

Research article

Assessing the Reliability of Predicted Decadal Surface Temperatures in Southeast Asia

Dara Kasihairani¹, Rahmat Hidayat^{1*}, Supari Supari²¹ Applied Climatology Program, Department of Geophysics and Meteorology, Faculty of Mathematics and Natural Sciences, IPB University, Bogor 16680, Indonesia² Directorate of Climate Change, Deputy for Climatology, BMKG, Indonesia^{*}) Correspondence: rahmath@apps.ipb.ac.id**Citation:**

Kasihairani, D., Hidayat, R., Supari, S. (2024). Assessing the Reliability of Predicted Decadal Surface Temperatures in Southeast Asia. *Forum Geografi*. 38(3), 413-425.

Article history:

Received: 15 June 2024

Revised: 15 October 2024

Accepted: 21 December 2024

Published: 27 December 2024

Abstract

Climate predictions spanning 10-year periods, known as Decadal Climate Predictions (DCPs), have become an important aspect of the latest Coupled Model Intercomparison Project (CMIP6). These DCPs have the capability to capture the El Niño-Southern Oscillation (ENSO) phenomena, which affects heatwave frequency in Southeast Asia over years to decades. This research assesses the ability of six General Circulation Model (GCM) DCPs to predict surface temperature over the Southeast Asian region, using the dcpp-A hindcast as the main product. The metrics of Anomaly Correlation Coefficient (ACC) and Mean Error (ME) are employed to assess the model outputs, with 51 hindcast datasets spanning initial years from 1960 to 2010 and ERA5 reanalysis data serving as the reference. The evaluation reveals that DCP model skill varies across lead times and subregions, with no single model consistently outperforming the others. The highest correlation values are observed during the September-October-November (SON) season, and the ENSEMBLE model demonstrates the ability to increase correlation values compared to the individual DCP models. However, the ENSEMBLE approach is unable to effectively reduce ME values due to the contrasting errors among individual models. PBIAS metric aligns with the ME, consistently identifying similar areas of underestimation (mainland) and overestimation (maritime continent) across the models. Despite these challenges, the evaluation results highlight the potential of DCPs in predicting surface temperature variability for the Southeast Asian region over decadal periods, particularly in capturing ENSO-related signals. Further improvements in model initializations, internal variability representation, and bias reduction are necessary to enhance the utility of CMIP6 decadal predictions for heatwave preparedness and mitigation strategies in this vulnerable region.

Keywords: ENSO; decadal; dcpp-A hindcast; CMIP6; evaluation.

1. Introduction

Southeast Asia has a tropical climate characterized by relatively small variations in surface temperatures throughout the year. However, Southeast Asia's mainland is experiencing an increasing trend of extreme temperature events, causing human discomfort and health concerns (Thirumalai *et al.*, 2017). Observational data shows that all heatwave-related characteristics, including frequency, intensity, and duration, display rising trends in the majority of Southeast Asia (Li *et al.*, 2022), although the magnitude and statistical significance of these trends vary across different regions (Guo *et al.*, 2017). There is substantial evidence linking high temperatures to significant health burdens (Arsad *et al.*, 2022), with heatwaves posing a greater mortality risk in moderately cold and moderately hot areas compared to extremely cold or hot regions (Guo *et al.*, 2017).

Furthermore, in mainland Southeast Asia, it has been discovered that the El Niño-Southern Oscillation climate variability mode is tightly related to heatwave events (Lin *et al.*, 2018). During El Niño years, an increase in heatwave frequency, duration, and amplitude is found. Meanwhile, during the La Niña years, there is a decrease in these characteristics. These ENSO events may be predicted by decadal climate prediction (DCP), which covers a multiannual to decades long prediction period to support mitigation or planning efforts to reduce the death toll from heatwave events. During El Niño years, an increase in heatwave frequency, duration, and amplitude is observed, while La Niña years tend to exhibit a decrease in these characteristics. The ability to predict ENSO events through DCP, which covers multiannual to decadeslong time scales, is crucial for supporting mitigation strategies and planning efforts to minimize the mortality and health consequences of heatwave occurrences in the region.

DCP, a new type of forecast, is categorized as a short-term climate prediction. The previous Coupled Model Intercomparison Project (CMIP5) protocol for coordinated climate change experiments strongly emphasized decadal predictability and prediction. This emphasis on decadal forecasting has transitioned into a significant aspect of the current Coupled Model Intercomparison Project (CMIP6) (Boer *et al.*, 2016). DCP is divided into three components. The primary component, dcpp-A hindcast, encompasses a retrospective forecast for ten years ahead of time, with the first initialization of hindcast from 1960 to the present. It provides enough dataset numbers for



Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

the evaluation study. Forecast products are the focus of the dcpp-B component. The objective of this component is to satisfy societal needs by developing a nearly operational process for forecasting, aggregating, and combining data for decadal climate prediction. The component adopts CMIP5, other models, and hindcast data sets to produce its real-time decadal forecasts. Component C is the final component and focuses on studies of the ability to forecast climate patterns and variability over a given decade, as well as specific cases and coordinated processes related to climate variability.

DCP is an emerging field of forecasting that attempts to fill the gap between short-term seasonal predictions and long-term projections of climate. Given the relatively new nature of this approach, it is crucial to assess the performance of DCP models to determine their capability to attain specific forecasting objectives. While previous global verification studies have demonstrated the skill and potential of DCP models (Corti *et al.*, 2012; Döscher *et al.*, 2022; Kadow *et al.*, 2016; Nicolì *et al.*, 2023), their reliability in predicting surface temperatures specifically for the Southeast Asia region remains largely unexplored. Consequently, this study's primary purpose is to comprehensively evaluate the reliability of DCP models in forecasting surface temperature variability on decadal periods across Southeast Asia. This assessment is essential for determining the suitability and limitations of these models in supporting climate adaptation strategies and decision-making processes within this climatically diverse and vulnerable region.

2. Research Methods

2.1. Study Area

Southeast Asia is the study region (Figure 1). The region consists of the maritime continent of Indonesia, Timor-Leste, The Philippines, Papua New Guinea, and the mainland of Southeast Asia (Myanmar, Cambodia, Lao, and Thailand), including the Malay Peninsula. The area spans 89,26 E-146,96 E and 27,26 N-15,14 S (CORDEX, 2015). Further analysis also divides the region into the mainland subregion (brown dash line) and maritime continent (blue dash line). The mainland subregion is from 93°E to 107°E and 22°N to 12°N while the maritime continent subregion spread from 95°E to 140°E and 8°N to 12°S.

