





OPEN

DATA DESCRIPTOR

# APEC climate center multi-model ensemble dataset for seasonal climate prediction

Yoo-Bin Yhang  , Chang-Mook Lim & Daeun Jeong

The multi-model ensemble (MME) technique is useful for obtaining reliable climate information. The Asia-Pacific Economic Cooperation Climate Center (APCC) has provided well-constructed MME data and individual model data included in the MME. This study examines the prediction skill of the APCC MME, focusing on 2-m temperature, precipitation, and sea surface temperature. For the overall evaluation of the APCC MME, the hindcast skill was assessed for the period of 1991–2010 for four sets of APCC MME representing the operational model suites from 2019 to 2022. The forecast skill was also evaluated to demonstrate actual predictability from 2019 to 2022. Hindcast validation showed a modest improvement, indicating continuous enhancement. In addition, it was demonstrated that the forecast skill was consistently maintained. These seasonal forecast data can provide valuable insights for decision-making across various sectors, helping mitigate risks and optimize resource management.

## Background & Summary

The Asia-Pacific Economic Cooperation Climate Center (APCC) has provided high-quality climate information and data since 2007. The APCC seasonal forecast is based on multi-model ensemble (MME) prediction and disseminated to APEC members every month. Currently, 15 operational centers and research institutes from 11 countries participate in the APCC MME, routinely providing predictions of global forecast fields (<https://www.apcc21.org/prediction/global/outlook>).

The MME technique has become a well-established method for improving forecast skill beyond that of a single model, as model diversity compensates for the uncertainties in individual models<sup>1–4</sup>. This skill improvement has been demonstrated in various studies using several MME data such as the APCC<sup>5–7</sup>, the Copernicus Climate Change Service (C3S), the North American Multi-Model Ensemble (NMME)<sup>8</sup>, and the World Meteorological Organization (WMO) Lead Centre for Long-Range Forecast Multi-Model Ensemble (LC-LRFMME)<sup>9</sup>. Alessandri *et al.*<sup>10</sup> assess all potential combinations of 16 seasonal prediction models from two MMEs and confirmed that the skill tends to increase due to the independence of the constituent models<sup>11</sup>.

Seasonal forecasts of atmospheric variables, such as near-surface temperature and precipitation, are useful in different socioeconomic sectors and risk situation, including hydropower and wind energy production<sup>12,13</sup>, management of water resource<sup>14</sup>, fire risk, agriculture, and hazardous weather which can cause serious economic damage<sup>15</sup>. In all these cases, early and reliable climate information can be a prerequisite for providing well-defined economic value<sup>16</sup>. As the first step toward this, the predictability of key climate variables (e.g., temperature and precipitation) needs to be validated.

Previous assessments of the APCC MME demonstrated skills for deterministic and probabilistic MME in terms of hindcast and real-time forecast<sup>6,7</sup>. However, the number of participating models has increased and many models have been upgraded or replaced in the recent years. Therefore, it is necessary to introduce the current composition of the APCC MME and verify its predictability. This study aims to examine the overall performance of the APCC MME through hindcast validation, and to provide users more direct information by validating the real-time forecast data used in actual forecasts. Quantifying forecast quality allows users to become aware of the current limitation in seasonal forecasting across various parts of the world and allows forecast producer to recognize the limitations of their products.

In this study, we investigated the changes in the hindcast skills of the APCC MME in four recent sets for the period 1991–2010. We assessed the relative prediction skill for the seasonal mean of 2-m temperature, precipitation at 1-month lead, and sea surface temperature at 1-month and 4-month lead. The verification of hindcasts

APCC Climate Center, Busan, Korea. ✉e-mail: [ybyhang@apcc21.org](mailto:ybyhang@apcc21.org)

can provide general knowledge on the skill of the prediction system, which progressed in the last few years. However, diagnoses of hindcasts may not be consistent with the skill of operational real-time forecasts due to several reasons (e.g., different configurations, periods of forecasts), which have strong effects on the seasonal forecast skill<sup>17–20</sup>. Therefore, both the skills of real-time forecast and hindcast were analysed for 2-m temperature, precipitation, and sea surface temperature from 2019 to 2022.

The skill of the APCC MME varies depending on the region and season. The prediction skill in terms of correlation is higher in the tropics and boreal winter than in the extratropic and boreal summer for both temperature and precipitation because the El-Niño-Southern Oscillation (ENSO) variability was one of the main sources of predictability. The skill improvement achieved by the MME technique is higher for precipitation than for temperature, which could be explained by the higher independence of individual model errors in precipitation forecasts.

Despite continuous advancements in climate models, the predictability in mid- and high-latitudes is still challenging due to high atmospheric variability and less impact of strong predictability sources (e.g., ENSO). Nevertheless, the MME approach in seasonal forecasting is a great step forward as compared with a single-model one. For regions with appropriate predictability, the use of MME forecasts provides a more reliable basis for decision-making than single-model forecasts, and the advantage of MME approach is clear for these regions. Meanwhile, for regions with low predictability, MME forecasts prevent us from making adverse decisions, which is a definite advantage; however, support for correct decisions based on them is relatively limited.

This study demonstrates that the APCC MME can provide important seasonal prediction information regardless of whether the region has high or low predictability. These results are expected to be useful for other regions not mentioned in this study. In addition, other combinations of different models are expected to provide more useful information than the use of a single model.

## Methods

**Multi-model ensemble.** The APCC issues monthly-mean and 3-month mean (seasonal) MME forecasts with a 1-month and 4-month lead time every month. The “1-month lead” seasonal forecast means the forecast for the first 3-month-average following the initialization month. For example, the 1-month lead forecast from January is February–April (FMA). The 4-month lead forecast from January is for the target period of May–July (MJJ). The APCC MME predictions are based on outputs from climate prediction models from 15 operational centers and research institutes (Table 1). Noted that the set of models participating in real-time MME forecasts may slightly vary from one forecast to another according to the operational situation at the time of issuing the given forecast. Generally, the MME skills increase as the number of models increases because the reduction of systematic and random errors is more effective. Therefore, fluctuations in the number of participating models can be a potential issue in obtaining reliable predictions. Typically, the skill of the MME improves rapidly with the addition of the first few models, often reaching a significant portion of the maximum potential skill with approximately 5–10 models. Beyond approximately 10–15 models, the rate of improvement in skills tends to diminish, indicating skill saturation<sup>21,22</sup>. The number of participating models in the APCC MME has remained at approximately 10 or more over the past four years, which is considered sufficient to maintain the reliability of the prediction. This consistency ensures the robustness and accuracy of the ensemble predictions. Individual model configurations may also change over time because of the implementation of new versions of the model. The history of the individual models is reported on the APCC website (<https://apcc21.org/prediction/global/model?lang=en>).

The model data were interpolated to a common resolution of a  $2.5^\circ \times 2.5^\circ$  grid and used for multi-model combination. The ensemble mean anomalies of the individual models were calculated by subtracting the model's own climatology during 1991–2010 to correct each model's systematic error in the mean. This method, which is used to produce real-time deterministic APCC MME predictions, does not correct systematic errors in the variances or other moments.

$$F'_i = F_{i,t} - \bar{F}_i \quad (1)$$

The APCC MME adopts a simple composite of the ensemble mean anomalies from each model as the official APCC seasonal prediction<sup>6</sup>.

$$\text{MME} = \frac{1}{N} \sum_{i=1}^N (F'_{i,t}) \quad (2)$$

where  $F'_i$  is a model's anomaly of the  $i$ th model at the forecast time  $t$ .  $F_{i,t}$  and  $\bar{F}_i$  are the forecast and climatology of the  $i$ th model, respectively.  $N$  is the number of individual models involved. In this method, each of the  $N$  participating models has the same weight of  $1/N$  at all grid points.

