Check for updates

Performance evaluation and ranking of CMIP6 global climate models over Vietnam

Tung Nguyen-Duy [a], Thanh Ngo-Duc [a,*] and Quentin Desmet [b]

[a] LOTUS Laboratory, University of Science and Technology of Hanoi (USTH), Vietnam Academy of Science and Technology (VAST), Hanoi, Vietnam
[b] LEGOS, OMP, UMR 5566 (CNES-CNRS-IRD-University of Toulouse) Toulouse, France
*Corresponding author. E-mail: ngo-duc.thanh@usth.edu.vn

TND, 0000-0002-0186-7146; ThND, 0000-0003-1444-7498; QD, 0000-0001-7438-5765

ABSTRACT

This study comprehensively assesses the performance of 29 Coupled Model Intercomparison Project Phase 6 (CMIP6) Global Climate Models (GCMs) and their ensemble mean (ENS_MEAN) over Vietnam. The spatiotemporal variability of near-surface temperature and precipitation is thoroughly evaluated for the 30-year historical period of 1985–2014. Results show that the models can reasonably reproduce the observational annual cycles and spatial distribution of temperature and precipitation, though their performances vary across the seven climatic sub-regions of Vietnam. Due to their coarse resolutions, the models produce warm biases over certain highland and mountainous areas, and many of them cannot reproduce the rainy season shift from summer in most sub-regions toward the end of the year in Central Vietnam. Finally, these results are summarized, and a performance score is assigned to each model, inferring a ranking. The top three models are EC-Earth3-Veg, EC-Earth3, and HadGEM3-GC31-MM. Although created by simply averaging all the models, ENS_MEAN can compete with the best GCMs and ranks 4th overall. In the same model family, the higher-resolution experiment exhibits a higher ranking than the lower-resolution one. The results of this study can provide a reference for the choice of appropriate CMIP6 GCMs in the upcoming downscaling activities in Vietnam.

Key words: climate change, CMIP6, global climate model, model ranking, Vietnam

HIGHLIGHTS

- The performance of 29 CMIP6 GCMs and their ensemble mean has been analyzed and ranked, for the first time over Vietnam.
- The models reasonably reproduce the observed annual cycles of temperature and rainfall, though their performances may vary across sub-regions.
- The performance in modeling precipitation has been a fairly key point for ranking the models.
- The most appropriate CMIP6 GCMs for Vietnam have been identified.

GRAPHICAL ABSTRACT

(a) Rank for 2m-temperature | (b) Rank for precipitation | (c) Final rank
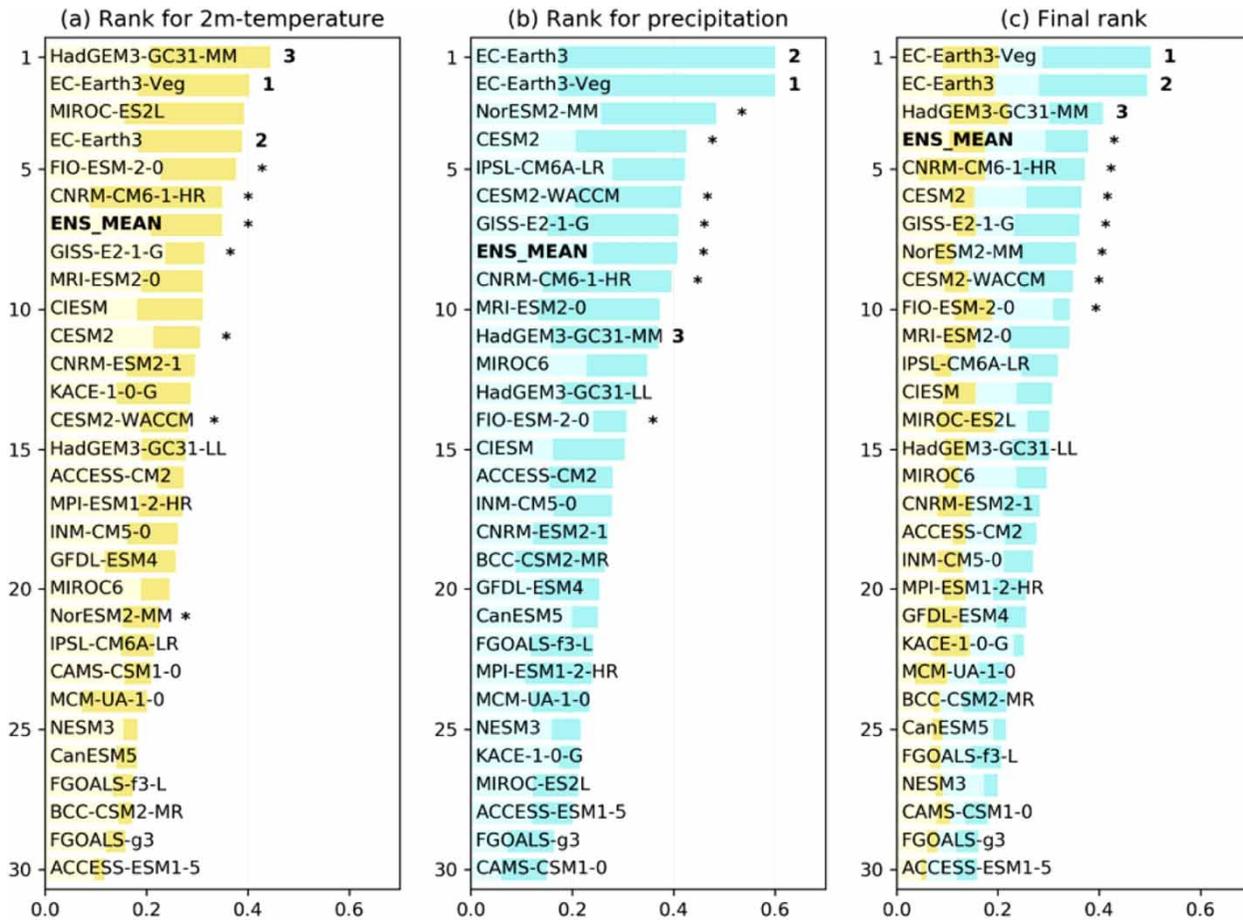
INTRODUCTION

Climate change is taking place worldwide and is resulting in numerous impacts across the planet (Arias *et al.* 2021). In order to conduct climate change impact assessments and devise effective response strategies for a specific region, generating future climate projections for the region is a prerequisite.

Climate projections at the global scale have generally been based on outputs from global climate models (GCMs) under the Coupled Model Intercomparison Project (CMIP). The model results of the CMIP phase 6 (CMIP6) (Eyring *et al.* 2016) have significantly contributed to the recent Sixth Assessment Report (AR6) of the Intergovernmental Panel on Climate Change (IPCC) (IPCC 2021). While CMIP6 GCMs have demonstrated adequate performance in representing historical climate features across various regions of the world (e.g. Seneviratne & Hauser 2020; Srivastava *et al.* 2020; Xin *et al.* 2020; Hong *et al.* 2021), they still exhibit systematic and region-specific biases arising from different sources. For instance, biases in CMIP6 models can be attributed to their representation of sea surface temperature (Wang *et al.* 2021; Tong *et al.* 2022; Rajendran *et al.* 2022), atmospheric circulation (Richter & Tokinaga 2020; Wang *et al.* 2021), land-atmosphere interaction (Abdelmoaty *et al.* 2021; Li *et al.* 2021), cloud processes (Cesana & Del Genio 2021; Wang *et al.* 2021), and other factors. Moreover, a model that performs well in one region may not necessarily perform as well in another region. Therefore, studies to rank the performance of individual CMIP6 models over specific regions have been initially performed (Papalexiou *et al.* 2020; Anil *et al.* 2021; Desmet & Ngo-Duc 2022; Gebresellase *et al.* 2022). Notably, Desmet & Ngo-Duc (2022), hereinafter referred to as DN22, have developed a novel method to rank CMIP6 models over Southeast Asia.