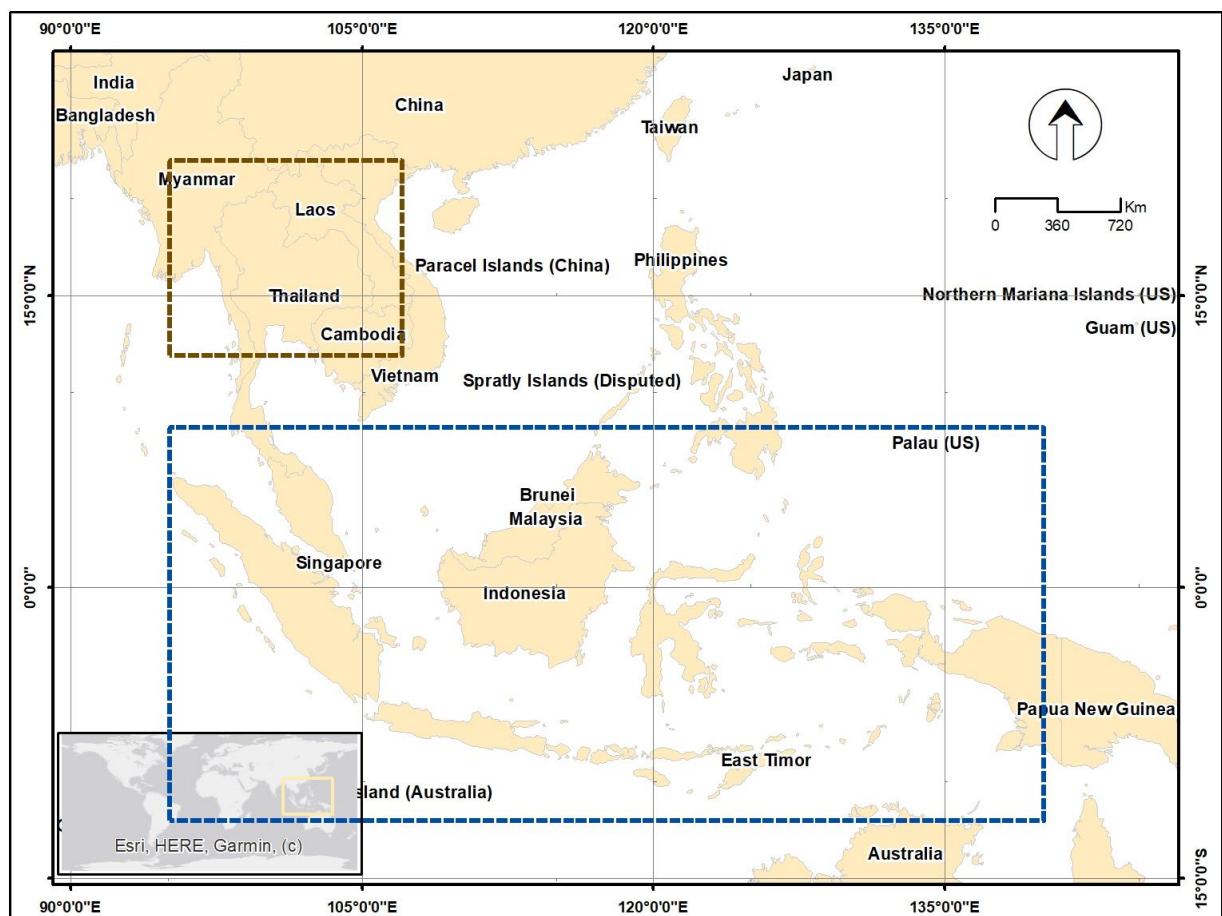


Figure 1. Southeast Asia as the region of the evaluation study.

2.2. Data and Method

The surface temperature of a 2-meter height from six CMIP6 DCPP-A Hindcast will be evaluated using reanalysis data of ERA5 (Hersbach et al., 2020), as a benchmark reference. The analysis utilizes each model's first ensemble member (r1i1p1f1) to ensure a consistent and reliable comparison. These data were then re-gridded to have equal spatial resolution. Each model has a ten-year prediction, with 1960 as the first initialization year. The prediction starts from 1961 to 2011 to meet the last ERA5 data of reference, 2020. Therefore, each model has 51 datasets of decadal climate prediction experiments. The requisite data were accessed and downloaded from the following website: <https://esgf-node.llnl.gov/search/cmip6/>. The measurement of the surface temperature prediction models will use the Anomaly Correlation Coefficient (ACC), a dimensionless metric, to measure and assess the phase agreement between anomalies of retrospective forecast and observation, ranging from -1 to 1. Two bias verification metrics Mean Error (ME) and Percent Bias (PBIAS) will incorporate the ACC. The ACC is given below Equation 1.

$$ACC = \frac{\sum_{i=1}^N (Y'_i - \bar{Y}_i)(X'_i - \bar{X}_i)}{\sqrt{\sum_{i=1}^N (Y'_i - \bar{Y}_i)^2 (X'_i - \bar{X}_i)^2}} \quad (1)$$

where \bar{Y}_i denotes forecast average value, \bar{X}_i denotes observation average value and $(Y'_i - \bar{Y}_i)^2$ means the squared standard deviations of the forecast anomalies, while $(X'_i - \bar{X}_i)^2$ means the squared standard deviations of analysis anomalies from the climate, respectively. The statistical significance of the ACC is ascertained through the implementation of a two-tailed t-statistic Equation 2 ($\alpha = 0.05, n = 51$).

$$t = \frac{ACC\sqrt{n-2}}{\sqrt{1-ACC^2}} \quad (2)$$

ME and PBIAS measure the models' tendencies to underestimate or overestimate. ME provides a deterministic quantification of the average magnitude of forecast errors while PBIAS illustrates the diverging ratio of the model's hindcast relative to the observed data (Mengistu et al., 2021; Viana et al., 2021). These two bias metrics complement each other. Let Y_i is the model forecast (or hindcast) and X_i is the observation data. Negative values indicate model underestimation, while positive values indicate overestimation in Equations 3 and 4.

$$ME = \frac{1}{N} \sum_1^N (Y_i - X_i) \quad (3)$$

$$PBIAS = 100 \times \frac{\sum (Y_i - X_i)}{\sum X_i} \quad (4)$$

The decadal climate prediction of ten years is then divided into three parts: the period called the first half decadal (FHDecad) for prediction year 1 to 5 average, the second half decadal (SHDecad) for prediction year 6 to 10 average, and the decadal (Decad) for ten years prediction average (year 1 to 10). The investigation will examine how the model's performance varies by season. The season division will be December to February (DJF), March to May (MAM), June to August (JJA), and September to November (SON). Table 1 lists the models that were used in this investigation. Apart from the six models, the ENSEMBLE model is the average of all studied models.

Tabel 1. CMIP6 DCPP-A Hindcast GCM.

Model	Institution	Spatial resolution
CMCC-CM2-SR5 (Nicolì et al., 2023)	CMCC (Canada)	0,9 x 1,25
MPI-ESM1-2-HR (Müller et al., 2018)	DWD (Germany)	0,9 x 0,9
IPSL-CM6A-LR (Boucher et al., 2020)	IPSL (France)	1,25 x 2,5
MIROC6 (Tatebe et al., 2019)	MIROC (Japan)	1,4 x 1,4
NorCPM1 (Bethke et al., 2021)	NCC (Norway)	1,9 x 2,5
FGOALS-f3-L (Hu et al., 2023)	CAS (China)	1,0 x 1,0

3. Results and Discussion

The annual time series of mean surface temperature hindcasts for Southeast Asia from each climate model, along with the corresponding ERA5 reanalysis data, are presented in Figure 2. The figure displays the temperature anomaly values for three distinct periods: the FHDecad, the SHDecad, and the Decad period. Consistent with previous studies examining global mean surface temperature trends (Corti et al., 2012; Hu et al., 2023; Nicolì et al., 2023), the model hindcasts

indicate a similar increasing pattern in surface temperatures for the Southeast Asia region over the analysed period. This similarity suggests that the region is experiencing temperature changes in line with the broader global warming trend.

The average annual anomaly hindcast in all models ranges from -0.61°K to 1.25°K for FHDecad and -0.55°K to 1.25°K for SHDecad. As for the Decad period, the range is from -0.58°K to 1.17°K . Although the Southeast Asia region is following the trend of global climate change, the retrospective predictions can capture the effect of volcanic forcing. The temperature anomaly decreased in value between 1982 and 1991 due to the cooling effect caused by the El Chichón and Pinatubo volcanic eruptions, as indicated by NorCPM1 (Bethke et al., 2021), IPSL-CM6A-LR (Boucher et al., 2020), MIROC6 (Kataoka et al., 2020), and CMCC-CM2-SR5 models (Nicolò et al., 2023). The graph also reveals that CMCC-CM2-SR5, MIROC6, and MPI-ESM1-2-HR models tend to overpredict surface temperatures, resulting in positive anomalies, while the FGOALS-f3-L, IPSL-CM6A-LR, and NorCPM1 models typically underpredict, leading to negative anomalies.

