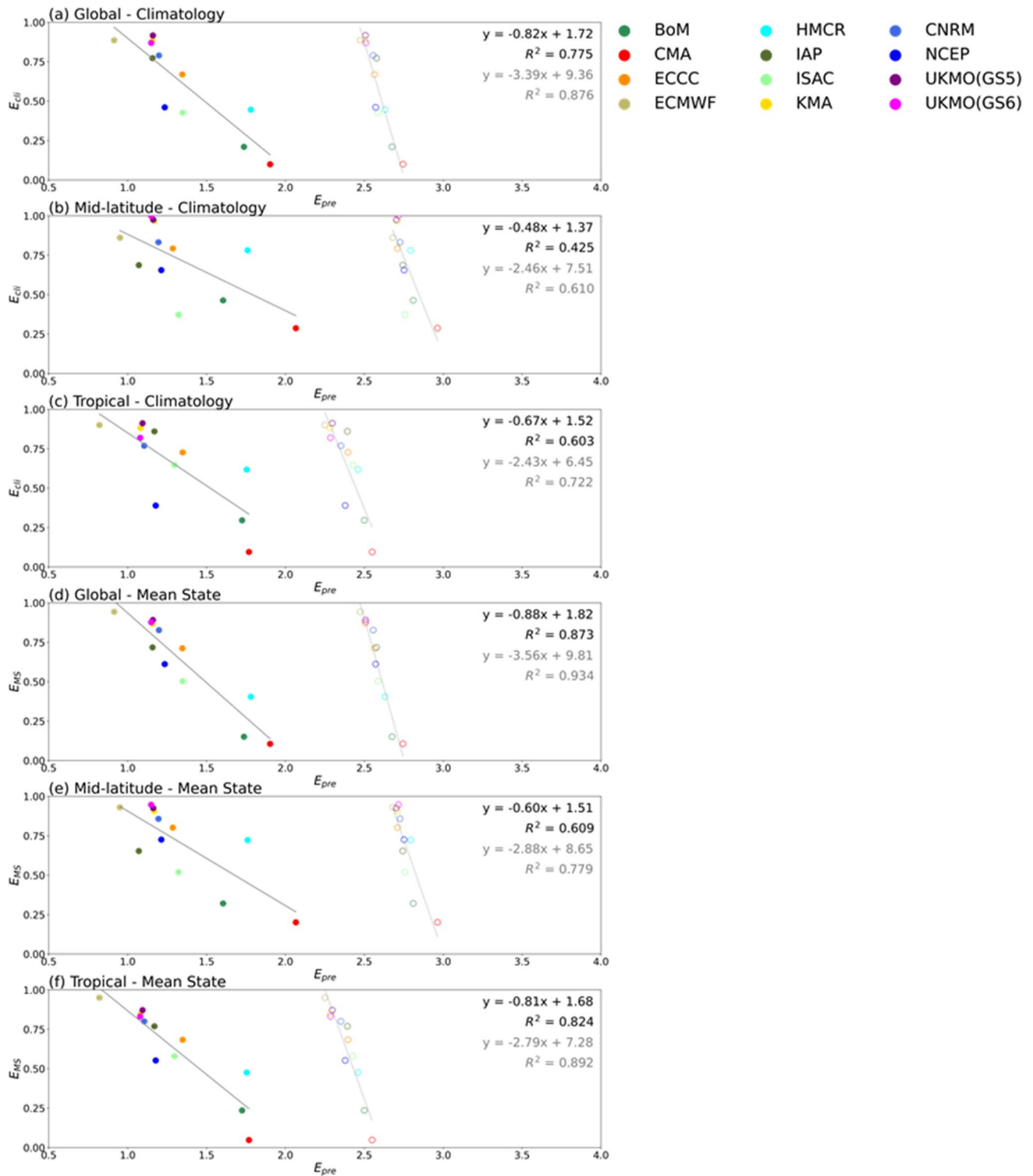


Fig. 7 Relationship between mean state performance and prediction skill in temperature. Solid lines, filled marks, and black equation and R^2 represent the weather scale, while dashed lines, non-scaled marks, and gray equation and R^2 represent the Sub-seasonal scale. **a–c** are

based on climatology, **d–f** considered both the climatology and seasonal cycles. **a** and **d** are global, **b** and **e** are mid-latitude regions, and **c** and **f** are tropical regions

state using simple metrics is increasingly difficult. As a result, this could obscure or present a false impression of the relationship between mean state performance and prediction skill. In a puzzling case, it was noted that prediction skill for sub-seasonal

precipitation forecasts in mid-latitude regions increased when there was a decrease in mean state performance.