Vietnam is among the countries strongly affected by climate change and sea level rise (Dasgupta *et al.* 2007; Ministry of Natural Resources & Environment 2020). Various studies on climate change in Vietnam have been conducted in recent years. The downscaling of climate information from GCMs for Vietnam has been carried out using statistical (Ministry of

Natural Resources & Environment 2009, 2012; Tran-Anh *et al.* 2022) and dynamical (Ministry of Natural Resources and Environment 2012, 2016, 2020; Ngo-Duc *et al.* 2014; Katzfey *et al.* 2016) approaches. Furthermore, the dynamical experiments that downscaled a number of CMIP5 GCMs under the Coordinated Regional Climate Downscaling Experiment – Southeast Asia (CORDEX-SEA) project (Tangang *et al.* 2020) have also been used to study future climate change characteristics in Vietnam (e.g. Trinh-Tuan *et al.* 2019; Tuyet *et al.* 2019; Nguyen-Thuy *et al.* 2021; Hoang-Cong *et al.* 2022; Vu Dinh *et al.* 2022). So far, the downscaled experiments for Vietnam have only been done with the model outputs of the CMIP phase 5 (CMIP5) and earlier. To our knowledge, some research groups are currently preparing to downscale the latest global outputs from CMIP6 GCMs for Vietnam.

To obtain detailed climate information at the local/regional scales, it is necessary to conduct downscaled experiments. However, this process, especially dynamical downscaling, requires significant computational resources. Therefore, choosing a suitable GCM for further downscaling is crucial to improve the accuracy of local-scale downscaled results and optimize computational resources. In such a context, the present study focuses on analyzing and ranking the performance of 29 CMIP6 GCMs over inland Vietnam. The results are expected to provide helpful information for the community in choosing adequate GCMs in upcoming downscaling studies for Vietnam.

STUDY AREA, DATA, AND RANKING METHODOLOGY

Study area

In this study, we perform the analysis only over inland Vietnam (Figure 1). Located in the eastern part of the Indochinese Peninsula, Vietnam has a long coastline and a complex terrain of which three-quarters are covered by mountainous areas. Nguyen & Nguyen (2004) classified Vietnam into seven climatic sub-regions based on temperature, rainfall, radiation, and sunshine hours. These sub-regions, namely the Northeast (hereafter denoted as R1), Northwest (R2), Red River Delta (R3), North Central (R4), South Central (R5), Central Highlands (R6), and South (R7) are widely accepted by the climate research community in Vietnam (e.g. Phan *et al.* 2009; Ngo-Duc 2014; Trinh-Tuan *et al.* 2019). While we analyze some aspects of the model data separately by sub-region, we also consider inland Vietnam as a whole for other features (see Section 'Model ranking methodology').
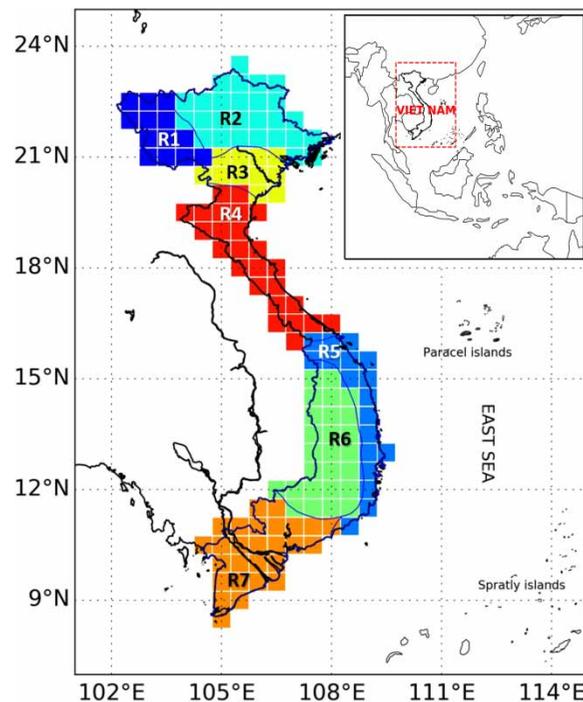

Figure 1 | Inland Vietnam and its seven climatic sub-regions. Grid cells with the size of 0.5° × 0.5° are displayed. The inset map in the upper right shows the location of Vietnam within Southeast Asia.

Data

The capability of 29 CMIP6 GCMs in representing climate features over Vietnam is evaluated through their historical experiment for the 30-year period of 1985–2014. Two variables, namely 2m-temperature (T2m) and precipitation (pr), are considered in this study. Corresponding monthly data are downloaded from the Earth System Grid Federation (ESGF) system. The models we focus on, along with their spatial resolution, are listed in Table 1. Two reference datasets are used (referred to as OBS) to assess the models' performance: the Global Precipitation Climatology Centre (GPCC) gauge-analysis product at the 0.5° × 0.5° resolution (Becker *et al.* 2013) for precipitation, and the Asian Precipitation-Highly Resolved Observational Data Integration Towards Evaluation of Water Resources (APHRODITE, version AphroTemp_V1808) dataset at the 0.25° × 0.25° resolution (Yasutomi *et al.* 2011) for temperature. Both datasets provide land-only information and are derived from quality-controlled station data.

Since the resolution of the GCMs varies significantly, ranging from 0.5° × 0.5° for CNRM-CM6-1-HR to 3.75° × 2.24° for MCM-UA-1-0, their outputs are bi-linearly interpolated to the 0.5° × 0.5° resolution before further analyses and rankings. Moreover, the ensemble mean of all models (referred to as ENS_MEAN) is computed for each variable at the same 0.5° × 0.5° resolution. The chosen resolution allows direct comparison with the GPCC dataset. The APHRODITE dataset is also upscaled to this grid for convenience.

Model ranking methodology

We apply the method proposed by DN22 to rank each model and the ensemble mean based on their ability to reproduce the regional features of near-surface temperature and precipitation (note that we exclude the evaluation of 850-hPa wind, which DN22 performed, due to the narrow study area being unsuitable for a value-added monsoon circulation analysis). One is invited to acquaint oneself with the original design and implications of the DN22 ranking method, which we summarized below.

For each model and variable, a 0-to-1 score is calculated based on the model's ability to represent the variable's observed temporal and spatial patterns. For the temporal aspect, we rely on the assessment of the climatological 12-month seasonal cycle, averaged over each climatic sub-region. We examine these cycles in relation to observations, i.e. GPCC for precipitation and APHRODITE for temperature. To evaluate the accuracy of the cycles, we use metrics including the correlation coefficient (CC), the normalized standard deviation (NSTD), and the mean bias (MB). Spatial patterns are evaluated throughout inland Vietnam as a whole by focusing on two seasonal means: winter (December–January–February; DJF) and summer (June–July–August; JJA). This aspect of the assessment is based on the spatial versions of CC and NSTD.

Table 1 | List of the 29 CMIP6 GCMs, their country of origin, and spatial resolution in degrees of longitude by latitude

ID	Abbreviation	Country	Resolution	ID	Abbreviation	Country	Resolution
1	ACCESS-CM2	Australia	1.88 × 1.25	16	GFDL-ESM4	USA	1.25 × 1.00
2	ACCESS-ESM1-5	Australia	1.88 × 1.25	17	GISS-E2-1-G	USA	2.50 × 2.00
3	BCC-CSM2-MR	China	1.13 × 1.12	18	HadGEM3-GC31-MM	UK	0.83 × 0.56
4	CAMS-CSM1-0	Europe	1.13 × 1.12	19	HadGEM3-GC31-LL	UK	1.88 × 1.25
5	CESM2	USA	1.25 × 0.94	20	INM-CM5-0	Russia	2.00 × 1.50
6	CESM2-WACCM	USA	1.25 × 0.94	21	IPSL-CM6A-LR	France	2.50 × 1.27
7	CIESM	China	1.25 × 0.94	22	KACE-1-0-G	S. Korea	1.88 × 1.25
8	CNRM-CM6-1-HR	France	0.50 × 0.50	23	MCM-UA-1-0	USA	3.75 × 2.24
9	CNRM-ESM2-1	France	1.41 × 1.40	24	MIROC-ES2L	Japan	2.81 × 2.79
10	CanESM5	Canada	2.81 × 2.79	25	MIROC6	Japan	1.41 × 1.40
11	EC-Earth3	Europe	0.70 × 0.70	26	MPI-ESM1-2-HR	Germany	0.94 × 0.94
12	EC-Earth3-Veg	Europe	0.70 × 0.70	27	MRI-ESM2-0	Japan	1.13 × 1.12
13	FGOALS-f3-L	China	1.25 × 1.00	28	NESM3	China	1.88 × 1.88
14	FGOALS-g3	China	2.00 × 2.03	29	NorESM2-MM	Norway	1.25 × 0.94
15	FIO-ESM-2-0	China	1.25 × 0.94				