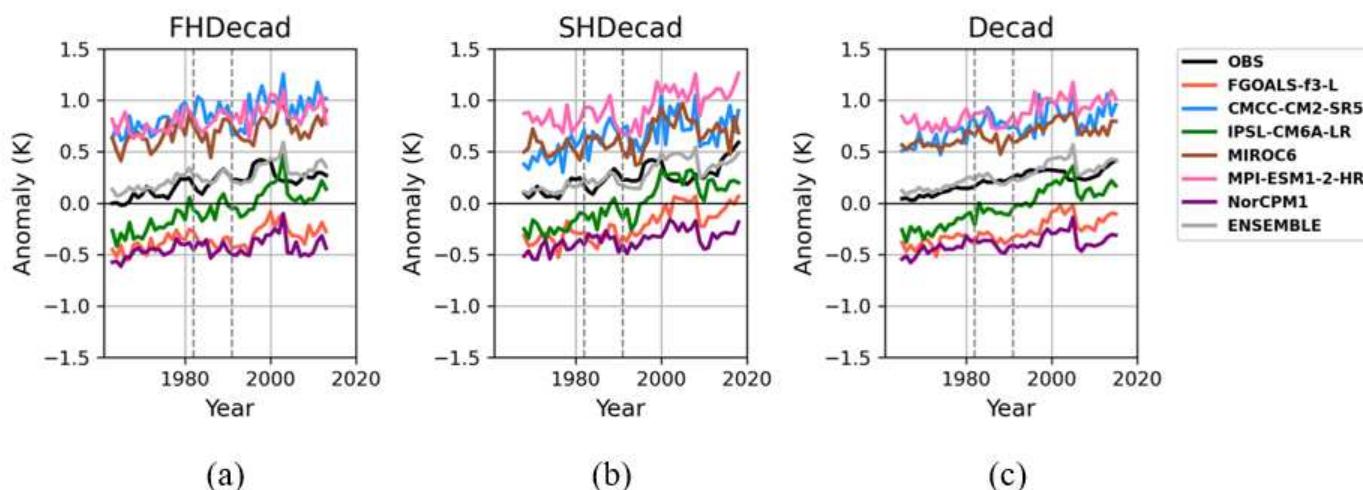


Figure 2. Time series graph of regional average surface temperature anomaly of Southeast Asia. (a) FHDecad average shown in, (b) is for the SHDecad average, and (c) is for the Decad average. The vertical dash line (grey) indicates 1982 and 1991.

3.1. Metrics of Verification

The direct evaluation of the models' proficiency utilized the ERA5 reanalysis data as a baseline reference. The relevant climatological period range was considered when calculating anomalies for the hindcast and reference data. Employing the ACC metric, an evaluation was conducted to determine the decadal climate prediction models' ability to capture the climatological pattern of the region. Furthermore, the ME metric will indicate part of the region is inclined towards warm or cold bias. At the same time, PBIAS will measure the relative degree to which the model's predictions diverge from the reference climate condition.

Southeast Asian climate models generally show positive ACC values (Figure 3). However, negative values are observed in mainland areas, particularly during winter (DJF) and spring (MAM) seasons and in some oceanic regions. Among the models, IPSL-CM6A-LR demonstrates superior performance with higher significant ACC values. In contrast, MIROC6 and NorCPM1 exhibit the lowest ACC values. Significant ACC values are rarely found in mainland areas across all models. Additionally, models show agreement in positive ACC during the fall (SON) season, with the highest values of positive ACC.

During SHDecad, the ACC exhibits a similar pattern to FHDecad. However, lower ACC values are observed in FGOALS-f3-L, CMCC-CM2-SR5, and MIROC6 (especially for MAM and JJA seasons), as well as IPSL-CM6A-LR (Figure 4). While NorCPM1 displays slightly higher values for this period, MPI-ESM1-2-HR shows a notable increase in ACC value.

In the Decad period category (figure not shown), the ACC values of the models are in the middle of FHDecad and SHDecad values range. The spatial pattern of ACC values in this period is similar to FHDecad and SHDecad. Notably, the ENSEMBLE model, except for a few areas of the maritime continent, increases the correlation value for each season in the region across the three time period categories. Additionally, it is crucial to note that the ACC value is more prevalent in the oceanic part than in the region's terrestrial area.

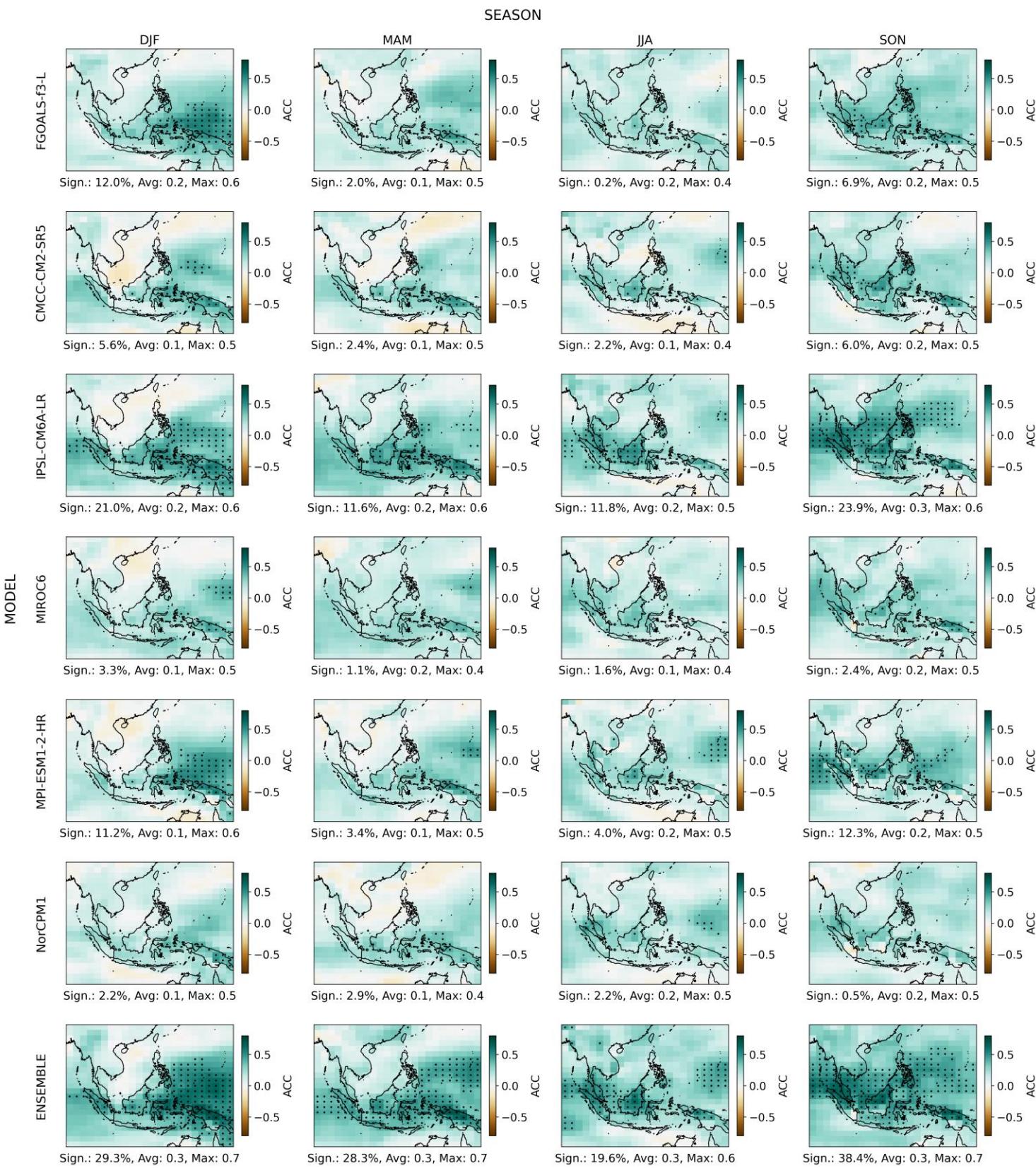


Figure 3. Anomaly Correlation Coefficient (ACC) for FHDecad period from DCP and ENSEMBLE of DCP models. The dots in the figure signify that the experimental results across all datasets satisfy the statistical significance criteria with 95% confidence. The ACC maximum and average values are shown below the subplots and the fraction of the significant grid area.