While this simple composite MME offers several advantages in terms of error reduction, robustness, and ease of implementation, it also has notable limitations such as bias retention, and the inability to account for quantiles. To overcome this limitation, more complex methods for optimally combining the models and calibrations have been described<sup>22–25</sup>. Calibration requires long hindcast data sets to examine the model's characteristics and a thorough assessment to select the areas and variables where calibration and/or combination are beneficial<sup>1</sup>. Therefore, a simple composite MME could be a practical way to enhance the predictive skill and manage computational resources effectively.

**Hindcast set.** In this study, four hindcast sets for four recent years were relatively compared. These four sets represent the most recent data that can be analyzed for the common hindcast period. H2019, which was active in

| Organization, Country  | System  | Years active                           |
|--|---|--|
| APEC Climate Center (APCC), Korea                            | SCoPS <sup>39</sup>                               | 2017–present                           |
| Beijing Climate Center (BCC), China                          | CSM1.1m <sup>40</sup>                             | 2016–present                           |
| Bureau of Meteorology (BOM), Australia                       | ACCESS-S1<br>ACCESS-S2 <sup>41</sup>              | 2019–2021<br>2022–present              |
| Centro Euro-Mediterraneo sui Cambiamenti (CMCC), Italy       | CMCC-SPS3<br>CMCC-SPS3.5 <sup>42</sup>            | 2020<br>2021–present                   |
| Central Weather Administration (CWA), Chinese Taipei         | GFST119L40<br>TCWB1Tv1.1 <sup>43</sup>            | 2017–2020<br>2020–present              |
| Environment and Climate Change Canada (ECCC), Canada         | CanSIPS<br>CanSIPsv2<br>CanSIPsv2.1 <sup>44</sup> | 2015–2019<br>2019–2021<br>2022–present |
| Hydrometeorological Research Center for Russia (HMC), Russia | SL-AV <sup>45</sup>                               | 2012–present                           |
| Japan Meteorological Agency (JMA), Japan                     | JMA/MRI-CPS2<br>JMA/MRI-CPS3 <sup>46</sup>        | 2015–2022<br>2022–present              |
| Korea Meteorological Administration (KMA), Korea             | GloSea5GC2<br>GloSea6GC3.2 <sup>47</sup>          | 2016–2022<br>2022–present              |
| Metero-France, France  | MF Sys7<br>MF Sys8 <sup>48</sup>                  | 2021–2021<br>2021–present              |
| Main Geophysical Observatory of Russia (MGO), Russia         | MGOAM-2 <sup>49</sup>                             | 2006–present                           |
| National Aeronautics and Space Administration (NASA), USA    | GEOS-S2S-2.1 <sup>50</sup>                        | 2018–present                           |
| National Center for Environmental Prediction (NCEP), USA     | CFSv2 <sup>51</sup>                               | 2011–present                           |
| Pusan National University (PNU), Korea                       | PNU CGCMv1.1<br>PNU CGCMv2.0 <sup>52</sup>        | 2010–2020<br>2020–present              |
| United Kingdom Met Office (UKMO), UK                         | GloSea5<br>GloSea6 <sup>53</sup>                  | 2016–2021<br>2021–present              |

**Table 1.** Participated models in APCC MME. The detailed information of models participating in APCC MME is found in <https://apcc21.org/prediction/global/model?lang=en>.

2019, included eight models: APCC, BOM, CWA, ECCC, JMA, NASA, NCEP, and PNU. H2020 consists of the APCC, BCC, BOM, CMCC, CWA, ECCC, HMC, JMA, KMA, NASA, NCEP, PNU, and UKMO, which were used in 2020. H2021, active in 2021, has 14 models: APCC, BCC, BOM, CMCC, CWA, ECCC, HMC, JMA, KMA, METFR, NASA, NCEP, PNU, and UKMO. The most recent set, H2022, includes the same models as H2021 except for individual model's version. The predictabilities for four recent hindcast sets were evaluated for the period of 1991–2010.

**Real-time forecast set.** The same models included in each hindcast set were used for the real-time forecast. The skill of real-time forecast was verified from 2019 to 2022.

**Verification metrics.** The latitude-weighted area-aggregated anomaly pattern correlation coefficient (ACC) is used as a measure of prediction skill for this analysis<sup>26,27</sup>. The anomaly correlation describes the strength of the linear relationship between the predicted and observed anomalies, and is widely used in seasonal prediction verification. The ACC can be calculated as:

$$ACC = \frac{\sum(M - \bar{M})(O - \bar{O})}{\sqrt{\sum(M - \bar{M})^2} \sqrt{\sum(O - \bar{O})^2}}$$

where  $M$  is the prediction, and  $O$  is the observation. Anomalies are calculated by subtracting climatology  $\bar{M}$  and  $\bar{O}$ , and weighted by latitude as multiplying the weighting factor of  $\cos\left(\text{latitude} \times \frac{\pi}{180}\right)$ . An ACC value of 1 indicates a perfect association of forecast anomalies with actual anomalies, and 0 refers to random forecasts. For the convenience of comparison, we also calculated the time-averaged anomaly pattern correlation coefficient over the globe and its subregions.

Temporal relationship is measured using the Pearson correlation coefficient. The Pearson correlation coefficient is the most common method used to measure linear correlations. It is a number between  $-1$  and  $1$  that measures the strength and direction of the relationship between two variables. If the temporal correlation coefficient (TCC) is  $-1$ , it indicates a strong negative relationship, while correlation coefficient of  $1$  implies a strong positive relationship.

Because the correlation coefficients are not sensitive to bias, they do not guarantee accuracy in terms of their values. Therefore, the root mean square error (RMSE) was used to measure average of error. It represents the square root of the average of the squared difference between the predicted and observed values. The RMSE puts greater influence on large errors than on smaller errors, which may be useful if large errors are especially undesirable, but may also encourage conservative forecast. The RMSE can be calculated as

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (M_i' - O_i')^2}$$

where,  $M'$  is the latitude-weighted anomaly of prediction,  $O'$  is the latitude-weighted anomaly of observation, and  $N$  is the number of predictions. A low RMSE indicates that the forecasts are close to the observed values, indicating that model has high accuracy. On the other hand, high RMSE indicates that the forecasts deviate significantly from the actual value, indicating lower accuracy.

### Data Records

The MME dataset presented in this paper is open access in the figshare<sup>28</sup>. The APCC MME and participating individual model data are also available for public use on APCC CLimate Information toolKit (CLIK) web page (<https://cliks.apcc21.org>). We provide 10 variables from 1991 to 2010 as hindcast and real-time forecast datasets since 2015. Seasonal forecast data are updated around the 15<sup>th</sup> of each month. Please note that the MME data are anomalies of monthly and seasonal means. On the other hand, individual model data are provided as real-time forecasts or hindcasts, not anomalies. The data period and variables from the individual models depend on each model's version. Some models cannot be redistributed through the APCC because of their security issues (e.g., JMA). All datasets are provided in NetCDF format.

### Technical Validation

We verified the relative prediction skill of four recent sets of the APCC MME for seasonal average of 2-m temperature (T2M), precipitation (PREC), and sea surface temperature (SST) at 1-month and 4-month leads. The variables selected for validation were one of the essential climate variables in atmosphere and ocean specified by the global climate observing system (GCOS; <https://gcos.wmo.int/en/essential-climate-variables/about>). The essential climate variable datasets provide empirical evidence to support climate science and better predict future changes. They can be used to guide mitigation and adaptation measures, assess climate risks, attribute climatic events to their underlying causes, and support climate services.