The statistics listed above constitute the basis of the scoring process. Each x value calculated for CC, NSTD, or MB is then transformed to X as follows:

$$X = \begin{cases} x - 1, & \text{for CC} \\ -|x - 1|, & \text{for NSTD} \\ -|x|, & \text{for MB} \end{cases} \tag{1}$$

As can be seen, X is of the interval $[-\infty, 0]$, and the greater X, the better performance. It allows the score $S(X)$ to vary in $[0, 1]$ and to increase with the model's reliability:

$$S(X) = \beta^{\sigma(X + \delta)} \tag{2}$$

where the tolerance δ, the scale factor σ, and the base β are all dependent on the ensemble E of the 30 X results (i.e. 29 GCMs plus the ensemble mean) given a variable, an aspect (temporal/spatial), a sub-region/season, and a statistic. Several conditions ruling the choice of δ and σ have been mentioned in DN22. We opt to use the same expressions as originally proposed:

$$\delta = |\max_E(X)| \tag{3}$$

$$\sigma = |\text{ave}_E(X + \delta)|^{-1} \tag{4}$$

This choice of δ implies that the best performance in E gets $S = 1$. β is selected at last in $[1, +\infty]$, so that the standard deviation of the scores for E is maximum.

At this stage, each statistical criterion has resulted in one 0-to-1 score per model. Global scores are ultimately obtained by successively averaging (arithmetic mean) the scores from the most to the least specific distinctions in the assessment, i.e. in the order of one score per statistic CC, NSTD, and MB; one score per sub-region (season); one score for the temporal (spatial) aspect; one score per variable; and eventually, one global model score. Note that these scoring results are only meaningful for ranking this specific model list, as the scoring formula depends on all outputs' characteristics and would give different scores to the same models in another ensemble.

RESULTS

Annual cycle

Figure 2 shows the 2m-temperature annual cycles over the seven sub-regions of Vietnam for the GCMs, ENS_MEAN, and OBS. In Figure 3, these temperature series' statistics are summarized using Taylor diagrams (Taylor 2001). Each symbol corresponds to one model's performance with regard to its NSTD (radial distance) and CC (polar angle) when compared with OBS (NSTD = 1; CC = 1). Additionally, the distance between a model's symbol and the OBS point is proportional to the centered root mean square difference (RMSD) between the two series: the closer to the OBS point, the better performance.

In general, the models do a relatively good job of representing OBS across sub-regions, especially by capturing the annual temperature range, whether it is relatively large in the north (R1–R4; around 15 °C), or small in the south (R5–R7; 5–10 °C; Figure 2). However, the Taylor diagrams (Figure 3) highlight an ensemble trend in which the annual range is simulated slightly too large, as most GCMs have an NSTD over 1. The ensemble mean's cycle in Figure 2 allows us to identify the causes of this ensemble increased variability: most GCMs underestimate autumn–winter temperatures (except in R6), and warmer summers are generally modeled in R1 and R6. Notably, a strong positive MB is computed for the ensemble in the last sub-region, Central Highlands, which lies in a series of plateaus with heights ranging around 500–1,500 m. This warm bias is caused by the fact that GCMs with coarse resolutions often represent lower-than-real topography over highland areas, hence failing to fully reflect the real environmental lapse-rate, i.e. the decrease of temperature with height, in their simulations.

On another note, some individual results distinguish themselves, such as the northern sub-regions' cycles (R1–R4) of MCM-UA-1-0, which show the lowest CCs (~0.7–0.8). Indeed, related curves in Figure 2 exhibit a significant overestimation in the first few months of the year, causing the annual peak to be reached 2 months earlier than observed. On the other hand, a few models show high CC ($\geq$0.95) and close-to-1 NSTD, albeit not uniformly across sub-regions: EC-Earth3 produces satisfactory
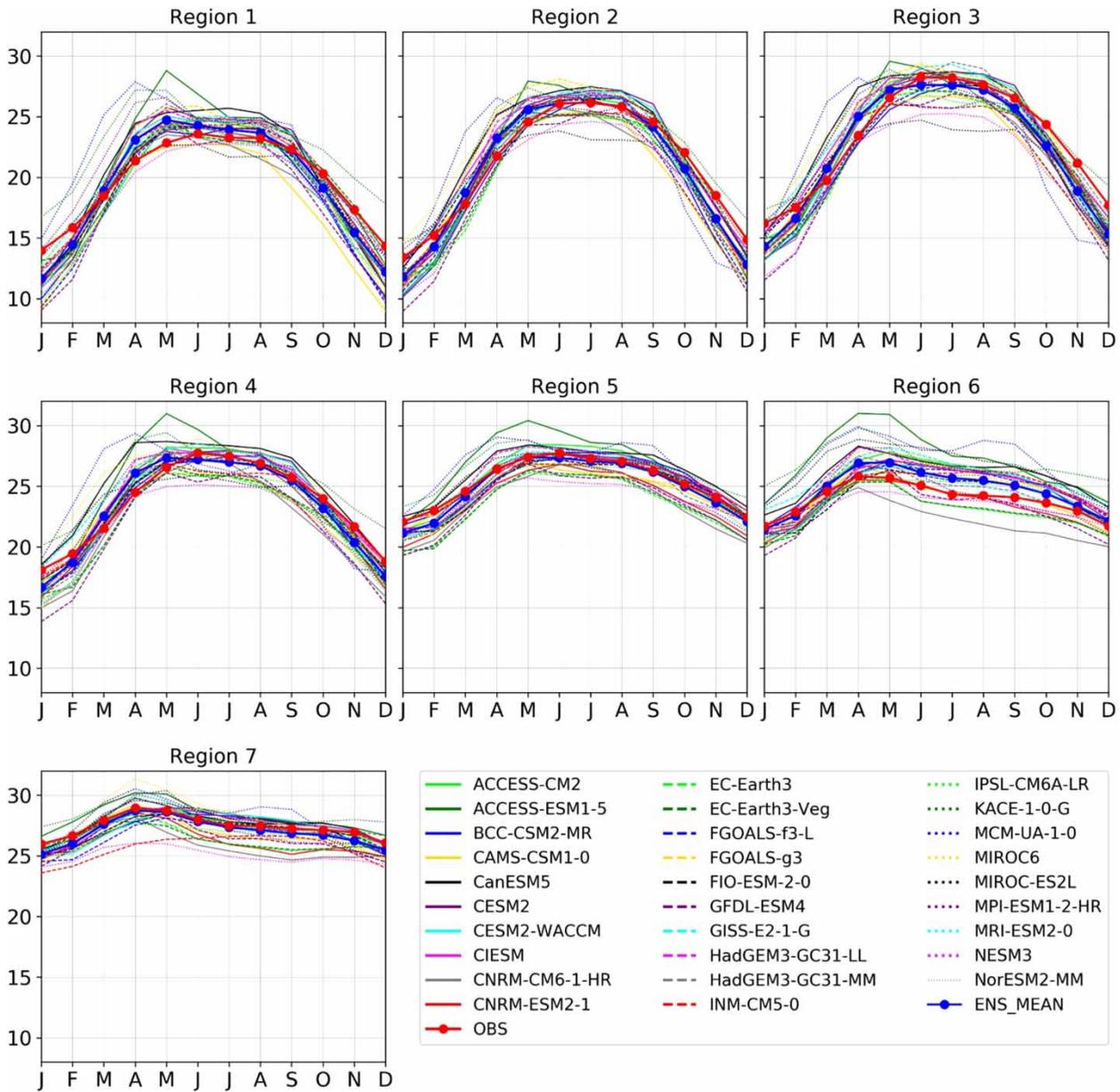
Figure 2 | Annual cycles of temperature over the seven sub-regions of Vietnam, represented by the GCMs and the reference dataset OBS.

results in R1–R3; FIO-ESM-2-0 outperforms in R4; so do GISS-E2-1-G in R5; NESM3 and EC-Earth3-Veg in R6; and ACCESS-CM2 in R7. Models with higher resolution, such as CNRM-CM6-1-HR, EC-Earth3, EC-Earth3-Veg, and HadGEM3-GC31-MM, seem to stand out in most sub-regions, except in R7, where the flat terrain does not particularly benefit from finer grids. Additionally, ENS_MEAN performs better than most models, especially in R5 and R7 (where it exhibits almost no bias; see Figure 2). This suggests that averaging the models can absorb models' outliers, thus improving the results. Overall, the best GCMs for reproducing the annual cycle for 2m-temperature are GISS-E2-1-G, FIO-ESM-2-0, and ACCESS-CM2.