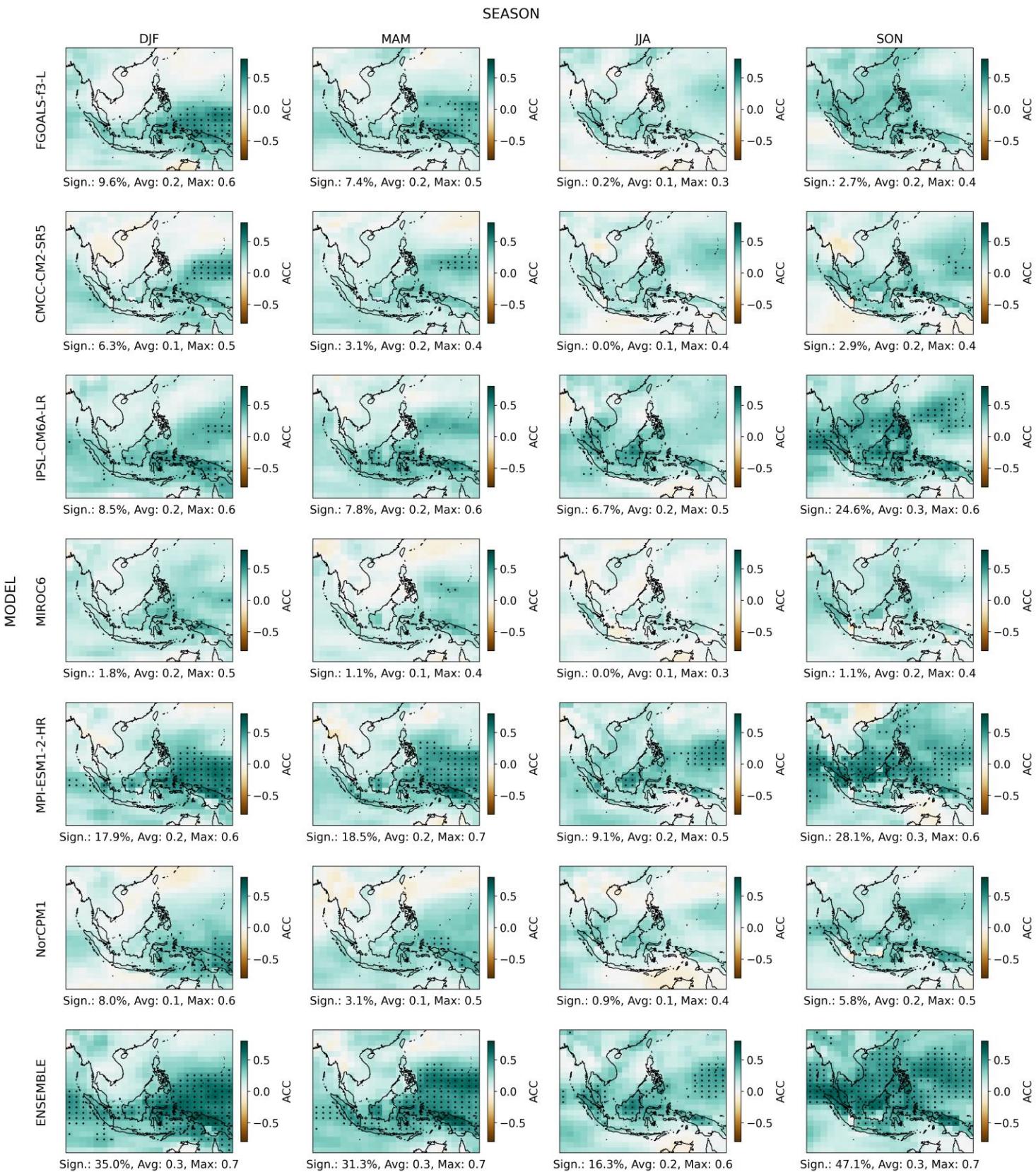


Figure 4. Anomaly Correlation Coefficient (ACC) for SHDecad period from DCP and ENSEMBLE of DCP models. The dots in the figure signify that the experimental results across all datasets satisfy the statistical significance criteria with 95% confidence. The ACC maximum and average values are shown below of the subplots, along with the fraction of the significant grid area.

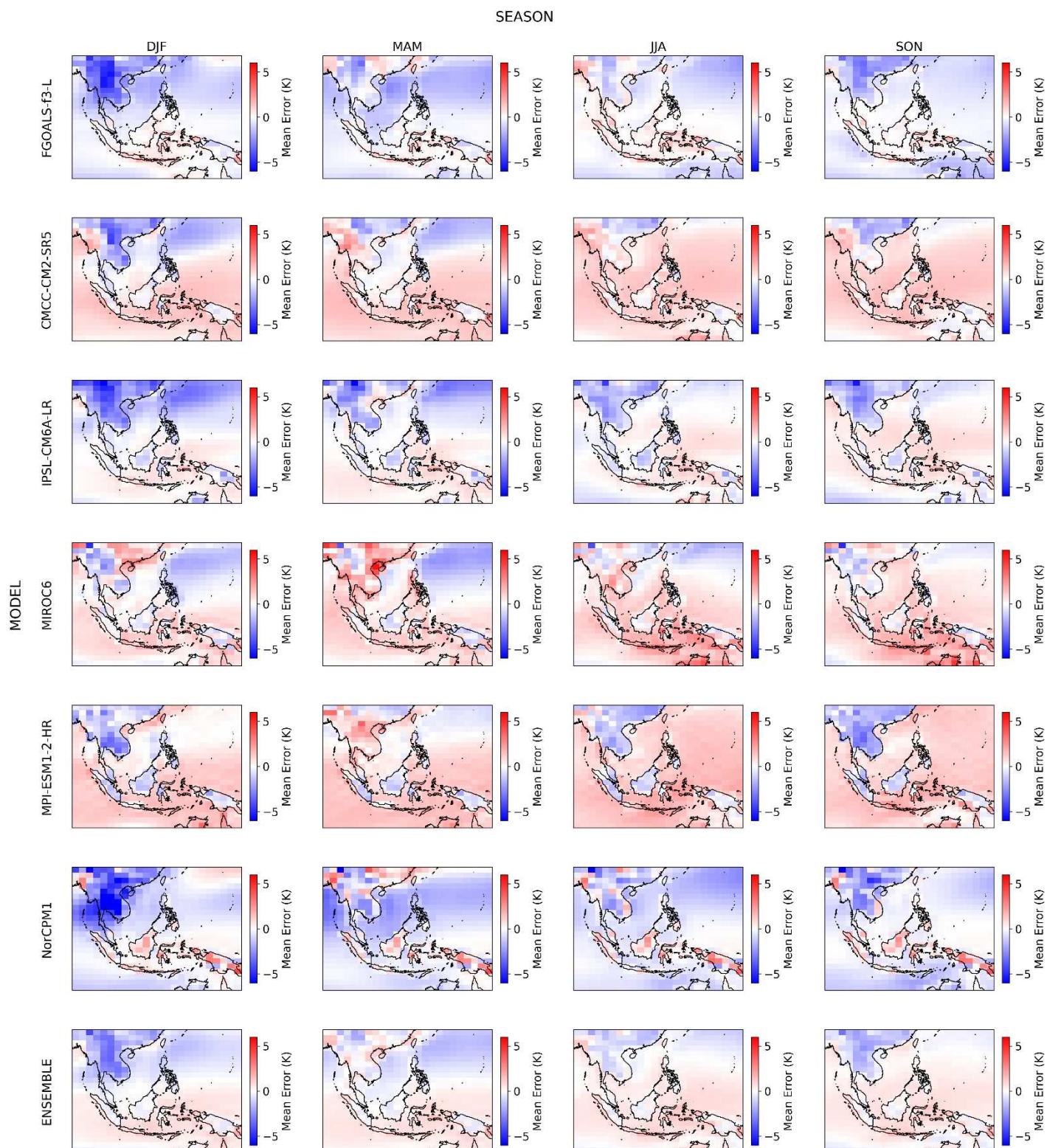


Figure 5. Mean Error value for Decad period from DCP and ENSEMBLE of DCP models. FHDecad and SHDecad are not shown, as their pattern resemble each other.