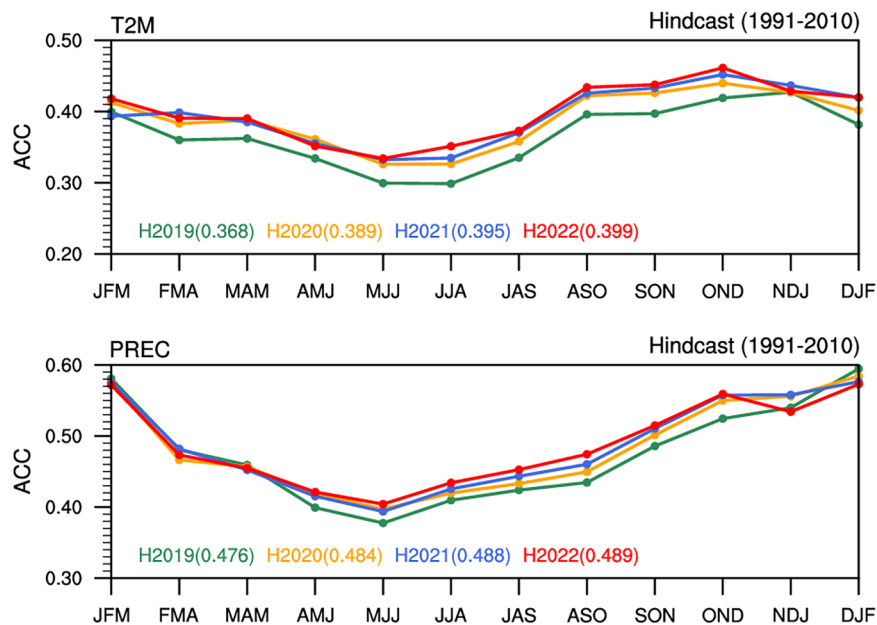
Following the APCC's official verification, the National Center for Environmental Prediction (NCEP)-Department of Energy (DOE) Reanalysis 2 data<sup>29</sup> were used for temperature. Precipitation forecasts were verified using the Climate Anomaly Monitoring System and Outgoing longwave radiation Precipitation Index data (CAM5 OPI<sup>30</sup>). For sea surface temperature, we used the optimum interpolation (OI) version 2 monthly mean SST<sup>31</sup>, obtained from the Climate Diagnostics Center (CDC) of NOAA.

**Hindcast.** Figures 1, 2 show the ACC of 2-m temperature and precipitation at 1-month lead. The ACC generally increases from H2019 to H2021 for both temperature and precipitation, but there are seasonal and regional variations. The skills of temperature and precipitation are relatively low in summer because the ENSO, the origin of seasonal predictability with global influence, has dissipated in summer<sup>5,32</sup>. This indicates that the predictability of 2-m temperature and precipitation is likely related to seasonality of the ENSO cycle: El Niño (La Niña) tends to warm (cool) the tropical atmosphere most predictably and strongly when it reaches maturity and afterward. Hindcast skill in H2022 was improved by 8.4% compared to H2019 in 2-m temperature and showed the highest score during autumn (ASO–NDJ). It is encouraging that there was an improvement of the hindcast skill in the season when the ACC was the lowest (MJJ, JJA). For precipitation, there was a little improvement in skill compared to temperature. The skill score increased by 2.7% from H2019 to H2022. Increased skill was observed in late summer and autumn.

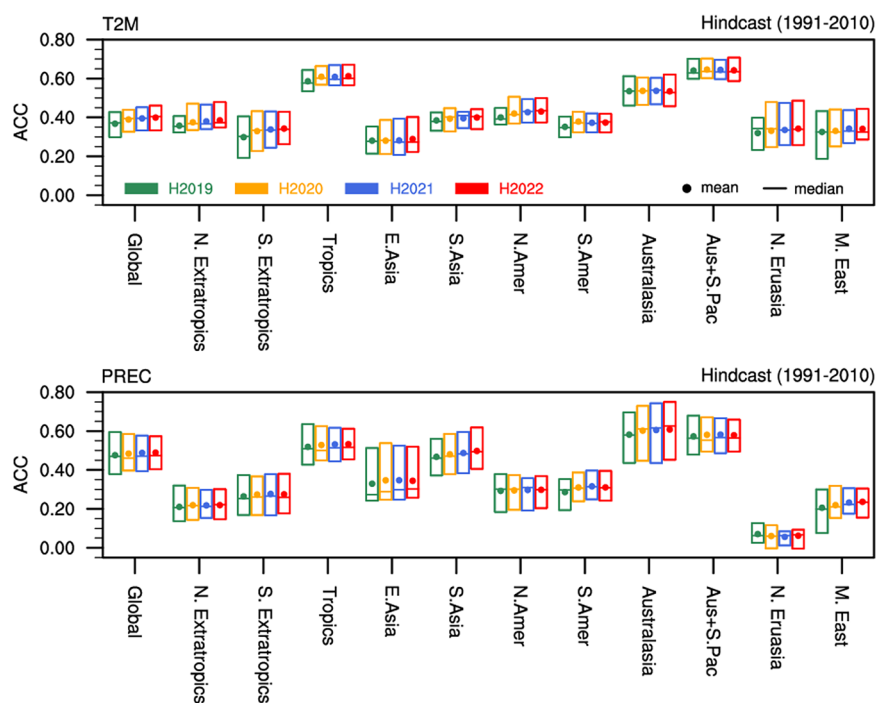
Domain-averaged ACCs were calculated for the subregions in Fig. 2. Definitions of the domains used for verification are provided on the APCC website (<https://apcc21.org/prediction/global/foreveri?lang=en>). The averaged value of ACC gradually increased from H2019 to H2022 in most subregions. Progress is distinct in the southern extratropic and the Middle East in terms of temperature, which means that the minimum value of the ACC has increased. In contrast, predictability near Australasia showed little improvement in H2022. Skill increases in precipitation occur predominantly in South America and the Middle East. The length of each bar represents the range of the prediction skill across the 12 seasons, which indicates the degree of variability in predictability. The seasonal variability in predictability was more pronounced for precipitation than for temperature. In East Asia and Australasia, there was apparent variability in the predictability of precipitation. In East Asia, the median was lower than the mean, indicating that the ACCs were distributed toward the lower end of the bar. In contrast, in the case of Australasia, the median and mean values are similar or the median may slightly higher than the mean.

Despite continuous advancements in climate models, the predictabilities for mid- and high-latitudes are still quite poor<sup>33–35</sup>. Mid-latitudes are characterized by high atmospheric variability. While teleconnection patterns, such as the ENSO, influence the mid-latitude climate, their impacts can be less direct, and more variable compared to those in tropical regions. In addition, mid-latitudes often interact with polar regions, where predictability is even more challenging due to factors such as sea ice variability and polar vortex dynamics. These interactions can further complicate seasonal predictions at mid-latitudes. Therefore, current climate models still have limitations in capturing the detailed processes and feedback mechanisms that are crucial for accurate predictions in mid-latitudes.