Since then, the analysis has been centered on evaluating the performance of the mean state in the S2S models,

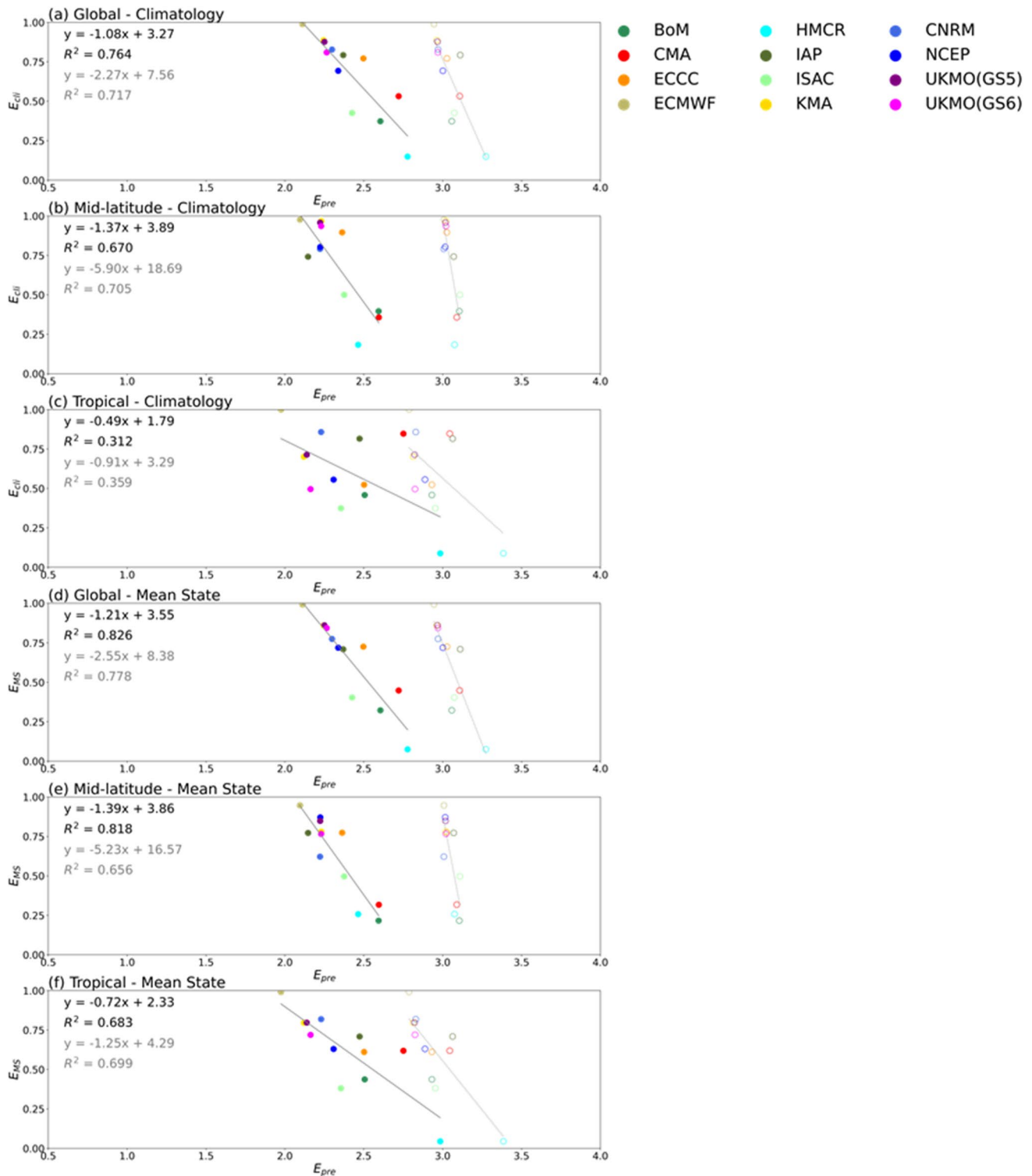


Fig. 8 Same as Fig. 7, but for the precipitation

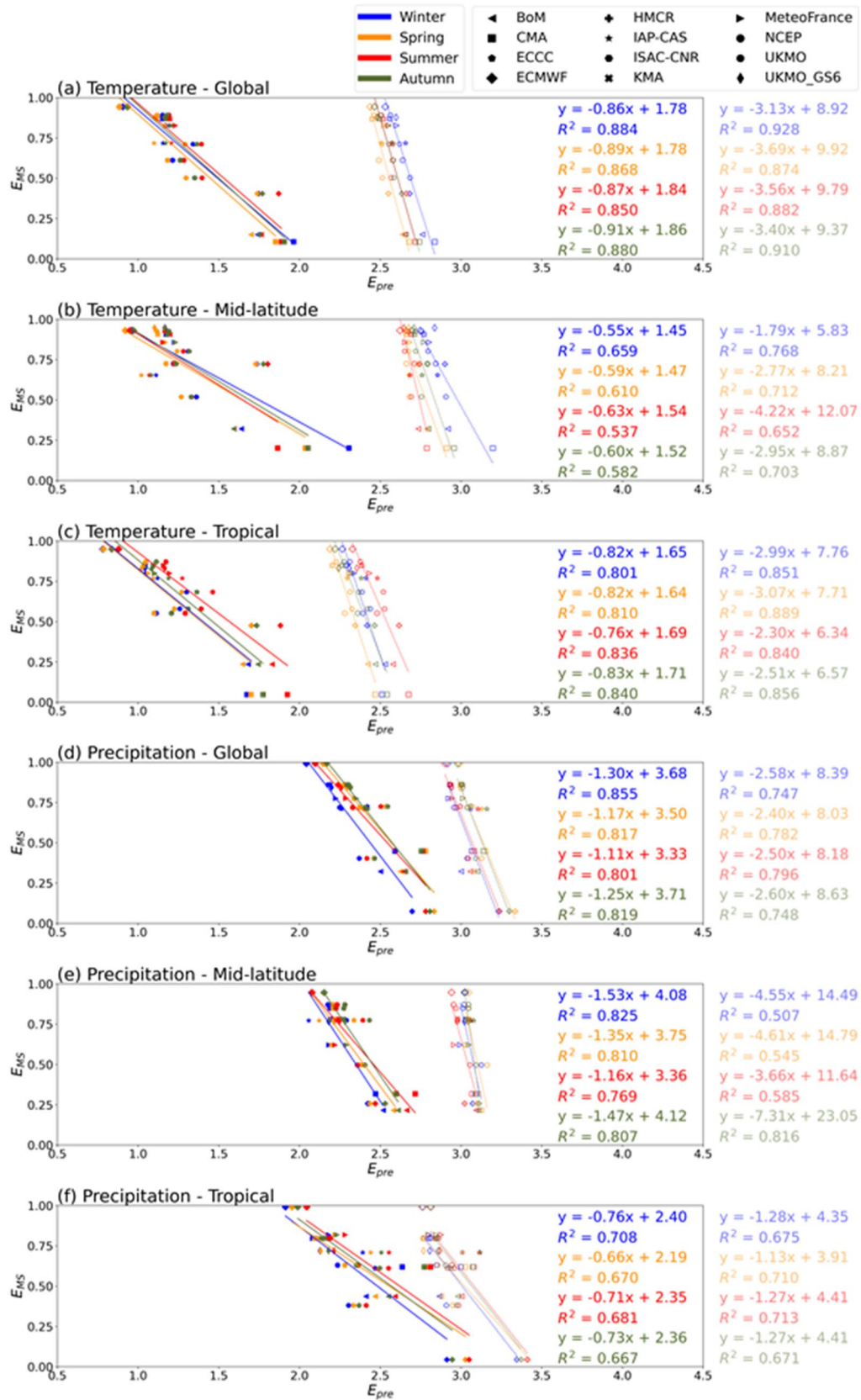


Fig. 9 Seasonal relationship between mean state performance and prediction skill. **a–c** is temperature, **d–f** is precipitation. **a** and **d** are global, **b** and **e** are mid-latitude regions, and **c** and **f** are tropical regions

with emphasis on climatology and the seasonal cycle as an important component of the mean state. Additionally, there has been a comparative analysis of three indicators for prediction skill: correlation coefficient, nRMSE, and E_{pre} . Here, the focus shifts to examining the relationship between the performance of the mean state and prediction skill. Further scrutiny is also applied to ascertain which metric of the mean state better captures this relationship: whether it is climatology alone or E_{MS} , which considers both climatology and the seasonal cycle.