Similarly, Figure 4 shows the annual precipitation cycles over the seven sub-regions of Vietnam for the GCMs and OBS, while Figure 5 gives a synthetic view of their statistics.

Over the sub-regions R1–R3, OBS values exhibit the summer monsoon's characteristics with its peak rainfall occurring in JJA (Figure 4). Most models show high correlation (CC > 0.9, Figure 5), but not all of them can reproduce the reference
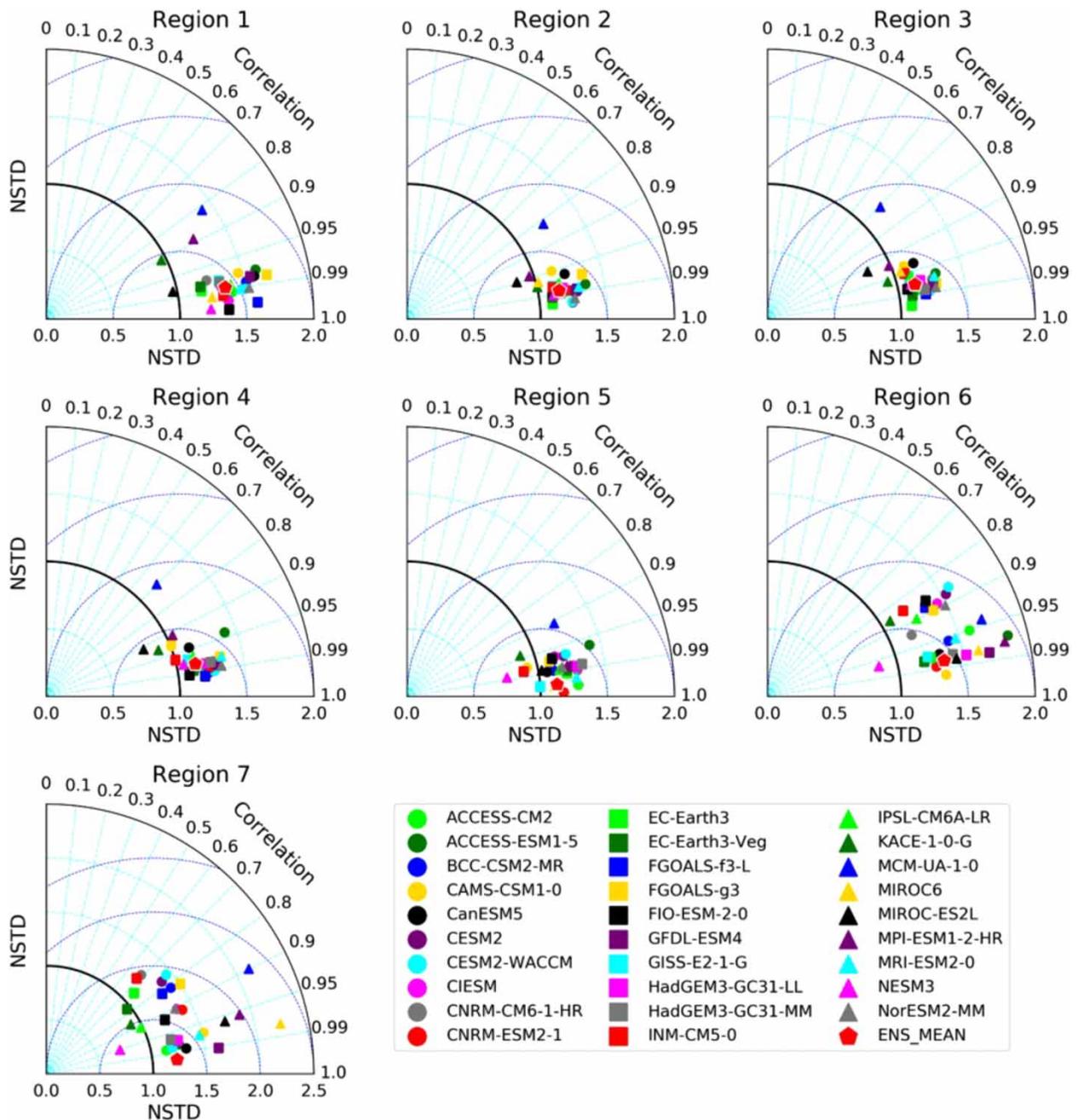
Figure 3 | Taylor diagrams of the GCMs' temperature annual cycles over the seven sub-regions of Vietnam.

amplitude well. Several models overestimate precipitation in summer, e.g. INM-CM5-0 and MCM-UA-1-0 (Figure 4), leading to high NSTDs of up to 1.4 (Figure 5). On the contrary, some outputs exhibit unreasonably low rainfall values, such as CAMS-CSM1-0 with NSTD less than 0.5. ENS_MEAN is among the best models in these regions, with CC > 0.95 and NSTD > 0.9.

In Central Vietnam (R4–R5), the rainy season gradually shifts towards the end of the year (Figure 4) because of the inter-action between the winter monsoon and the nearby mountain range (Ngo-Duc 2014; Nguyen-Le *et al.* 2015). By contrast with the overall satisfying representations of temperature analyzed above, in the area, the GCMs face a challenge in simulating observed precipitation patterns (particularly this rainy season shift), as many models have CCs lower than 0.7, or even down to 0.3 (Figure 5). For example, MCM-UA-1-0 still depicts a rainfall peak in May and June (Figure 4). A possible reason lies in its coarse spatial resolution (3.75° × 2.24°) leading to the failure to capture the fine-scale interaction between
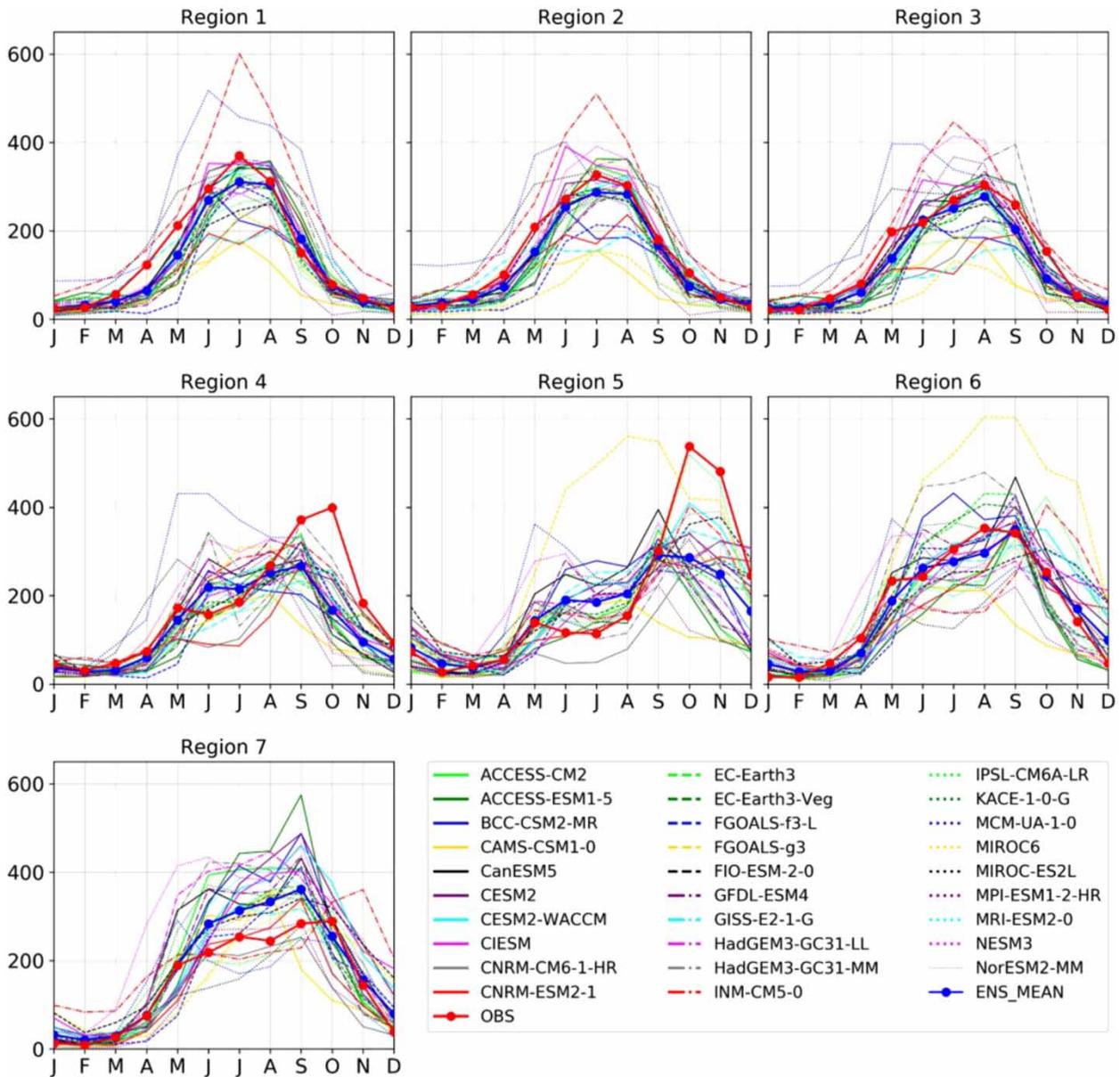
Figure 4 | Same as Figure 2 but for rainfall.