To illustrate the spatial distribution of prediction skills, the differences in temporal correlations between H2019 and H2022 at 1-month lead are shown in Fig. 3. For temperature, the predictability was improved in most regions except for northern China and southern part of South America. The global averaged TCC is 0.483 in H2022, which is approximately 8% higher than the value of 0.447 in H2019. For precipitation prediction, the increase in predictability was not clear in H2022 compared to H2019 in terms of the global mean. The global



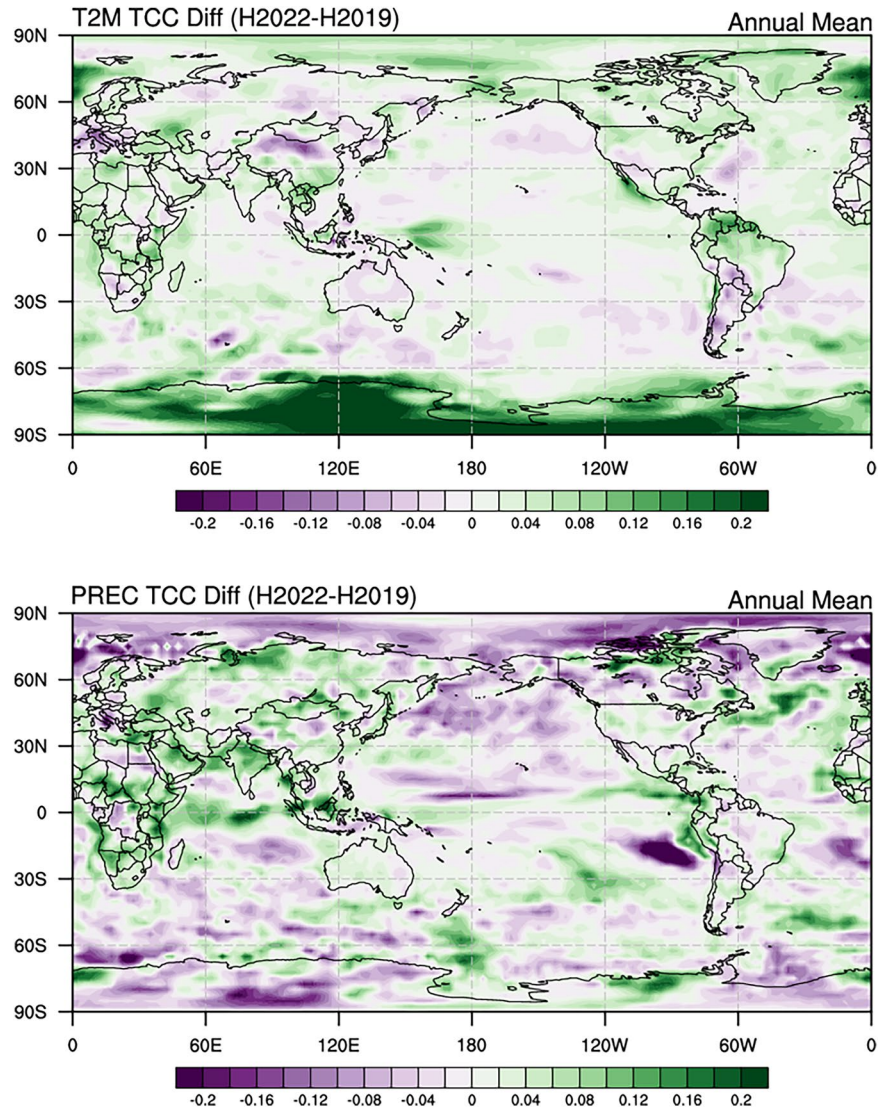
**Fig. 1** Anomaly correlation coefficients of 2-m temperature (top) and precipitation (bottom) at 1-month lead. Annual mean anomaly correlations for each hindcast set are shown in parentheses.



**Fig. 2** Anomaly correlation coefficients of 2-m temperature (top) and precipitation (bottom) for the regions. The length of the bar represents the range of the 12 seasons' predictability. Dots inside the bars are mean of the 12 seasons, and the horizontal lines are the median.

averaged TCC decreased slightly from 0.137 in H2019 to 0.134 in H2022. Improved skill was mainly observed in parts of Africa, Europe, and Asia, whereas correlations decreased in the North Pacific and South Pacific near Peru.

The difference in the temporal correlation of SST showed little change between H2019 and H2022 (Fig. 4). The skill score of SST in H2022 tends to be similar to that in H2019 or lower at 1-month lead, presumably because the predictability of some newly included models is relatively low as the hindcast period changes. Positive values are presented near the Antarctic, but a decline is also shown near 45°S. Improvement is prominent at 4-month lead. The global averaged TCC was 0.413 in H2022, which is approximately 6.7% higher than the value of 0.387

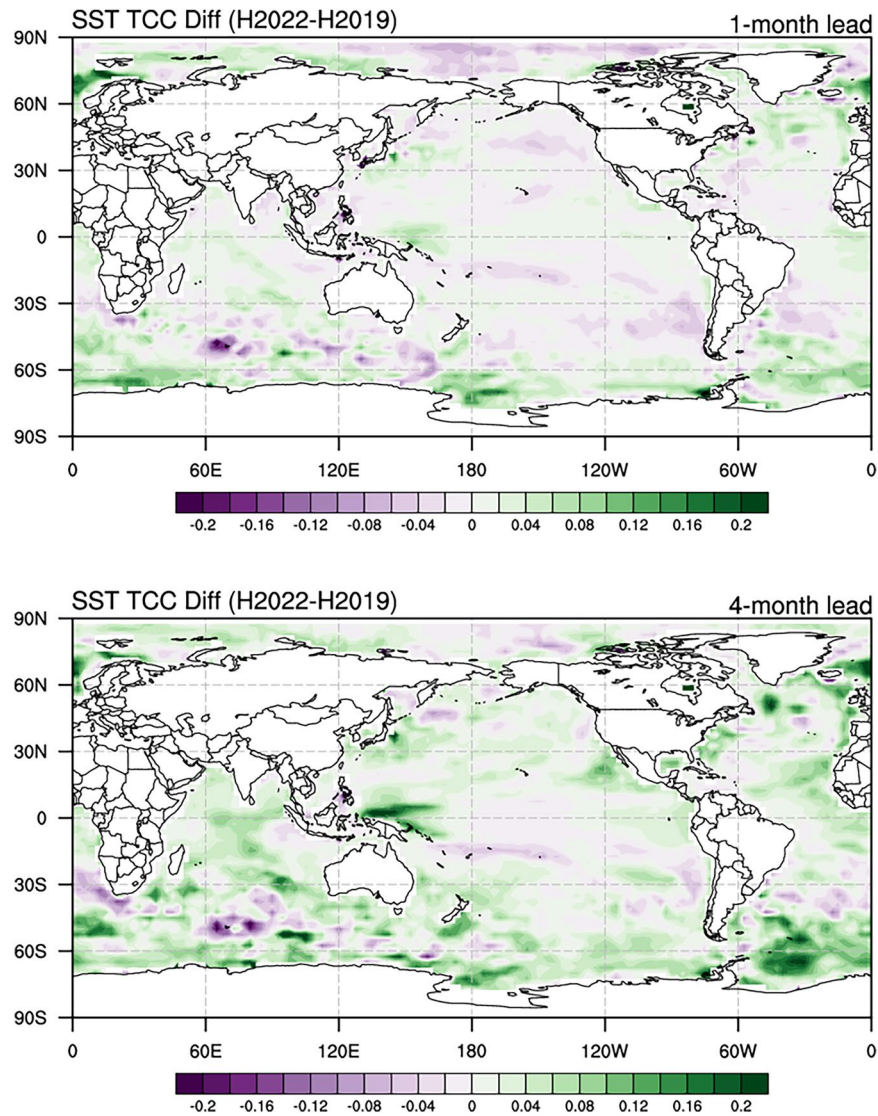


**Fig. 3** Temporal correlation coefficient differences of 2-m temperature (top) and precipitation (bottom) between H2019 and H2022 at 1-month lead.

in H2019. This indicates that predictability is maintained even with longer lead time in the recent hindcast set as the climate models are improved. However, negative values remained at approximately 45°S.