The value pattern of the ME metric exhibits similarity across all time categories (FHDecad, SHDecad, and Decad), underscoring the importance of understanding how each model's hindcast responds to different surface features of land and ocean in the region. Figure 5, the FGOALS-f3-L model's ME value varies over terrestrial areas but remains invariable above the ocean, displaying a cold bias. Similarly, the NorCPM1 model exhibits a clear hot bias over the maritime continent area. Conversely, the MPI-ESM1-2-HR model's pattern is almost the opposite of FGOALS-f3-L and NorCPM1, showing a cold bias for the terrestrial area of the region (except for the central mainland area in the MAM season). At the same time, the temperature hindcast above the ocean uniformly exhibits a hot bias.

The CMCC-CM2-SR5, MIROC6, and IPSL-CM6A-LR models share a similar pattern above the ocean part of the region, where they exhibit a hot bias on the equatorial region and a cold bias in the upper equatorial region. Notably, the CMCC-CM2-SR5 model tends to display a cold bias for the eastern part of the mainland, while the western part and maritime continent predominantly have a hot bias. The MIROC6 ME value varies across the region, and the IPSL-CM6A-LR model tends to have a cold bias over the terrestrial area of the region. The fact that one model has a relatively large error and another has a relatively small error prevents the ENSEMBLE model from concluding that it has effectively reduced the ME value. On the other hand, it can detect the distinct pattern of mean error values between the region's upper and equatorial sections.

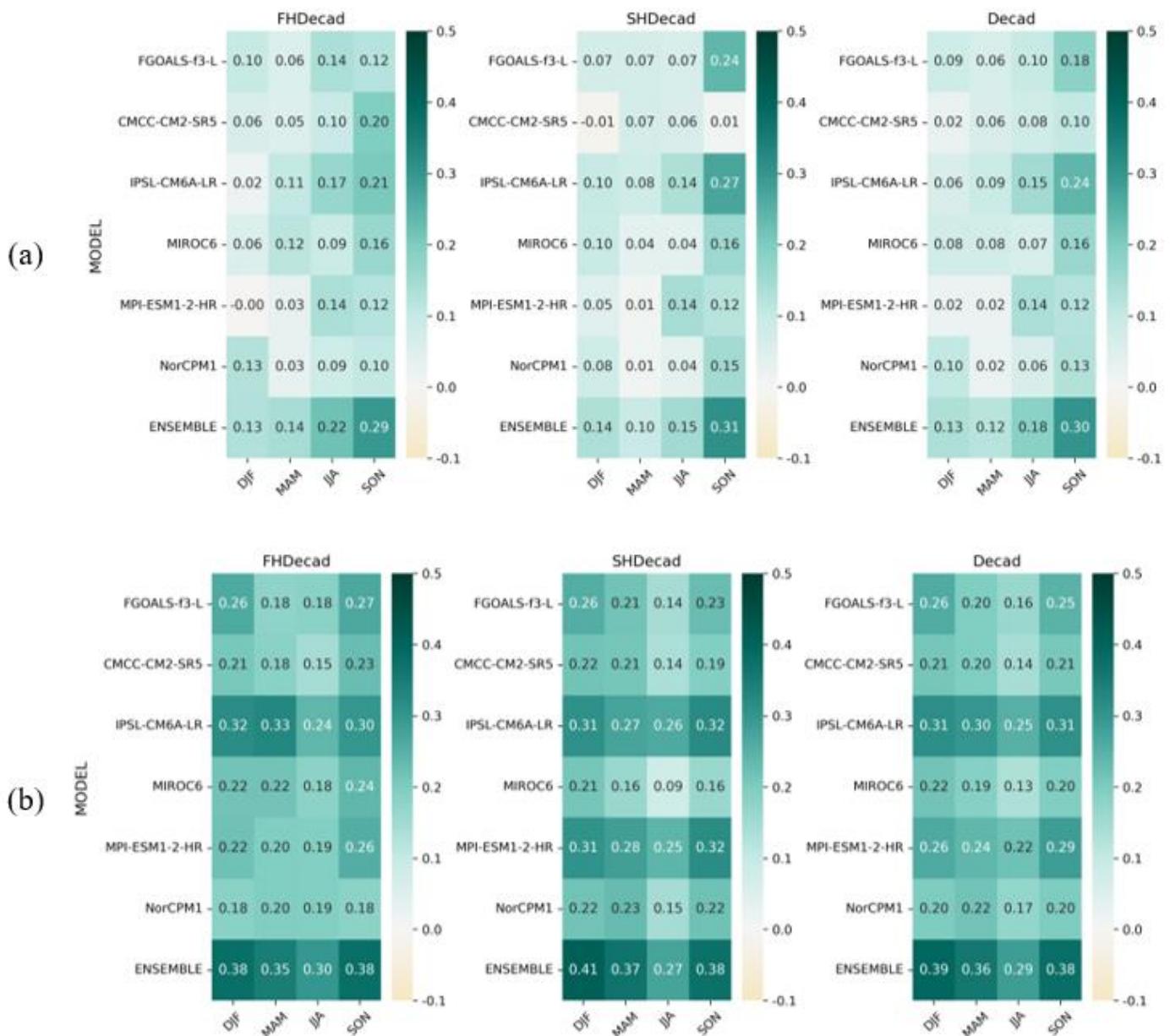


Figure 6. Anomaly Correlation Coefficient (ACC) heatmap for the two subregions (a) mainland and (b) maritime continent.

The PBIAS metric values align closely with the patterns observed in the ME metric. Both metrics consistently identify similar areas where each hindcast model either underestimates or overestimates. Notably, the PBIAS values remain within a narrow range, not exceeding $\pm 2\%$ for any of the models.

The spatial analysis shows that the models have different abilities to predict the two distinct areas, the mainland of Southeast Asia and the maritime continent. Therefore, subsequent examination is carried out by dividing the evaluation value metric into two subregions: the mainland area and the maritime continent area. The mainland subregion is from 93°E to 107°E and 22°N to 12°N while

the maritime continent subregion spread from 95°E to 140°E and 8°N to 12°S. Each model's prediction performance for the two subregions can be observed from the heatmap in Figures 6 and 7.

ACC values in the mainland subregion are more modest compared to those of the maritime continent. Autumn (SON) consistently shows the highest ACC values across models and periods for both subregions. The maritime continent exhibits higher and more consistent ACC values over time, with slight improvements in SHDecad. Among individual models, IPSL-CM6A-LR demonstrates the highest performance. Notably, the ENSEMBLE model successfully increases ACC values in both subregions.