the monsoon and the local topography. On another note, most models underestimate the precipitation variability, with NSTD < 1, except for KACE-1-0-g, NESM3 and MCM-UA-1-0 over R4, and for FGOALS-g3 over R5 (Figure 5). The highest performance models in R4 are CNRM-CM6-1-HR, FIO-ESM-2-0, and IPSL-CM6A-LR, with CC ~ 0.95 and NSTD ~ 0.7; and in R5, IPSL-CM6A-LR stands out with CC = 0.99 and NSTD ~ 1. Ensemble averaging smooths the rainfall values from models that well represent the interaction between the winter monsoon and local topography, thus ENS_MEAN does not outperform the above models over these sub-regions.

In R6–R7, the OBS values show the monsoon with the peak rainfall occurring in summer and maintaining high values for a longer period than in R1–R3 (Figure 4). Overall correct phase performances are observed over these sub-regions, with CCs mostly greater than 0.8, except for INM-CM5-0 (Figure 5). However, simulated amplitudes are diverse. In R6, while some models strongly overestimate this feature (e.g. FGOALS-g3 with NSTD ~ 1.8), others show underestimation (e.g. MPI-ESM1-2-HR with NSTD ~ 0.6). Additionally, in R7, a strong trend of overestimated variability is observed: NSTDs are
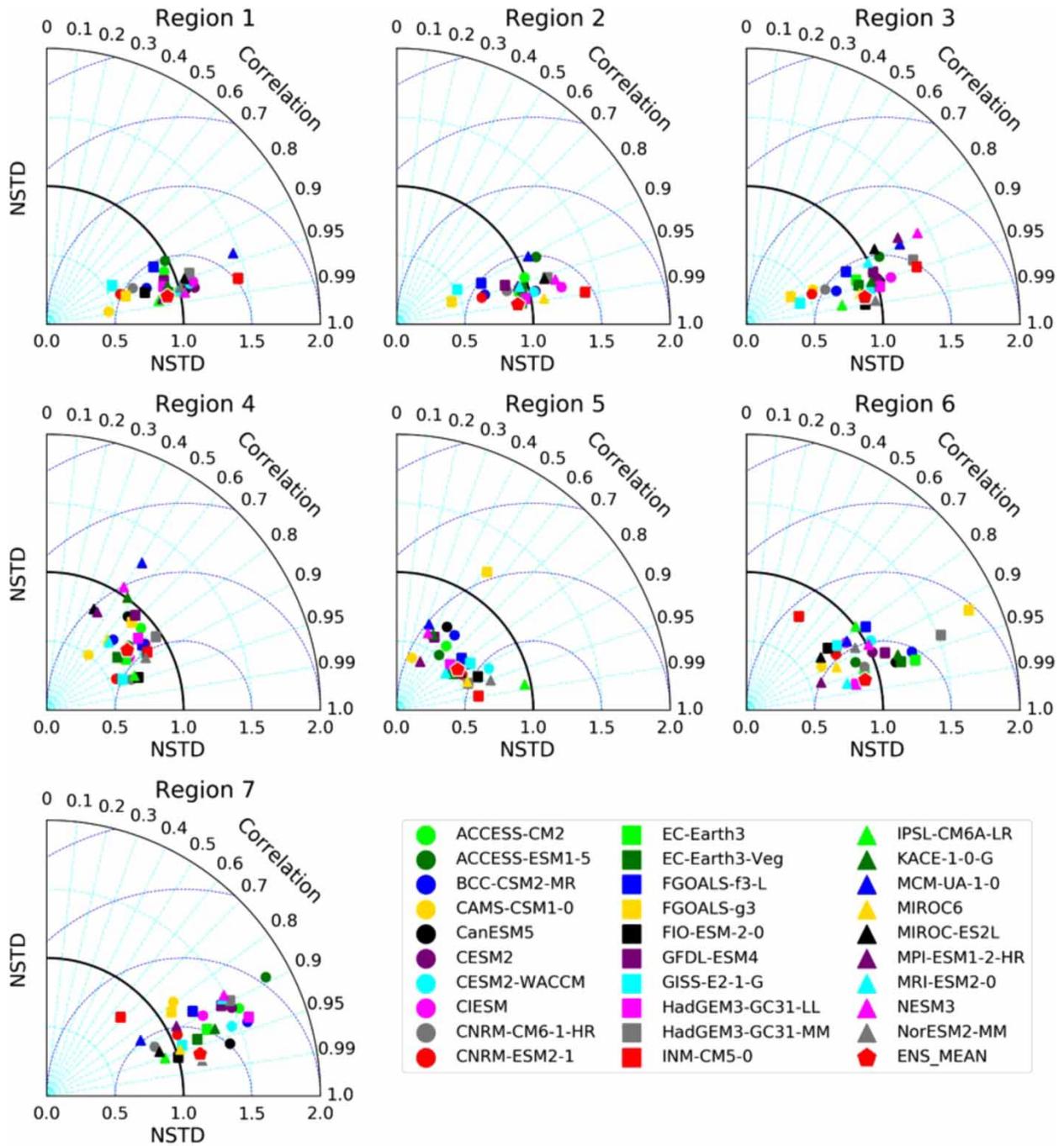
Figure 5 | Same as Figure 3 but for rainfall.

mostly over 1 (up to 1.8 with ACCESS-ESM1-5), and this is naturally the case of the ensemble mean as well. This trend is linked to an ensemble overestimation of rainfall in summer (Figure 4).

Overall, some models with good simulation results across all seven sub-regions (i.e. high CC and close-to-1 NSTD) particularly stand out: CESM2-WACCM, FIO-ESM-2-0, IPSL-CM6A-LR, and NorESM2-MM. On the other hand, some models are reliable in one sub-region but less in another. For instance, HadGEM3-GC31-LL displays satisfactory results over R6 but exhibits much larger variability than OBS over R7 (NSTD > 1.6). Lastly, ENS_MEAN is among the best models in most of the sub-regions, except in R4 and R5 where many GCMs have difficulty representing the rainy season shift (Figure 4). When

taking into account MB, GCMs with highest temporal accuracy for rainfall are IPSL-CM6A-LR, NorESM2-MM, and FIO-ESM-2-0.

Spatial variability

The spatial analysis has been conducted identically for each model, so that final rankings could be drawn and the best GCM identified for both variables (Section 'Model ranking and discussions'). Yet, the models' spatial performances are presented in this section with a focus on the best GCM, ENS_MEAN, and the ensemble spread.