Because the correlation coefficients are not sensitive to bias, we examined the RMSE to investigate the errors in regions and seasons (Fig. 5). Generally, the RMSE values are larger in boreal winter than in boreal summer because of higher variability in winter in extratropical Northern Hemisphere land regions, particularly for the 2-m temperature<sup>36</sup>. In contrast, the RMSE of precipitation was higher in the summer hemisphere owing to a greater variability in summer. Although the ACC increased from H2019 to H2022, the RMSE did not significantly decrease. This indicates that MME is better at capturing the spatial and temporal patterns of deviation from the mean state, even if it does not precisely predict the exact values. However, the RMSE may remain high if there are substantial differences between the predicted and observed values, particularly in terms of intensity or exact amounts.

Figure 6 presents the skills of the individual models, averaged skill of participating models, and MME in each hindcast set. In the case of the 2-m temperature, the highest ACC of the individual model barely changed since H2020, but the lowest score increased. Therefore, the averaged skill of the individual models has been gradually improved and the range of scores has narrowed over time. The MME skill is generally proportional to the mean skill of the individual models. However, the relation between the averaged skill of single models and the MME skill is not linear, and MME skill is superior to the averaged skill of participated models. The effectiveness of the MME approach is mainly attributable to error cancellation and nonlinearity of the skill metrics applied<sup>2</sup>. The effectiveness of the MME technique, which is represented by the red box in Fig. 6, is estimated by the difference between the MME skill and the averaged skill of all models. The efficiency obtained from the MME is the largest in H2020 and has declined since H2020 as the averaged skill of the individual models increased. The range of prediction skills of the individual models is broader for precipitation. The difference between the averaged skill



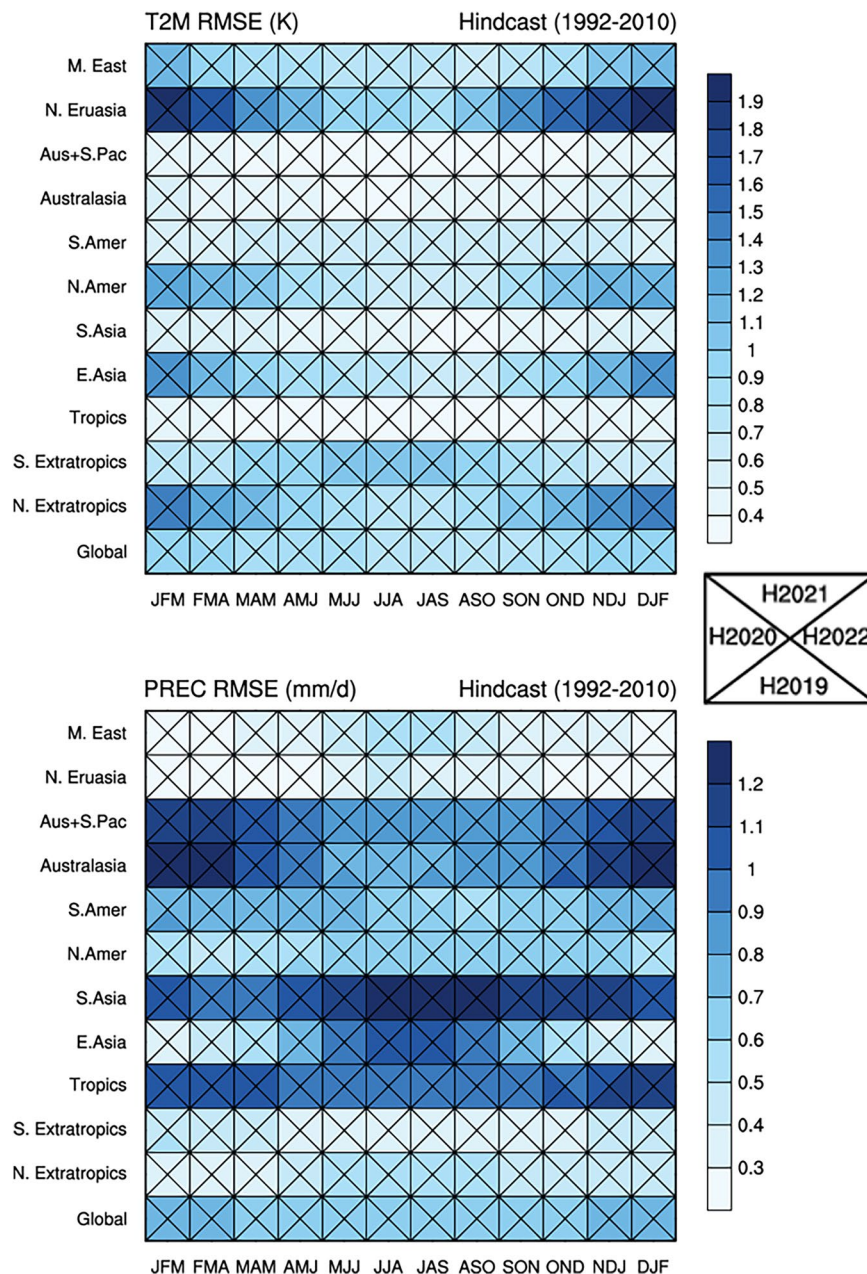
**Fig. 4** Temporal correlation coefficient differences of SST between H2019 and H2022 at 1-month lead (top) and 4-month lead (bottom).

of the individual models and the MME skill was distinct in precipitation compared to 2-m temperature. This indicates that the MME technique is more effective for precipitation than for temperature, mainly because of the mutual independence of the individual models' precipitation forecasts or the independence of the individual models' errors in precipitation forecasts<sup>5,37</sup>. In SST prediction, the ranges of skills from individual models remain from H2019 to H2022. Because the predictabilities of the models for SST were higher than those for temperature and precipitation, the efficiency of the MME was relatively low. However, the predictability of MME is gradually increasing.

**Real-time forecast.** The skill of real-time forecasts has been evaluated relatively rarely in the literature. Meanwhile the assessment of hindcast is necessary to determine systematic errors in the prediction system, real-time operational forecasts provide skill information to the user community<sup>19</sup>.

To assess the predictability of real-time forecasts for SST, the seasonal NINO 3.4 index was analysed in Fig. 7. The NINO 3.4 index is the most commonly used index to define ENSO events by averaging the SST over the Pacific Ocean (5°S–5°N, 120°W–170°W). The ENSO is one of the most predictable climate modes and provides an important source of predictability for seasonal prediction owing to its worldwide effects. The APCC MME generally follows the trend of observation although it sometimes simulates the disagreement of timing or magnitude with the observed peak. The correlation coefficients for the four years from 2019 to 2022 were 0.89 at 1-month lead and 0.72 at 4-month lead.