Firstly, Fig. 7 displays the results for temperature. The filled marks with solid lines denote the weather regime, while the hollow markers with dashed lines represent the sub-seasonal regime. At the global level, as in Fig. 7a and d, a linear relation is observed for both metrics at both the weather and the sub-seasonal regimes. This linearity is more distinct at the weather regime, while the convergence of prediction skill in the S2S models becomes quite clear at the sub-seasonal regime, resulting in less noticeable differences among models. On a regional scale, a linear relationship is similarly evident across all cases. Moreover, the incorporation of both climatology and the seasonal cycle (E_{MS}) leads to higher R-square values for the trend line, which indicates an advantage in explaining the relationship between mean state and prediction skill.

Analogous to temperature, analysis results of precipitation are illustrated in Fig. 8, and it generally exhibits similar patterns, albeit with a few notable differences. Particularly in the tropics, the differences in prediction skill are more pronounced compared to temperature. In Fig. 8c, when solely utilizing climatology, the R-square values at the weather and sub-seasonal regimes are considerably low, standing at 0.312 and 0.359, respectively. Conversely, Fig. 8f, which employs E_{MS} by integrating both climatology and the seasonal cycle, reveals a substantial escalation in the R-square values, recording 0.683 and 0.699. Consequently, in analyzing the interrelationship between the mean state and prediction skill across various seasons, only E_{MS} was employed.

In order to discern seasonal variations, prediction skill was divided into four seasons as in Fig. 9. For mid-latitudes, the Southern Hemisphere experiences seasons opposite to those in the Northern Hemisphere, and hence the computations were adjusted accordingly. For instance, the term ‘summer’ in the mid-latitudes denotes June-July-August in the Northern Hemisphere, and December-January-February in the Southern Hemisphere. In contrast, the tropics do not exhibit pronounced seasonality, so no distinction was made between the Southern and Northern Hemispheres, and instead the seasons of the Northern Hemisphere were adopted. As such, ‘summer’ in the tropics refers to June-July-August in both hemispheres. When examined in terms of R-square values, winter emerges as the season with the

highest values on the weather scale, except for temperature in the tropical region. On the other hand, it is difficult to find a clear seasonal feature for sub-seasonal forecasts. The relatively robust connection observed between the performance of mean state and prediction skill during winter can likely be attributed to the MJO, as previously mentioned. Notably, it has been reported that MJO exhibits greater predictive potential during the winter season compared to other seasons (Liu et al. 2017).

Our analysis initially focused on comparing the simulation performance of the mean state across various models. We then proceeded with a within-model comparison for each S2S model, applying min-max normalization across 36 regions. The outcomes are presented in Figs. S4 and S5 for temperature and precipitation, respectively. For temperature, we did not find a clear relationship, which we speculate may be due to the models reaching a saturation point in their ability to simulate the mean state. In contrast, for precipitation, we discovered a correlation between mean state simulation and predictability, likely due to the considerable variability in mean state simulation performance across regions.

Improving the mean state in climate models has a significant impact on forecast predictability (Lee et al. 2010). Specifically, enhancing the mean state leads to improved simulation accuracy of the MJO and ENSO, a critical pattern of climate variability that influences the global climate system (Bayr et al. 2018; Kang et al. 2020). The improvement of the MJO and ENSO representation plays a crucial role in the enhancement of bias correction, hydroclimate, and thermocline within the models (Kim et al. 2014, 2017; Lim et al. 2018). These improvements ensure that the models more accurately reflect the process and mechanisms of the actual climate system, thereby comprehensively enhancing the predictive skill for weather and seasonal climate variability. Therefore, conducting a deeper analysis and understanding of how the mean state affects the sub-seasonal forecast skill is a critical step towards increasing the accuracy of climate modeling and predictions.

4 Conclusions

We investigated the relationship between the performance of mean state simulation and the prediction skill of anomalies in weather and sub-seasonal forecasting. We concluded that a linear relationship is present in the mean state and prediction skill in weather and sub-seasonal forecasting. When the seasonal cycle is considered, this linear relationship is more distinctly exhibited compared to climatology. Nevertheless, in sub-seasonal forecasting, the disparity in prediction skill among models is insignificant, meaning that the variability in anomaly prediction skill relative to the performance of

mean state simulation is not as pronounced as it is on the weather timescale. The interrelationship between the accuracy of mean state simulation and prediction skill has been studied from various angles. For instance, an earlier study (Richter et al. 2018) using atmospheric model intercomparison project models indicated that the RMSE has a relationship with the mean state, whereas the ACC does not. In the context of seasonal forecasts, it has been demonstrated that both the annual mean and annual cycle have a positive correlation with precipitation anomalies (Lee et al. 2010). Additionally, at the subregional level, the ability to simulate the tropical Pacific cold tongue is linked to mean state bias (Ding et al. 2020).