Figure 6 shows the spatial variability of winter (DJF) and summer (JJA) temperatures given by OBS, the best GCM for temperature (HadGEM3-GC31-MM as depicted in Figure 8(a), denoted as GCM_BEST) and ENS_MEAN, along with the standard deviation of the 29 GCMs. During the DJF months, the observational temperature increases from approximately 14 °C in the north to around 28 °C in the south. The lower temperature in the north of Vietnam is attributed to the cold north-easterly wind during the winter monsoon, while in the south, the temperature is influenced by the equator climate. The DJF spatial temperature pattern is well represented by GCM_BEST and ENS_MEAN, with NSTDs of 1.08 and 1.09, and CCs of 0.98 and 0.97, respectively. Both GCM_BEST and ENS_MEAN underestimate temperature over Vietnam, particularly over the northern regions R1–R4, by around 2 °C, which has been previously highlighted with sub-regional means in Figure 2. The models cannot accurately capture the low northwest-southeast temperature stripe in between the sub-regions R1 and R2, which coincides with the location of Hoang Lien Son, the highest mountain range in Vietnam. The warm bias of the models over this mountain range, similar to what happens in R6 (Figure 2), also owes to their coarse spatial resolutions; hence the decrease in temperatures due to the environmental lapse-rate cannot be simulated effectively. The ensemble
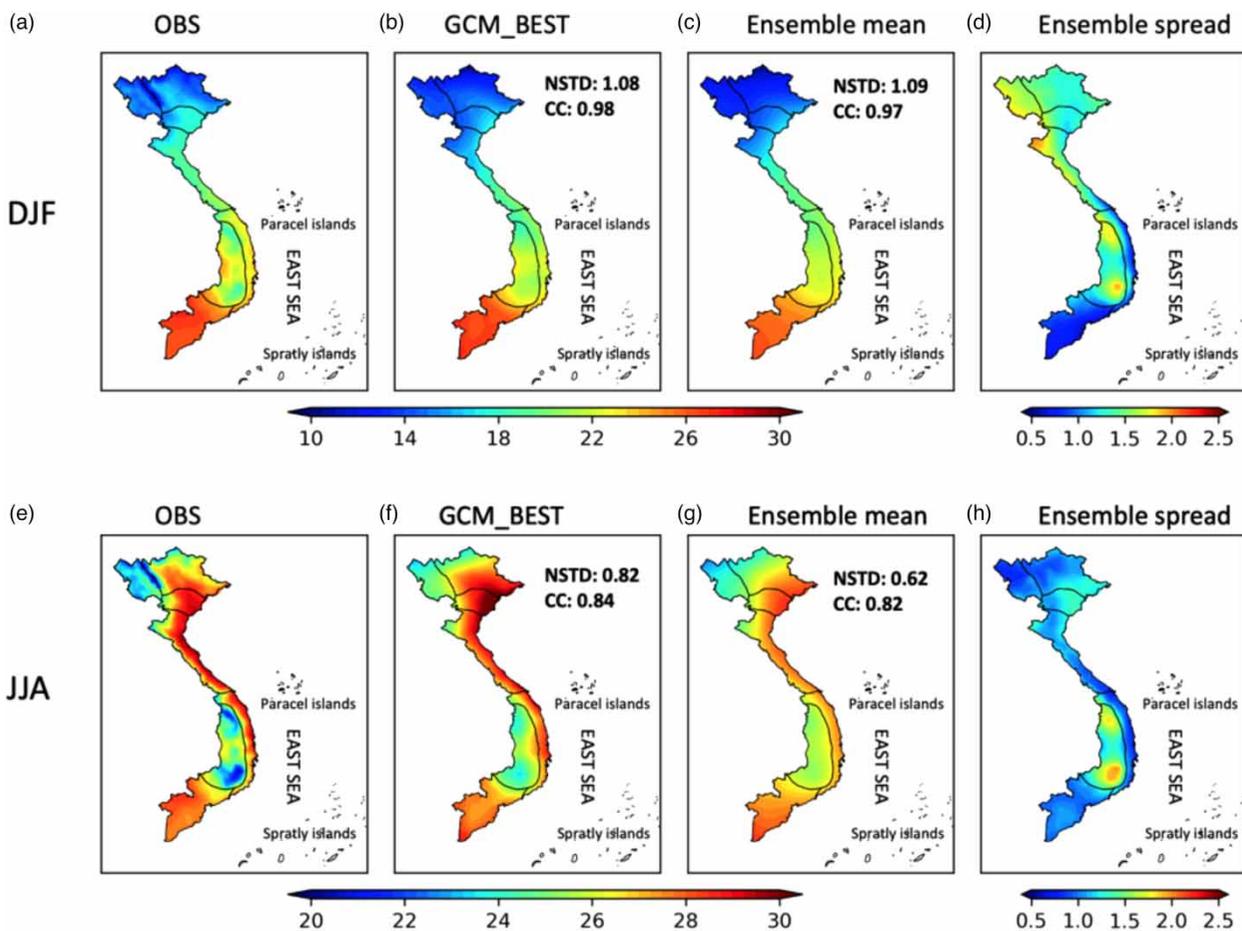

Figure 6 | Spatial distribution of DJF (upper panel) and JJA (lower panel) temperatures for (a,e) OBS, (b,f) the GCM that best represents temperature, (c,g) the ensemble mean, and (d,h) the standard deviation of all 29 models. Units in °C.

spread (Figure 6(d)), which can be used as an indicator of the similarity among models, indicates that temperature outputs are highly consistent in R5 and R7, where the ensemble spread is less than 1 °C, but more varied in other regions with more complex topography, such as Central Highlands and the western mountainous areas of Vietnam.

During the JJA months, the observed temperature is high (28–30 °C), except in sub-regions R1 and R6 (22–26 °C). Although GCM_BEST tends to overestimate temperature in certain sub-regions, such as R1, R3, and R6, by 1–3 °C, and still cannot accurately resolve the cold temperature features induced by the mountain ranges in the northwest and Central Highlands, its performance is generally good with 0.82 NSTD and 0.84 CC. The ensemble mean tends to smooth out the local temperature features, leading to the overall decrease of NSTD to 0.62 though its CC is still comparable with GCM_BEST (0.82). The ensemble spread indicates higher values of 1.5–2 °C in Central Highlands (R6), suggesting the high uncertainty of the GCMs in this sub-region whatever the season. On the contrary, the spread in JJA differs by around 1 °C from that of DJF in northern Vietnam: there is a broader consensus among the models regarding summer temperature in sub-regions R1–R4.

Figure 7 is similar to Figure 6 but for precipitation. In DJF, the observed precipitation in Vietnam is typically less than 50 mm/month, except for the central coastal region of R4–R5, where monthly rainfall can reach as high as 160–200 mm (Figure 7(a)). GCM_BEST (i.e. EC-Earth3 depicted in Figure 8(b)) demonstrates good agreement with the observations, with NSTD and CC values of 0.92 and 0.93, respectively, though it slightly underestimates R4–R5 precipitation peaks (Figure 7(b)). The ensemble mean, on the other hand, cannot represent these maximums at all, as the amount of rainfall appears to be spread over nearby sub-regions and even down to R7 (Figure 7(c)). This goes along with lower values of both NSTD (0.61) and CC (0.72). The lower performance of ENS-MEAN is further supported by the high ensemble standard deviation in southern Vietnam. In particular over R6 and R7, the uncertainty reaches 30–50 mm/month, which is in the same order of magnitude as the OBS values.
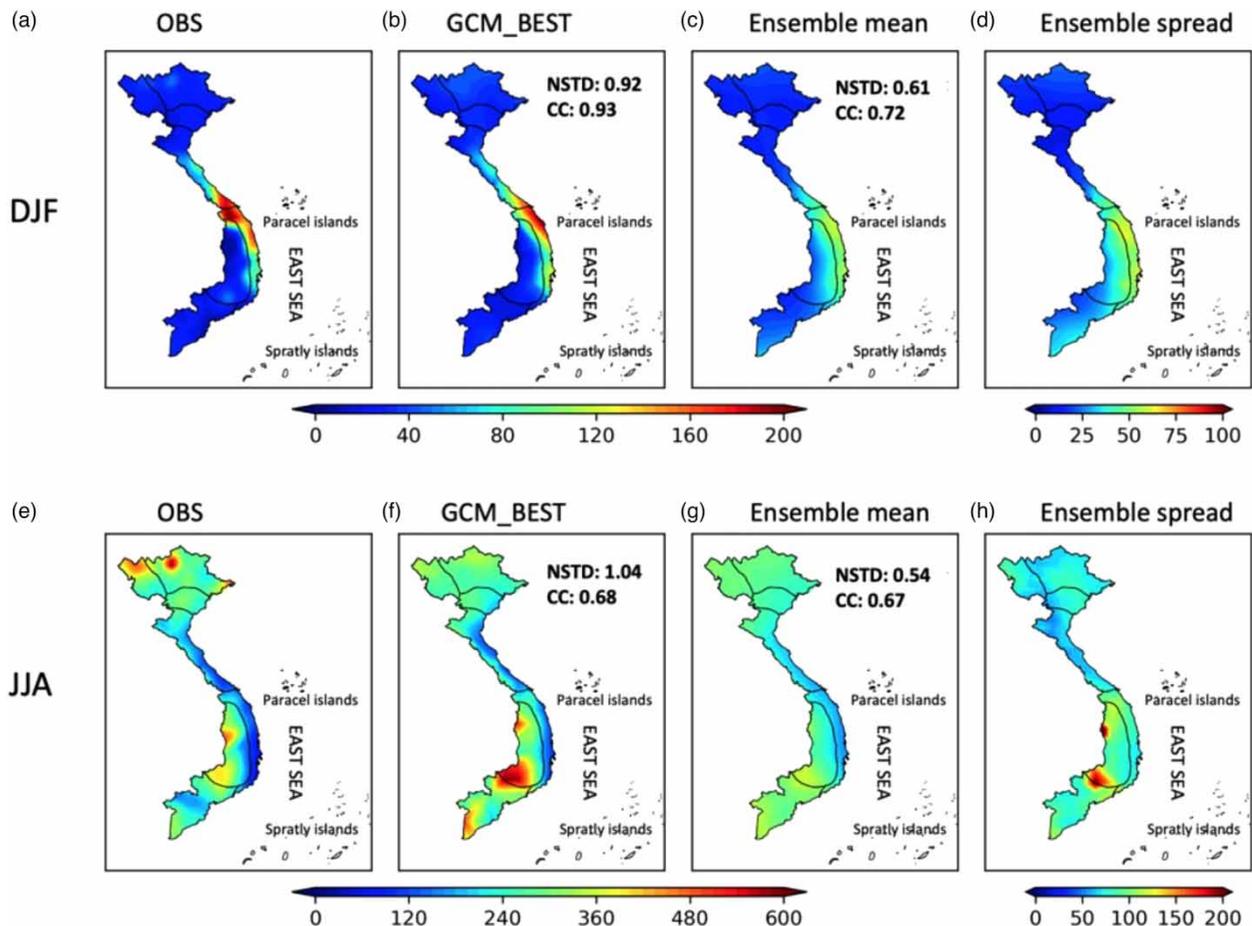