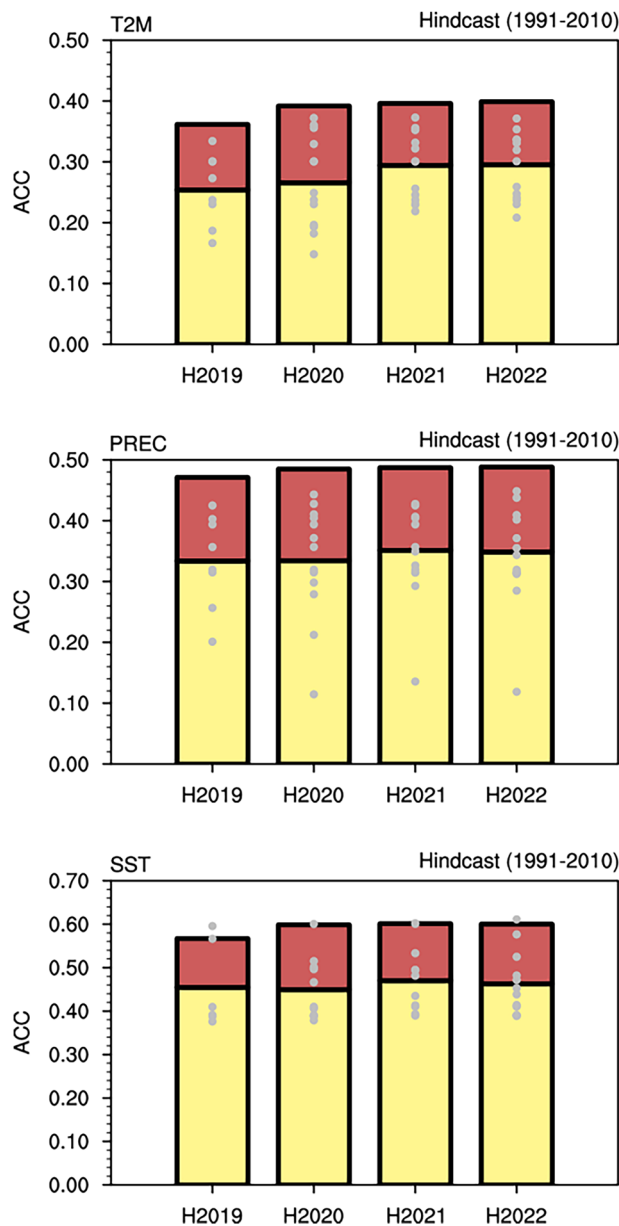
Figure 8 shows the ACC of the 2-m temperature and precipitation for the seasonal mean from 2019 to 2022. The predictability of the real-time MME fluctuates, but shows a reasonably good skill of 0.49 for temperature and 0.43 for precipitation on average. Forecast skill of seasonal precipitation is strongly related to the ENSO variability because it is a reliable source of seasonal predictability. The correlation between the ACC of precipitation



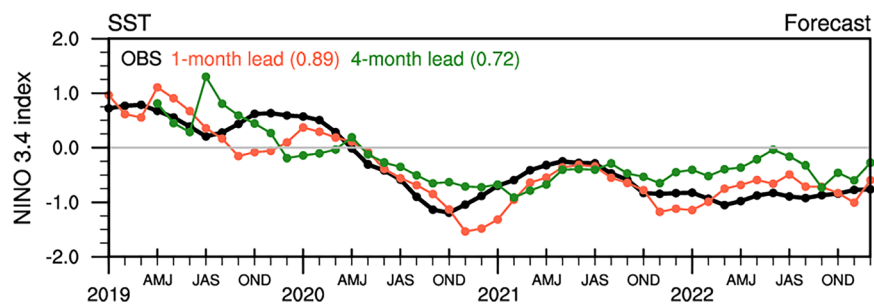
**Fig. 5** Root mean square errors of 2-m temperature (top) and precipitation (bottom) for the regions. Each square is divided into four triangles, and each triangular section represents a hindcast set.

and amplitude of NINO 3.4 index was 0.72. The correlation between the amplitude of NINO 3.4 and the ACC of precipitation was higher than that for 2-m temperature regardless of seasons. This finding agrees with the results of Barnston *et al.*<sup>32</sup> who demonstrated that the greatest impact of ENSO on the forecast skill of temperature occurs four months following the ENSO peak for both ENSO phases due to a delayed temperature response in both the tropics and extratropics<sup>38</sup>. Meanwhile, a simultaneous positive relationship between precipitation and both phases of ENSO was noted. This supports the notion that ENSO variability is likely the sole source of predictability of seasonal precipitation<sup>5,32</sup>.

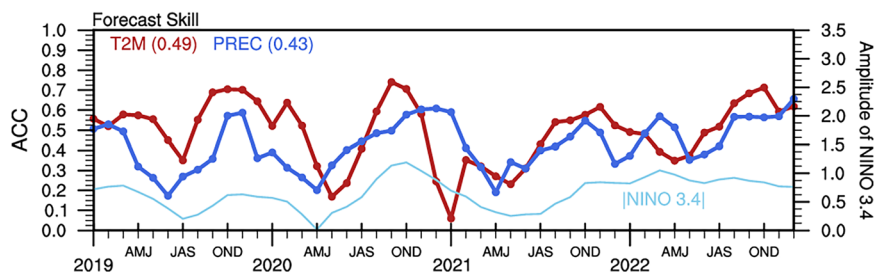
Further analysis for the subregions is shown in Fig. 9. The averaged anomaly correlation for the 48 seasons and temporal correlation from 2019 to 2022 are represented on the X and Y axes. To represent the relative RMSE in the regions, the RMSE range for 2-m temperature and precipitation was divided into four categories, which are illustrated by the size of the circles. For the 2-m temperature, the correlations and RMSE generally have a linear relationship. For example, TCC and ACC are high and RMSE is low in the tropics, Australasia, and South Pacific (Aus + S. Pac). By contrast, in the regions with low correlations such as East Asia and Northern Eurasia, RMSE was also high. However, this positive linear relationship was not observed for precipitation. Regions with high correlations tended to have a high RMSE (e.g., S. Asia, tropics), while regions with low correlations tended



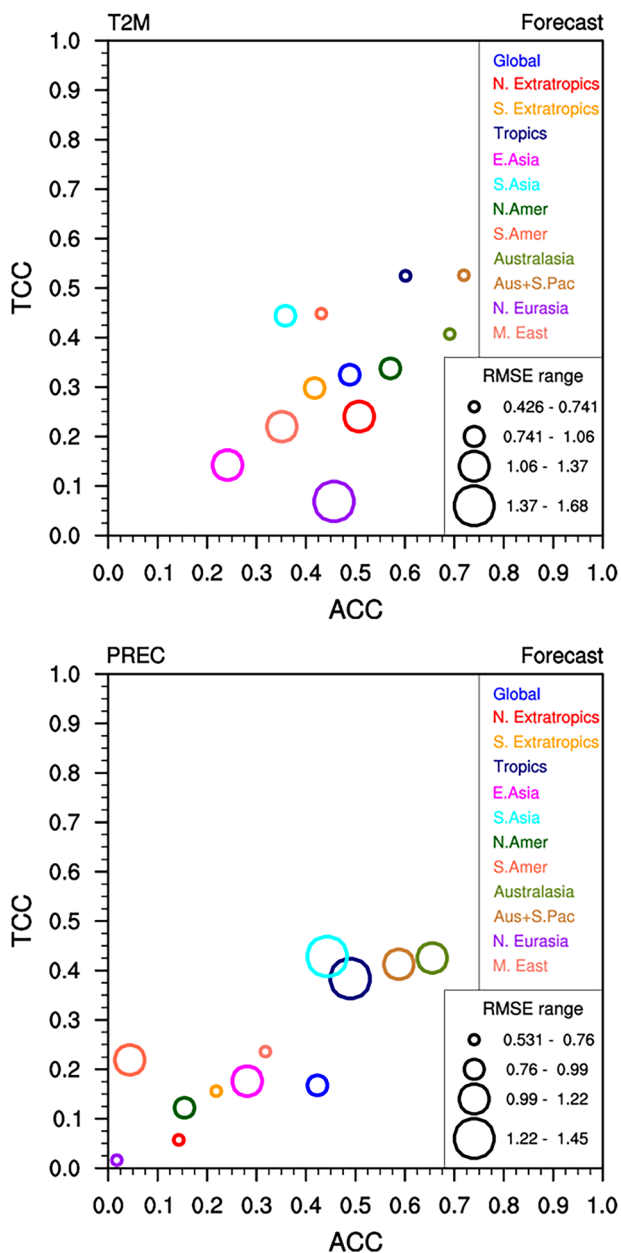
**Fig. 6** Annual mean anomaly correlation coefficients of 2-m temperature (top), precipitation (center), and sea surface temperature (bottom). Yellow bar is the ACC mean of individual models and gray dots are ACCs of each model. Red bars indicate the effectiveness of the MME technique for each hindcast set.