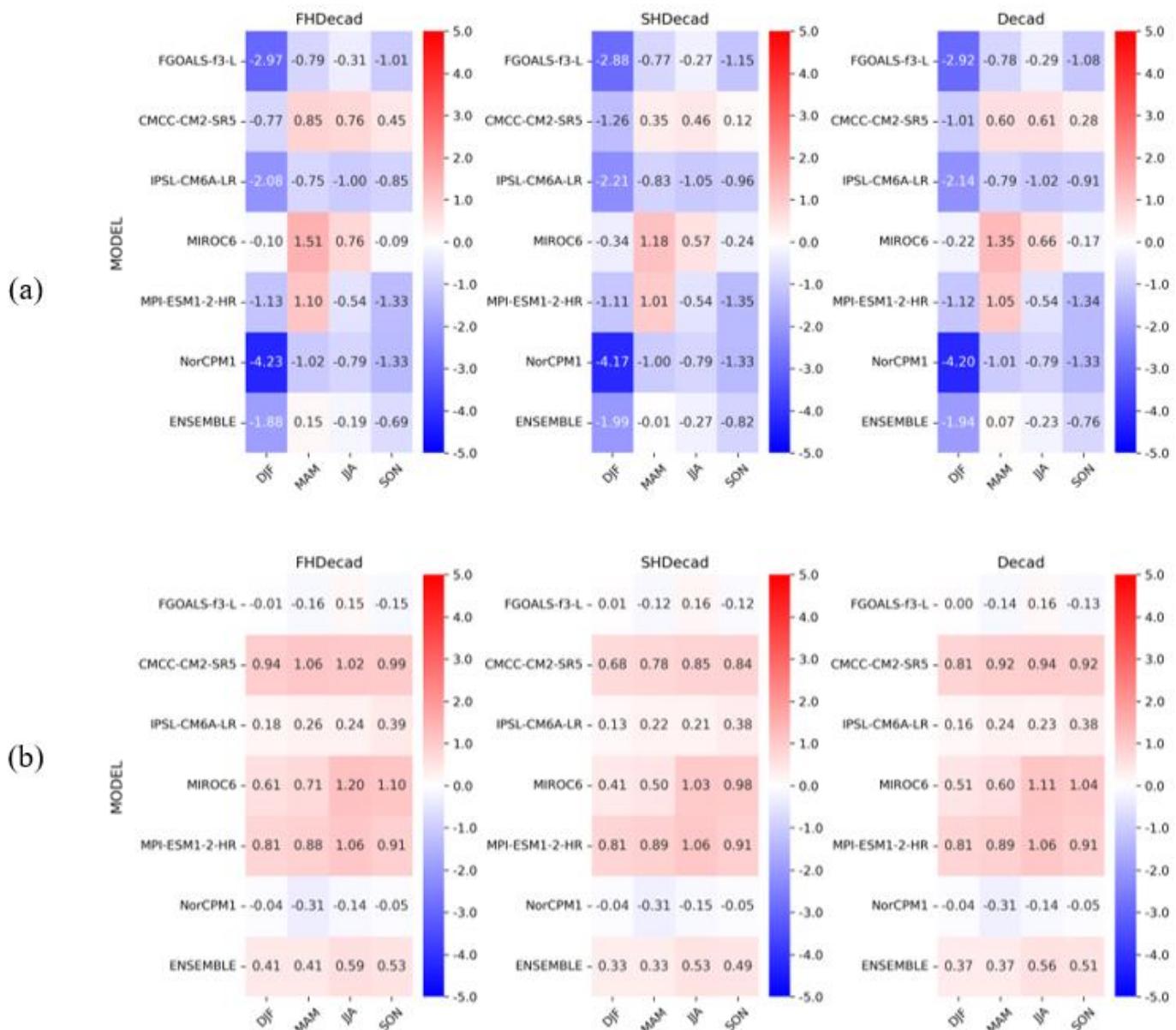


Figure 7. ME heatmap for the two subregions (a) mainland and (b) maritime continent. Unit is in Kelvin.

The ME metric reveals contrasting values for Southeast Asia's mainland and maritime continent subregions (Figure 7). The hindcast datasets tend to predict the area to have a colder temperature than the climatology on the mainland. Opposite to that, in the maritime continent, the models tend to predict warmer conditions compared to the climatology persistently. Winter is the season with the highest ME value, and the ME value for both subregions is marginally higher in the FHDecad period. The model with the highest error for the mainland area is NorCPM1, especially during the DJF season when its error reaches more than 4°K cooler than its climatology reference. While the model of MPI-ESM1-2-HR particularly during the MAM season exhibits a warm bias, MIROC6 tends to have a warm bias during spring and summer, and CMCC-CM2-SR5 extends it to autumn.

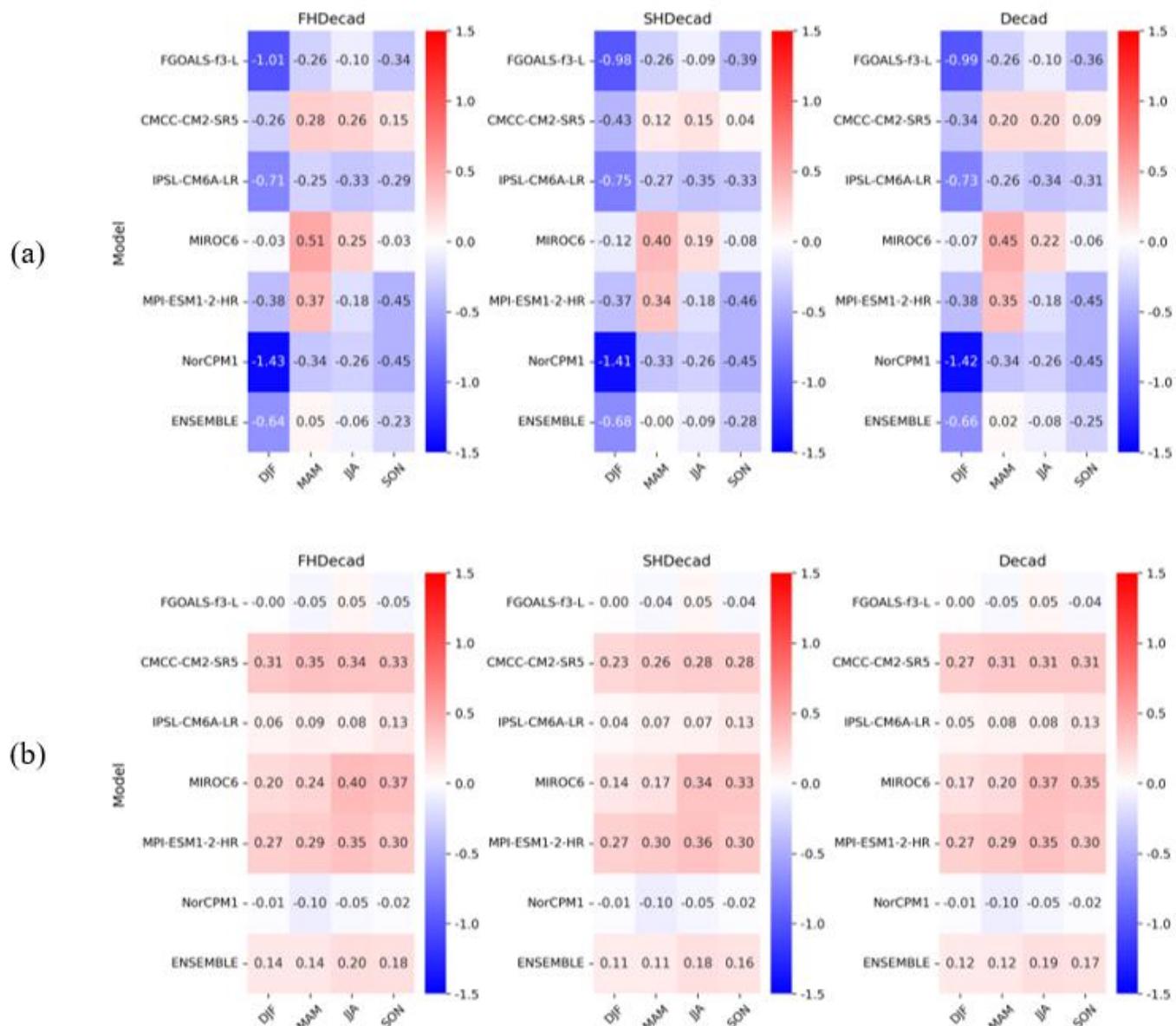


Figure 8. PBIAS heatmap for the two subregions (a) mainland and (b) maritime continent. Unit in percent (%). Negative values represent underestimation, and positive values represent overestimation.