The results of evaluating the mean field and prediction skill respectively are as follows: first, the performance of the mean state was undertaken from two perspectives: climatology and the seasonal cycle. The S2S model exhibits strong performance in replicating the climatology of temperature, with minimal divergence among the models. For precipitation, on the other hand, there is a more marked variation among the models compared to temperature. Secondly, prediction skill was examined utilizing the correlation coefficient, nRMSE, and E_{pre} . The S2S model demonstrates a higher prediction accuracy for temperature as opposed to precipitation across all the metrics assessed. A common characteristic of both temperature and precipitation is that their prediction skill appears to coalesce at a forecast lead time of roughly 15 days. This finding aligns with previous research that indicates a limit of prediction skill of 3–4 weeks compared to that of 1–2 weeks (de Andrade et al. 2019, 2021).

Forecasting within the sub-seasonal timescale continues to be a formidable task, and endeavors must be directed toward enhancing prediction skill. In the course of model development, the mean state is constantly evaluated in terms of rather simpler metrics focusing on how close it is to observation. Nonetheless, the findings of this study strongly imply that the mean state in the S2S models holds potential as a prime indicator of prediction skill, and that improving the simulation of the mean state could contribute to augmenting prediction skill, especially when it is more comprehensively measured, such as including the seasonal cycle.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00382-024-07315-x>.

Acknowledgements This research is funded by the National Research Foundation of Korea.

Author contributions The study was conceptualized by Jihun Ryu, and Jin-Ho Yoon. Jihun Ryu has done the data analysis, visualization, and writing the original draft. Reviewing and editing the manuscript is done by Shih-Yu (Simon) Wang, Jee-Hoon Jeong, and Jin-Ho Yoon.

Funding This research is funded by the National Research Foundation of Korea under NRF-2021R1A2C1011827 and the Brain Pool program under RS-2023-00283239, the Korean Meteorological Agency under the grant KMI2018-07010, and the GIST Research project funded by the GIST in 2024. S.-Y. Wang acknowledges the funding from U.S. Department of Energy/Office of Science under Award Number DE-SC0016605 and the U.S. SERDP project RC20-3056.

Data availability The S2S datasets can be downloaded from <https://apps.ecmwf.int/datasets/data/s2s/levtype=sfc/type=cf/>. The GPCP precipitation can be downloaded from <https://www.ncei.noaa.gov/products/climate-data-records/precipitation-gpcp-daily>. The ERA5 temperature can be downloaded from <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=form>.

Declarations

Competing interests The authors declare no competing interests.

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

- Ardilouze C, Batté L, Déqué M (2017) Subseasonal-to-seasonal (S2S) forecasts with CNRM-CM: a case study on the July 2015 West-European heat wave. *Adv Sci Res* 14:115–121
- Bayr T, Latif M, Dommenges D, Wengel C, Harlaß J, Park W (2018) Mean-state dependence of ENSO atmospheric feedbacks in climate models. *Clim Dyn* 50:3171–3194
- Coelho CA, Firpo M, de Andrade FM (2018) A verification framework for south American sub-seasonal precipitation predictions. *Meteorol Z* 27:503–520
- de Andrade FM, Coelho CA, Cavalcanti IF (2019) Global precipitation hindcast quality assessment of the Subseasonal to Seasonal (S2S) prediction project models. *Clim Dyn* 52:5451–5475
- de Andrade FM, Young MP, MacLeod D, Hiron LC, Woolnough SJ, Black E (2021) Subseasonal precipitation prediction for Africa: Forecast evaluation and sources of predictability. *Weather Forecast* 36:265–284
- Ding H, Keenlyside N, Latif M, Park W, Wahl S (2015) The impact of mean state errors on equatorial Atlantic interannual variability in a climate model. *J Geophys Res* 120:1133–1151
- Ding H, Newman M, Alexander MA, Wittenberg AT (2020) Relating CMIP5 model biases to seasonal forecast skill in the tropical Pacific. *Geophys Res Lett* 47:e2019GL086765
- Hersbach H et al (2020) The ERA5 global reanalysis. *Q J R Meteorol Soc* 146:1999–2049
- Hsu C-PF, Wallace JM (1976) The global distribution of the annual and semiannual cycles in precipitation. *Mon Weather Rev* 104:1093–1101