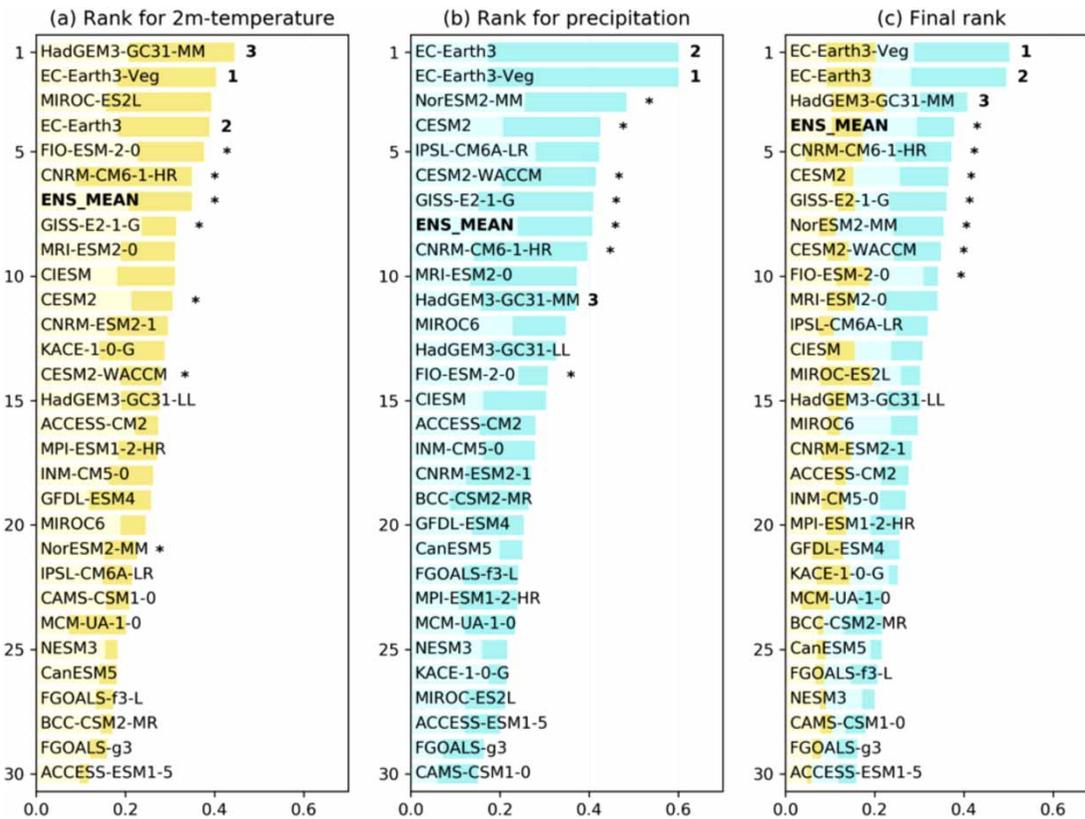**Figure 7** | Same as Figure 6 but for precipitation. Units in mm/month.

Figure 8 | Scores and ranks of 29 CMIP6 GCMs and their ensemble mean for (a) 2m-temperature (b) precipitation, and (c) all. Light and dark yellow highlight the contribution of temporal and spatial scores for temperature, respectively, while two shades of cyan are used similarly for precipitation. The top three models for Vietnam are labeled with numbers and the next seven models are marked with an asterisk. Please refer to the online version of this paper to see this figure in colour: https://dx.doi.org/10.2166/wcc.2023.454.

In summer, the precipitation is much higher with a minimum of 100 mm/month in R4 and R5 sub-regions, which is the opposite of the winter spatial pattern. The maximum precipitation of 500 mm/month can be observed in R1, R2, and R6. GCM_BEST demonstrates a reliable large-scale distribution; however, it fails to reproduce local scale patterns, with the local maximums in R1 and R2 being ignored, while those of R6 are overestimated. This discrepancy results in a relatively low CC of 0.68. Furthermore, the ensemble mean exhibits even poorer applicability as it smoothes out most of the spatial variability (NSTD = 0.54) (Figure 7). The ensemble spread shows high uncertainty everywhere, ranging from a minimum of 70 mm/month to 200 mm/month in R6.

Model ranking and discussions

The final ranking of the 29 CMIP6 GCMs and the ensemble mean for inland Vietnam is presented in Figure 8. The figure clearly shows the top two models, which are EC-Earth3-Veg and EC-Earth3, followed by HadGEM3-GC31-MM and ENS_MEAN (Figure 8(c)). It is worth noting that of the three best models, one excels in temperature (HadGEM3-GC31-MM), one in precipitation (EC-Earth3), and one is the second-best for both temperature and precipitation (EC-Earth3-Veg). The top ten models are generally among the best for both variables' rankings (Figure 8(a) and 8(b)). However, there are exceptions due to offsets between the two variable scores of a model. For instance, NorESM2-MM ranks 8th overall because of its top-three precipitation performance, compensating for its lower temperature score (ranking 21st). On the other hand, HadGEM3-GC31-MM ranks poorly for precipitation (11th) but catches up with a top-one temperature performance, so it eventually ranks 3rd. Among the top ten models, those ranking in the top ten for both variables are EC-Earth3-Veg, EC-Earth3, ENS_MEAN, CNRM-CM6-1-HR, and GISS-E2-1-G.

It is worth mentioning that the model resolution appears to have an impact on the final ranking results. Models with higher resolution, such as EC-Earth3-Veg (0.7° × 0.7°), EC-Earth3 (0.7° × 0.7°), HadGEM3-GC31-MM (0.83° × 0.56°), and

CNRM-CM6-1-HR (0.5° × 0.5°) are in the top-ranked models. Moreover, in the same model family, the higher-resolution experiment tends to achieve a higher ranking. For example, HadGEM3-GC31-MM (0.83° × 0.56°) ranks 1st, 11th, and 3rd, respectively for temperature, precipitation, and globally, while HadGEM3-GC31-LL (1.88° × 1.25°) ranks 15th, 13th, and 14th, in the same order. This aspect is consistent with the findings of DN22.