**Fig. 7** Seasonal NINO 3.4 index from 2019 to 2022. Black solid line is observation, orange is the MME at 1-month lead, and green is the MME at 4-month lead. Values in parentheses are temporal correlation coefficients with observation during 2019-2022.



**Fig. 8** Forecast skill of 2-m temperature (red) and precipitation (blue) from 2019 to 2022. Amplitude of NINO 3.4 index is represented in skyblue. Values in parentheses are averaged ACCs during 2019–2022. Temporal correlation coefficient between amplitude of NINO 3.4 and 2-m temperature (precipitation) is 0.42 (0.72).



**Fig. 9** Diagram of 2-m temperature (top) and precipitation (bottom) for the regions from 2019 to 2022. The X axis is anomaly correlation coefficients. The Y axis is temporal correlation coefficients. Sizes of circles indicate magnitude of RMSE. Each color represents a region.

to have low RMSE values (e.g., N. Eurasia). Because the RMSE is a metric influenced by the difference from observed values, regions such as Northern Eurasia where annual precipitation is small, may have a low RMSE; however, the correlation can be low if the MME does not capture the relative variations and patterns.

### Code availability

The MME data were produced using the NCL (version 6.5.0) code by applying Eqs. (1, 2). The code files used to produce the anomalies of the individual models and the MME are available at <https://github.com/bybyhang/MME>. The codes used to verify predictability consisted of NCL scripts.

Received: 14 March 2024; Accepted: 13 February 2025;

Published online: 20 February 2025

### References

- Doblas-Reyes, F. J., Hagedorn, R. & Palmer, T. N. The rationale behind the success of multi-model ensembles in seasonal forecasting: II. Calibration and combination. *Tellus* **57A**, 234–252, <https://doi.org/10.1111/j.1600-0870.2005.00104.x> (2005).
- Hagedorn, R., Doblas-Reyes, F. J. & Palmer, T. N. The rationale behind the success of multi-model ensembles in seasonal forecasting: I. Basic concept. *Tellus* **57A**, 219–233, <https://doi.org/10.1111/j.1600-0870.2005.00103.x> (2005).
- Kirtman, B. P. & Min, D. Multimodel ensemble ENSO prediction with CCSM and CFS. *Mon. Wea. Rev.* **137**, 2908–2930, <https://doi.org/10.3137/Ao1101.2009> (2009).
- Palmer, T. N. *et al.* Development of a European multi-model ensemble system for seasonal to inter-annual prediction. *Bull. Am. Meteorol. Soc.* **85**, 853–872, <https://doi.org/10.1175/BAMS-85-6-853> (2004).
- Wang, B. *et al.* Advance and prospectus of seasonal prediction: assessment of the APCC/ClipAS 14-model ensemble retrospective seasonal prediction (1980–2004). *Clim. Dyn.* **33**, 93–117, <https://doi.org/10.1007/s00382-008-0460-0> (2009).
- Min, Y. M., Kryjov, V. N. & Oh, S. M. Assessment of APCC multimodel ensemble prediction in seasonal climate forecasting: Retrospective (1983–2003) and real-time forecasts (2008–2013). *J. Geophys. Res.* **199**, 12132–12150, <https://doi.org/10.1002/2014JD022230> (2014).
- Min, Y.-M., Kryjov, V. N., Oh, S. M. & Lee, H.-J. Skill of real-time operational forecasts with the APCC multi-model ensemble prediction system during the period 2008–2015. *Clim. Dyn.* **49**, 4141–4159, <https://doi.org/10.1007/s00382-017-3576-2> (2017).
- Becker, E., Kirtman, B. P. & Pegion, K. Evolution of the North American multi-model ensemble. *Geophys. Res. Lett.* **47**, e2020GL087408, <https://doi.org/10.1029/2020GL087408> (2020).
- Kim, G. *et al.* Global and regional skill of the seasonal predictions by WMO lead centre for long-range forecast multi-model ensemble. *Inter. J. Climatol.* **36**(4), 1657–1675, <https://doi.org/10.1002/joc.4449> (2016).
- Alessandri, A. *et al.* Grand European and Asian-Pacific multi-model seasonal forecasts: Maximization of skill and of potential economical value to end-users. *Clim. Dyn.* **50**(7–8), 2719–2738 (2018). 10.1007/s00382-017-3766-y.
- Bell, R. & Kirtman, B. P. Seasonal forecasting of wind and waves in the North Atlantic using a grand multimodel ensemble. *Weather and Forecasting* **34**(1), 31–59 (2019). 10.1175/WAF-D-18-0099.1.
- Torralba, V., Doblas-Reyes, F. J., MacLeod, D., Christel, I. & Davis, M. Seasonal climate prediction: a new source of information for the management of wind energy resources. *J. Appl. Meteorol. Climatol.* **56**(5), 1231–1247, <https://doi.org/10.1175/JAMC-D-16-0204.1> (2017).
- Clark, R. T., Bett, P. E., Thornton, H. E. & Scaife, A. A. Skilful seasonal predictions for the European energy industry. *Environ Res Lett.* <https://doi.org/10.1088/1748-9326/aa57ab> (2017).
- Svensson, C. *et al.* Longrange forecasts of UK winter hydrology. *Environ Res Lett.* <https://doi.org/10.1088/1748-9326/10/6/064006> (2015).
- Morss, R. E. *et al.* Societal and economic research and applications for weather forecasts: priorities for the North American THORPEX Program. *Bull. Am. Meteorol. Soc.* **89**(3), 335–346, <https://doi.org/10.1175/BAMS-89-3-335> (2008).
- Bruno Soares, M., Daly, M. & Dessai, S. Assessing the value of seasonal climate forecasts for decision-making. *Wiley Interdiscip Rev Clim Change.* <https://doi.org/10.1002/wcc.523> (2018).
- Tang, Y., Deng, Z., Zhou, X., Cheng, Y. & Chen, D. Inter-decadal variation of ENSO predictability in multiple models. *J. Clim.* **21**, 4811–4833 (2008).
- DelSole, T. & Shukla, J. Artificial skill due to predictor screening. *J. Clim.* **22**, 331–345, <https://doi.org/10.1175/2008JCLI2414.1> (2009).
- Wang, W., Chen, M. & Kumar, A. An assessment of the CFS real-time seasonal forecasts. *Weather Forecast* **25**, 950–969, <https://doi.org/10.1175/2010WAF2222345.1> (2010).
- Tippett, M. K., Barnston, A. G. & Li, S. Performance of recent multimodel ENSO forecasts. *J. Appl. Meteorol. Clim.* **9**, 637–654, <https://doi.org/10.1175/JAMC-D-11-093.1> (2012).
- Kang, I.-S. *et al.* Intercomparison of GCM simulated anomalies associated with the 1997–98 El Niño. *J. Clim.* **15**, 2791–2805 (2002).
- Kug, J.-S., Lee, J.-Y., Kang, I.-S., Wang, B. & Park, C.-K. Optimal multi-model ensemble method in seasonal climate prediction. *Asia-Pacific J. Atmos. Sci.* **44**, 259–267 (2008).
- Krishnamurti, T. N. *et al.* Improved weather and seasonal climate forecasts from multi-model superensemble. *Science* **285**, 1548–1550 (1999).
- Pavan, V. & Doblas-Reyes, F. J. Multi-model seasonal hindcasts over the Euro-Atlantic: skill scores and dynamical features. *Clim. Dyn.* **16**, 611–625 (2000).
- Rajagopalan, B., Lall, U. & Zebiak, S. E. Categorical climate forecasts through regularization and optimal combination of multiple GCM ensembles. *Mon. Wea. Rev.* **130**, 1792–1811 (2002).
- Wilks, D. *Statistical methods in the atmospheric sciences*, (p. 627). London: Academic Press (2006).
- Van den Dool, H. *Empirical methods in short-term climate prediction*, (p. 215). Oxford: Oxford University Press (2007).
- APEC climate center multi-model ensemble dataset for seasonal climate prediction. *figshare*, <https://doi.org/10.6084/m9.figshare.28321397> (2025).
- Kanamitsu, M. *et al.* NCEP-DOE AMIP-II reanalysis (R-2). *Bull. Am. Meteorol. Soc.* **83**, 1631–1643. 10.1175/BAMS-83-11-1631(2002)083<1631:NAR>2.3.CO;2 (2002).
- Janowiak, J. E., Xie, P. CAMS\_OPI: A global satellite-raingauge merged product for real-time precipitation monitoring applications. *J. Clim.* **12**, 3335–3342. 10.1175/1520-0442(1999)012<3335:COAGSR>2.0.CO;2 (1999).
- Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C. & Wang, W. An improved *in situ* and satellite SST analysis for climate. *J. Clim.* **15**, 1609–162 (2002).
- Barnston, A. G. *et al.* Verification of the first 11 years of IRI's seasonal climate forecast. *J. Appl. Meteorol. Climatol.* **49**, 493–520 (2010).
- Kharin, V. V. & Zwiers, F. W. Skill as function of time scale in ensemble of seasonal hindcast *Clim. Dyn.* **17**, 127–41 (2001).