- Huffman GJ et al (2001) Global precipitation at one-degree daily resolution from multisatellite observations. *J Hydrometeorol* 2:36–50
- Kang D, Kim D, Ahn MS, Neale R, Lee J, Gleckler PJ (2020) The role of the mean state on MJO simulation in CESM2 ensemble simulation. *Geophys Res Lett* 47:e2020GL089824
- Kim ST, Cai W, Jin F-F, Yu J-Y (2014) ENSO stability in coupled climate models and its association with mean state. *Clim Dyn* 42:3313–3321
- Kim ST, Jeong H-I, Jin F-F (2017) Mean bias in seasonal forecast model and ENSO prediction error. *Sci Rep* 7:6029
- Kim H, Vitart F, Waliser DE (2018) Prediction of the Madden–Julian oscillation: a review. *J Clim* 31:9425–9443
- Kim H, Richter JH, Martin Z (2019) Insignificant QBO–MJO prediction skill relationship in the SubX and S2S subseasonal reforecasts. *J Geophys Res: Atmos* 124:12655–12666
- Kim H, Son SW, Yoo C (2020) QBO modulation of the MJO-related precipitation in East Asia. *J Geophys Res: Atmos* 125:e2019JD031929
- Kumi N, Abiodun BJ, Adefisan EA (2020) Performance evaluation of a subseasonal to seasonal model in predicting rainfall onset over West Africa. *Earth Space Sci* 7:e2019EA000928-T
- Lee J-Y et al (2010) How are seasonal prediction skills related to models’ performance on mean state and annual cycle? *Clim Dyn* 35:267–283
- Li S, Robertson AW (2015) Evaluation of submonthly precipitation forecast skill from global ensemble prediction systems. *Mon Weather Rev* 143:2871–2889
- Lim Y, Son S-W, Kim D (2018) MJO prediction skill of the subseasonal-to-seasonal prediction models. *J Clim* 31:4075–4094
- Liu X et al (2017) MJO prediction using the sub-seasonal to seasonal forecast model of Beijing Climate Center. *Clim Dyn* 48:3283–3307
- Meyer JDD, Wang SYS, Gillies RR, Yoon JH (2021) Evaluating NA-CORDEX historical performance and future change of western US precipitation patterns and modes of variability. *Int J Climatol* 41:4509–4532
- Moron V, Robertson AW (2021) Relationships between subseasonal-to-seasonal predictability and spatial scales in tropical rainfall. *Int J Climatol* 41:5596–5624
- Richter I, Doi T, Behera SK, Keenlyside N (2018) On the link between mean state biases and prediction skill in the tropics: an atmospheric perspective. *Clim Dyn* 50:3355–3374
- Robertson AW, Kumar A, Peña M, Vitart F (2015) Improving and promoting subseasonal to seasonal prediction. *Bull Am Meteorol Soc* 96:ES49–ES53
- Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res: Atmos* 106:7183–7192
- Vitart F et al (2017) The subseasonal to seasonal (S2S) prediction project database. *Bull Am Meteorol Soc* 98:163–173
- Vitart F, Robertson AW (2018) The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. *NPJ Clim Atmos Sci* 1:3
- Vitart F, Robertson AW, Anderson DL (2012) Subseasonal to seasonal prediction project: bridging the gap between weather and climate. *Bull World Meteorol Organ* 61:23
- Wang L et al (2021) Multiple metrics informed projections of future precipitation in China. *Geophys Res Lett* 48:e2021GL093810
- Yang B et al (2013) Uncertainty quantification and parameter tuning in the CAM5 Zhang–McFarlane convection scheme and impact of improved convection on the global circulation and climate. *J Geophys Res: Atmos* 118:395–415
- Zhang C, Dong M (2004) Seasonality in the Madden–Julian oscillation. *J Clim* 17:3169–3180
- Zhu H, Wheeler MC, Sobel AH, Hudson D (2014) Seamless precipitation prediction skill in the tropics and extratropics from a global model. *Mon Weather Rev* 142:1556–1569

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