The CMCC-CM2-SR5 and MPI-ESM1-2-HR models display the most pronounced errors when considering the maritime continent subregion. However, the average ME in the maritime continent subregion ranges only from 0.31°K to 1.20°K. The FGOALS-f3-L and NorCPM1 models have a relatively minor cold bias from the climatology reference. In this subregion, the models tend to have more significant biases during the JJA and SON seasons. PBIAS values align with the ME verification metric results. Spatially, PBIAS exhibits the same pattern as ME (maps not shown), indicating areas of underestimation or overestimation in past predictions. Each model consistently underestimates or overestimates across the analysed periods.

PBIAS values in the two subregions range from -1.43% to 0.51% (Figure 8), falling within the "very good" category. The mainland shows more prominent PBIAS values than the maritime continent, especially during the DJF season. Interestingly, the NorCPM1 model, with an ME of -4.23, has a PBIAS of only -1.43%. Similar results are observed for IPSL-CM6A-LR (ME=-2.08, PBIAS=-0.71) and FGOALS-f3-L (ME=-2.97, PBIAS=-1.01). The fact that a 2-4°K error translates to only about 1% bias suggests that the average temperatures in the hindcast datasets are quite low. Although models capture temperature patterns well, as evidenced by small PBIAS values, a 2-4°K underestimation is significant for surface temperature prediction related to human health, particularly for the region of Southeast Asia (Sun et al., 2022).

Furthermore, the analysis examining the influence of initialization on prediction skill reveals that the ACC values of models FGOALS-f3-L, CMCC-CM2-SR5, and MIROC6 (for MAM and JJA

seasons) are higher during FHDecad than SHDecad. In contrast, IPSL-CM6A-LR, MPI-ESM1-2-HR, and NorCPM1 demonstrate improved performance in SHDecad. ME metric indicates higher error values during FHDecad in mainland subregions and a slight decrease in the maritime continent during SHDecad. A consistent cool bias throughout the period across models in the mainland and warm bias in the maritime continent suggests systematic issues in how each model's physical parameterizations or processes respond to these distinct geographical subregions. The PBIAS remains within a very good range throughout the entire period for all models despite these regional biases.

3.2. Regional Predictability of Decadal Climate Prediction Models

There are two sources of forecast predictability in climate models: internal variability and external forcing. The external forcing components stem from factors such as fluctuations in solar irradiance, aerosols from anthropogenic activity and volcanoes, concentration alterations of greenhouse gases, and other external influences. On the other hand, the estimation of climate predictability depends on how climate models respond to these factors and the regions where accurate forecasts emerge on relevant timescales (Meehl et al., 2014).

Notably, the maritime continent subregion exhibits the highest average ACC value over the entire decadal period. In Southeast Asia, the ENSO stands out as the most significant source of internal climate variability, crucial in driving the increasing trend of heatwave events (Lin et al., 2018). Accurately capturing the ENSO signal is essential for climate models aiming to provide reliable

projections for this area. A model's predictive capability in Southeast Asia hinges heavily on its ability to appropriately simulate key oceanic features in the Pacific basin, not only sea surface temperatures (SST) but also ocean heat content, as well as the associated regional variability patterns linked to ENSO dynamics. The proficiency of climate models in representing ENSO-driven internal variability is a critical determinant of their forecast skill for the Southeast Asia region.

Considering FGOALS-f3-L and CMCC-CM2-SR5 models exhibit significant predictability for ocean heat content in regions closely linked to ENSO dynamics, including the tropical Pacific and the subtropical North Atlantic (Hu et al., 2023; Nicoli et al., 2023), they display large errors in the equatorial region. These errors could impact their ability to capture ENSO variability accurately, which is crucial for reliable projections in Southeast Asia. Thus, despite their strengths in specific ocean basins, the limitations in simulating equatorial ocean conditions may hinder the performance of these models in representing ENSO-driven climate variability and its impacts on the Southeast Asia region.

The metrics of ACC for both subregions indicate that models have a relatively similar range of values throughout the decadal period (Figure 6). While ACC values over the maritime continent exhibit greater stability, models FGOALS-f3-L, CMCC-CM2-SR5, and MIROC6 show diminished ability to replicate climatology patterns in the mainland subregion during the latter half of the decadal period (SHDecad) compared to the earlier half (FHDecad).

The decrease in performance can be attributed to the models' diminishing ability to represent the ENSO signal beyond the FHDecad accurately. For instance, the CMCC-CM2-SR5 model's ability to forecast ENSO is limited beyond one year (Nicoli et al., 2023), while the MIROC6 SST hindcast shows a delayed response to ENSO evolution until the second forecast year and overestimates the Niño 3.4 index at longer leads (Kataoka et al., 2020). In contrast to the models showing declining performance, the MPI-ESM1-2-HR and NorCPM1 models demonstrate more consistent predictive skill throughout the decadal period. The MPI-ESM1-2-HR model, however, tends to overestimate ENSO behaviour (Boucher et al., 2020) and NorCPM1 skill degrades for the eastern Pacific area at longer lead times (Bethke et al., 2021). Despite their individual limitations, both MPI-ESM1-2-HR and NorCPM1 models exhibit better overall performance in maintaining their predictive skill across the entire decadal period compared to the other models discussed.

Although ENSO is a crucial internal variability factor for Southeast Asia, climate models also recognize broader timescale climate patterns in the Pacific basin, such as the Decadal Variability of Pacific (PDV) and Multidecadal Variability of Atlantic (AMV), which exhibit distinct teleconnections to the region (Boucher et al., 2020). However, it is worth noting that extreme temperatures in Southeast Asia are not significantly related to these multidecadal variabilities (Fan et al., 2022). Nonetheless, the ability of models to replicate variability patterns in the Pacific region, including ENSO and other decadal and multidecadal climate variables, is crucial for their overall skill in simulating the Southeast Asian climate accurately.

The maritime continent verification metrics display better values than the mainland of Southeast Asia (Figures 6 and 7). Prior research in the region also suggests that models simulate the surface

temperature in the maritime continent subregion more accurately than in the mainland subregion (Kamworapan & Surussavadee, 2019). This improved performance for the maritime continent can be attributed to its proximity to the Indian Ocean, where models exhibit relatively higher decadal predictive skills than other ocean basins (Guemas et al., 2013). This conclusion is further supported by the narrow seasonal variability of surface temperature in the maritime continent region.

4. Conclusion

The assessment of decadal surface temperature predictions provides critical insights for mitigating heatwave events in Southeast Asia. This study evaluated six hindcast models initialized with observational data to examine initialization effects across different prediction timeframes: FHDecad, SHDecad, and Decad. Through comprehensive analysis using ACC, ME, and PBIAS metrics, our findings revealed distinct subregional patterns between mainland and maritime continent.

The results demonstrated superior predictive skill during the first half of the decade (FHDecad), particularly during the SON season, with stronger surface temperature correlations over oceanic regions. This enhanced performance in FHDecad underscores the positive impact of initialization on predictive capabilities. However, we observed a general deterioration in predictive skill over longer lead times, likely due to challenges in representing ocean features and internal variability. Notably, some models maintained or showed improved performance in the latter half of the decade, although this phenomenon warrants careful interpretation regarding its physical basis.