34. Yun, W.-T., Stefanova, L. & Krishnamurti, T. N. Improvement of the multimodel superensemble technique for seasonal forecast. *J. Clim.* **16**, 3834–40 (2003).
35. Kryjov, V. N. Seasonal climate prediction for North Eurasia. *Environ. Res. Lett.* **7**, 015203 (2012).
36. Kim, G. *et al.* A. Assessment of MME methods for seasonal prediction using WMO LC-LRFMME hindcast dataset. *Inter. J. Climatol.* **41**, E2462–E2481 (2021).
37. Yoo, J. H. & Kang, I. S. Theoretical examination of a multi-model composite for seasonal prediction. *J. Geophys. Res.* **32**, L18707, <https://doi.org/10.1029/2005GL023513> (2005).
38. Kumar, A. & Hoerling, M. P. The nature and causes for the delayed atmospheric response to El Niño. *J. Clim.* **16**, 1391–1403, <https://doi.org/10.1175/2771.1> (2003).
39. Ham, S., Lim, A., Kang, S., Jeong, H. & Jeong, Y. A newly developed APCC SCoPS and its prediction of East Asia seasonal climate variability. *Clim. Dyn.* **52**(11), 6391–6410 (2019).
40. Wu, T. *et al.* An overview of BCC climate system model development and application for climate change studies. *J. Meteor. Res.* **28**, 34–56 (2014).
41. Wedd, R. *et al.* ACCESS-S2: the upgraded Bureau of Meteorology multi-week to seasonal prediction system. *J. Southern Hemisphere Earth Systems Science* **72**(3), 218–242 (2022).
42. Gumdi, S. *et al.* The new CMCC operational seasonal prediction system, CMCC Technical Note (RP0288), <https://doi.org/10.25424/CMCC/SPS3.5> (2020).
43. Paek, H., Yu, J.-Y., Hwu, J.-W., Lu, M.-M. & Gao, T. A source of AGCM bias in simulating the western Pacific subtropical high: Different sensitivities to the two types of ENSO. *Mon. Wea. Rev.* **143**(6), 2348–2362 (2015).
44. Lin, H. *et al.* The Canadian seasonal to interannual prediction system version 2 (CanSIPSv2). *Weather and Forecasting* **35**(4), 1317–1343 (2020).
45. Fadeev, R. Y., Tolstykh, M. A. & Volodin, E. M. Climate version of the SL-AV atmospheric model: development and preliminary results. *Russian Meteorology and Hydrology* **44**, 13–22 (2019).
46. Hirahara, S. *et al.* Japan Meteorological Agency/Meteorological Research Institute Coupled Prediction System Version 3 (JMA/MRI-CPS3). *J. Meteor. Soc. Japan. Ser. II* **101**(2), 149–169 (2023).
47. Kim, H., Lee, J., Hyun, Y. K. & Hwang, S. O. The KMA Global Seasonal forecasting system (GloSea6)-Part 1: operational system and improvements. *Atmosphere* **31**(3), 341–359 (2021).
48. Batté, L., Dorel, L., Ardilouze, C., & Guérémy, J. F. Documentation of the METEO-FRANCE seasonal forecasting system 8. C3S\_D330\_3.3.1, <https://www.umr-cnrm.fr/IMG/pdf/system8-technical.pdf> (2021).
49. Meleshko V. P. *et al.* MGO atmospheric general circulation model (version MGO-03 T63L25), *MGO Proc.*, **571**, 5–87 (2014).
50. Molod, A. *et al.* GEOS-S2S version 2: The GMAO high-resolution coupled model and assimilation system for seasonal prediction. *J. Geophys. Res: Atmospheres* **125**(5), e2019JD031767 (2020).
51. Saha, S. *et al.* The NCEP climate forecast system version 2. *J. Clim.* **27**(6), 2185–2208 (2014).
52. Ahn, J. B., Lee, J. & Jo, S. Evaluation of PNU CGCM ensemble forecast system for boreal winter temperature over South Korea. *Atmosphere* **28**(4), 509–520 (2018).
53. Williams, K. D. *et al.* The Met Office global coupled model 3.0 and 3.1 (GC3.0 and GC3.1) configurations. *J. Adv. Modeling Earth Systems* **10**(2), 357–380 (2018).

## Acknowledgements

This research was funded by the Korea Meteorological Administration Research and Development Program “APEC Climate Center for Climate Information Services” under Grant (KMA2013-07510). The authors acknowledge that the APCC Multi Model Ensemble (MME) Producing Centers for making their hindcast/forecast data available for analysis and the APEC Climate Center for collecting and archiving them and for organizing APCC MME prediction.

## Author contributions

Y.-B. Y. designed the research, analysed the results, contributed to the graphics and wrote the main text of the manuscript. C.-M. L. and D. J. collected model data and performed processes related to MME. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to Y.-B.Y.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

© The Author(s) 2025