Acknowledgements

The authors would like to thank the Indonesia Endowment Fund for Education Agency (LPDP) for funding this research and the anonymous reviewer for valuable comments.

Author Contributions

Conceptualization: Kasihairani, D., Hidayat, R., Supari, S.; **methodology:** Kasihairani, D., Hidayat, R., Supari, S.; **investigation:** Kasihairani, D., Hidayat, R., Supari, S.; **writing—original draft preparation:** Kasihairani, D.; **writing—review and editing:** Kasihairani, D.; Hidayat, R., Supari, S.; **visualization:** Kasihairani, D.; Hidayat, R., Supari S. All authors have read and agreed to the published version of the manuscript.

Conflict of interest

All authors declare that they have no conflicts of interest.

Data availability

Dataset is provided in this URL: <https://esgf-node.llnl.gov/search/cmip6/>.

Funding

This research was funded by the Indonesia Endowment Fund for Education Agency (LPDP) as part of the given Master's Degree Scholarship Program.

CMIP6 dcpp-A hindcast models demonstrated enhanced predictability of sea surface temperature and ocean heat content in key regions. While the models successfully captured decadal Pacific variabilities and showed promise in replicating ENSO variability patterns, their reliability remains limited in the FHDecad period. Despite achieving favorable ACC values for Southeast Asia, several challenges persist. Future work should focus on refining initialization techniques, improving the representation of internal variability, and reducing systematic biases to enhance the reliability of heatwave projections in this climatically vulnerable region.

References

Bethke, I., Wang, Y., Counillon, F., Keenlyside, N., Kimmritz, M., Fransner, F., Samuelsen, A., Langehaug, H., Svendsen, L., Chiu, P.-G., Passos, L., Bentsen, M., Guo, C., Gupta, A., Tjiputra, J., Kirkevåg, A., Olivé, D., Seland, Ø., Solsvik Vågåne, J., ... Eldevik, T. (2021). NorCPM1 and its contribution to CMIP6 DCPP. *Geoscientific Model Development*, 14(11), 7073–7116. <https://doi.org/10.5194/gmd-14-7073-2021>

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., ... Vuichard, N. (2020). Presentation and Evaluation of the IPSL-CM6A-LR Climate Model. *Journal of Advances in Modeling Earth Systems*, 12(7), e2019MS002010. <https://doi.org/10.1029/2019MS002010>

Corti, S., Weisheimer, A., Palmer, T., Doblas-Reyes, F., & Magnusson, L. (2012). Reliability of decadal predictions. *Geophysical Research Letters*, 39. <https://doi.org/10.1029/2012GL053354>

Fan, Y., Li, J., Zhu, S., Li, H., & Zhou, B. (2022). Trends and variabilities of precipitation and temperature extremes over Southeast Asia during 1981–2017. *Meteorology and Atmospheric Physics*, 134(4), 78. <https://doi.org/10.1007/s00703-022-00913-6>

Guemas, V., Corti, S., García-Serrano, J., Doblas-Reyes, F. J., Balmaseda, M., & Magnusson, L. (2013). The Indian Ocean: The Region of Highest Skill Worldwide in Decadal Climate Prediction. *Journal of Climate*, 26(3), 726–739. <https://doi.org/10.1175/JCLI-D-12-00049.1>

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellán, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., ... Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/qj.3803>

Hu, S., Wu, B., Wang, Y., Zhou, T., Yu, Y., He, B., Lin, P., Bao, Q., Liu, H., Chen, K., & Zhao, S. (2023). CAS FGOALS-f3-L Model Datasets for CMIP6 DCPP Experiment. *Advances in Atmospheric Sciences*, 40(10), 1911–1922. <https://doi.org/10.1007/s00376-023-2122-x>

Kamworapan, S., & Surussavadee, C. (2019). Evaluation of CMIP5 Global Climate Models for Simulating Climatological Temperature and Precipitation for Southeast Asia. *Advances in Meteorology*, 2019, e1067365. <https://doi.org/10.1155/2019/1067365>

Kataoka, T., Tatebe, H., Koyama, H., Mochizuki, T., Ogochi, K., Naoe, H., Imada, Y., Shiogama, H., Kimoto, M., & Watanabe, M. (2020). Seasonal to Decadal Predictions With MIROC6: Description and Basic Evaluation. *Journal of Advances in Modeling Earth Systems*, 12(12), e2019MS002035. <https://doi.org/10.1029/2019MS002035>

Lin, L., Chen, C., & Luo, M. (2018). Impacts of El Niño–Southern Oscillation on heat waves in the Indochina peninsula. *Atmospheric Science Letters*, 19(11), e856. <https://doi.org/10.1002/asl.856>

Meehl, G. A., Goddard, L., Boer, G., Burgman, R., Branstator, G., Cassou, C., Corti, S., Danabasoglu, G., Doblas-Reyes, F., Hawkins, E., Karspeck, A., Kimoto, M., Kumar, A., Matei, D., Mignot, J., Msadek, R., Navarra, A., Pohlmann, H., Riener, M., ... Yeager, S. (2014). Decadal Climate Prediction: An Update from the Trenches. *Bulletin of the American Meteorological Society*, 95(2), 243–267. <https://doi.org/10.1175/BAMS-D-12-00241.1>

Mengistu, A. G., Woldesenbet, T. A., & Dile, Y. T. (2021). Evaluation of the performance of bias-corrected CORDEX regional climate models in reproducing Baro–Akobo basin climate. *Theoretical and Applied Climatology*, 144(1), 751–767. <https://doi.org/10.1007/s00704-021-03552-w>

Müller, W. A., Jungclaus, J. H., Mauritzen, T., Baehr, J., Bittner, M., Budich, R., Bunzel, F., Esch, M., Ghosh, R., Haak, H., Ilyina, T., Kleine, T., Kornblueh, L., Li, H., Modal, K., Notz, D., Pohlmann, H., Roeckner, E., Stemmler, I., ... Marotzke, J. (2018). A Higher-resolution Version of the Max Planck Institute Earth System Model (MPI-

ESM1.2-HR). *Journal of Advances in Modeling Earth Systems*, 10(7), 1383–1413. <https://doi.org/10.1029/2017MS001217>

Nicolò, D., Bellucci, A., Ruggieri, P., Athanasiadis, P. J., Materia, S., Peano, D., Fedele, G., Hénin, R., & Gualdi, S. (2023). The Euro-Mediterranean Center on Climate Change (CMCC) decadal prediction system. *Geoscientific Model Development*, 16(1), 179–197. <https://doi.org/10.5194/gmd-16-179-2023>

Sun, X., Ge, F., Fan, Y., Zhu, S., & Chen, Q. (2022). Will population exposure to heat extremes intensify over Southeast Asia in a warmer world? *Environmental Research Letters*, 17(4), 044006. <https://doi.org/10.1088/1748-9326/ac48b6>

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., Sudo, K., Sekiguchi, M., Abe, M., Saito, F., Chikira, M., Watanabe, S., Mori, M., Hirota, N., Kawatani, Y., Mochizuki, T., Yoshimura, K., Takata, K., O’ishi, R., ... Kimoto, M. (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geoscientific Model Development*, 12(7), 2727–2765. <https://doi.org/10.5194/gmd-12-2727-2019>

Viana, J. F. de S., Montenegro, S. M. G. L., da Silva, B. B., da Silva, R. M., Srinivasan, R., Santos, C. A. G., Araujo, D. C. dos S., & Tavares, C. G. (2021). Evaluation of gridded meteorological datasets and their potential hydrological application to a humid area with scarce data for Pirapama River basin, northeastern Brazil. *Theoretical and Applied Climatology*, 145(1–2), 393–410. <https://doi.org/10.1007/s00704-021-03628-7>