While ENS_MEAN was created by simply averaging all the models, it competes with the best GCMs and ranks 4th overall. Interestingly, its final rank is even higher than its component ranks (7th and 8th for temperature and precipitation, respectively). Yet, it does not outperform all individual models, which differs from the conclusion of Li *et al.* (2022). We also computed additional integrated models using their method, but these models did not outperform any of the available GCMs in Vietnam. Most of the time, the integrated models underestimated spatial variability and had a lower correlation, resulting in a lower score. This divergence could be due to the design of our ranking methods and the fact that our study area is smaller, containing only 147 grid points, while Li *et al.* (2022) used more than 87,000 grid points over global land.

Figure 8 also reveals the temporal and spatial scores of each GCM. For temperature, the highest-applicability models seem to be supported by a balanced contribution of both aspects. However, from the 14th GCM, the temporal score appears to dominate. This is linked with the spatial/temporal difference in score variability observed for both variables with the shadings in Figure 8(a) and 8(b). In particular, EC-Earth3 and EC-Earth3-Veg stand out in simulating precipitation, owing to their substantially higher spatial performance, weighing more than two-thirds of their precipitation score and around 43% of their final one. These above comments could lead to the conclusion that the models' spatial performance unfairly weights more than the temporal aspect. Indeed, it is noteworthy that the temporal branch of the scoring process goes through more assessment criteria than the spatial branch (notably: seven sub-regions against two seasons), making it statistically less probable to score high after the multiple averages of the former branch. Yet, the selected statistical coefficients convey interrelated features of the GCMs' outputs, and more ranking criteria provide more opportunities to stand out. In that respect, when a scoring process results in a smaller range (i.e. here the temporal scoring), it means that the assessed features do not allow one model to differentiate, indicating a strong consensus among the ensemble members or compensations of good and bad performances throughout the evaluation. As a result, this scoring branch may account for less than the others. Furthermore, the precipitation score's contribution to the final rank fluctuates more than that of temperature assessment (precipitation scores vary in a larger range), thus having more influence. But in this case, the number of ranking criteria is the same for both variables, and the unbalance necessarily comes from the GCMs themselves. Therefore, in line with Sections 'Annual cycle' and 'Spatial variability' where several challenges in modeling precipitation over Vietnam have been discussed, it can be concluded that the performance in modeling precipitation (in particular its spatial features) has been a key point for ranking this GCM ensemble.

EC-Earth3-Veg, EC-Earth3, HadGEM3-GC31-MM, CNRM-CM6-1-HR, CESM2, GISS-E2-1-G, and FIO-ESM-2-0 are all in the top ten of both SEA (DN22) and Vietnam ranking. This introduces diverse links between this study and the results from DN22 (in this paragraph, rank comparisons are made with adjustments when ignoring the positions of ENS-MEAN and HadGEM3-GC31-LL, which were not part of the ranking in DN22). In particular, DN22 conducted an analysis of the same models' performance in capturing the regional wind circulation, which should be considered to complete this study's ranking. As such, satisfying representation of SEA general circulation supports the good Vietnam ranks of EC-Earth3-Veg, EC-Earth3, CNRM-CM6-1-HR, and CESM2; but not for HadGEM3-GC31-MM, which showed below-average wind performance. However, reliable wind should not outweigh the other variables in decision making. MPI-ESM1-2-HR is a good example of a model that produced excellent wind (ranked 3rd for SEA in DN22) but ranks average in the end (11th for SEA and 18th for Vietnam). Furthermore, the top six wind patterns of GFDL-ESM4 and FGOALS-f3-L have not been particularly helpful for scoring high over Vietnam, as one can notice a significant downgrade between their SEA and Vietnam final positions (7th to 19th and 4th to 24th, respectively). In addition to the fact that in the present study, they do not benefit from their wind score anymore, these rank changes also seem to originate from a substantial performance gap when looking at their annual cycles between the Maritime Continent (i.e. the near-equatorial region constituted with Malaysia and Indonesia) and the Indochinese Peninsula (these models performing better over the former). Conversely, reverse performance asymmetries (i.e. better applicability over Indochina than over the Maritime Continent) have been responsible for upgrading the ranks of, notably, CESM2-WACCM (17th to 8th) and NorESM2-MM (15th to 7th), whose general circulation performances were average. Lastly, if one is interested in using the outputs from IPSL-CM6A-LR and MRI-ESM2-0 in the region, their spatial performance over SEA should be taken into account, as precipitation and wind spatial

patterns scored below the average. These GCMs have been upgraded from 20th and 19th over SEA to 11th and to 10th over Vietnam, respectively.

CONCLUSIONS

In this study, we examined and ranked the performances of 29 CMIP6 GCMs and their ensemble mean in representing near-surface temperature and precipitation over the seven climatic sub-regions constituting the mainland of Vietnam.

In general, the models reasonably reproduce the observational annual cycles of temperature and precipitation, although their performance varies across sub-regions. Yet, the GCMs tend to overestimate the annual temperature variability, explained by either an autumn–winter temperature underestimation and/or a warm bias in summer. The uncertainties between the GCMs are more pronounced for precipitation than for temperatures, both in the annual cycle and in spatial variability. In complex topographical regions, the models face a challenge in simulating observed patterns due to their coarse resolutions, resulting in warm biases over certain highland and mountainous areas and many of them are incapable of reproducing the rainy season shift from summer in most sub-regions towards the end of the year in Central Vietnam.

Following the methodology of DN22 though only for temperature and precipitation, a performance rank is assigned to each model. The top three models are EC-Earth3-Veg, EC-Earth3, and HadGEM3-GC31-MM. ENS_MEAN competes with the best GCMs and ranks 4th overall, explained by the fact that the averaging operator may absorb models' outliers, thus improving the results. The model resolution seems to affect the final ranking: in the same model family, the higher-resolution experiment exhibits a higher ranking than the lower-resolution one.

Furthermore, seven out of the top ten models in Vietnam (excluding ENS_MEAN) are also among the top ten GCMs in Southeast Asia. These models are EC-Earth3-Veg, EC-Earth3, HadGEM3-GC31-MM, CNRM-CM6-1-HR, CESM2, GISS-E2-1-G, and FIO-ESM-2-0. This result should strengthen the confidence in the models on this list for simulating the regional climate, although each model produces local specificities that should be taken into account. In future work, it would be beneficial to examine the performance of the top-ranking GCMs both before and after applying a bias-correction technique. Such a study would provide insights into the degree to which bias correction improves model performance, as well as whether it affects the ranking of the models.

The Ministry of Natural Resources and Environment of Vietnam plans to update the national report on climate change and sea level rise scenarios by 2025 based on a new set of downscaling experiments from CMIP6 GCMs (P.T.T. Nga from the Vietnam Institute of Meteorology, Hydrology and Climate Change, personal communication, 28 September 2022). Since dynamical downscaling often requires large computational resources, it is not possible to downscale all CMIP6 GCMs for a limited area like Vietnam. The results of this study can help to identify the most appropriate GCMs to be considered in the upcoming downscaled experiments in Vietnam.

AUTHORS' CONTRIBUTIONS

Th.N.-D. and Q.D. conceptualized the research. Th.N.-D. processed data and performed the analysis. All authors discussed the results and contributed to the writing of the manuscript.

FUNDING

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Abdelmoaty, H. M., Papalexiou, S. M., Rajulapati, C. R. & AghaKouchak, A. 2021 Biases beyond the mean in CMIP6 extreme precipitation: a global investigation. *Earth's Future* **9**, e2021EF002196. https://doi.org/10.1029/2021EF002196.

Anil, S., Manikanta, V. & Pallakury, A. R. 2021 Unravelling the influence of subjectivity on ranking of CMIP6 based climate models: a case study. *International Journal of Climatology* **41** (13), 5998–6016. https://doi.org/10.1002/joc.7164.

Arias, P. A., Bellouin, N., Coppola, E., Jones, R. G., Krinner, G., Marotzke, J., Naik, V., Palmer, M. D., Plattner, G. K., Rogelj, J., Rojas, M., Sillmann, J., Storelvmo, T., Thorne, P. W., Trewin, B., Achuta Rao, K., Adhikary, B., Allan, R. P., Armour, K., Bala, G., Barimalala, R., Berger, S., Canadell, J. G., Cassou, C., Cherchi, A., Collins, W., Collins, W. D., Connors, S. L., Corti, S., Cruz, F., Dentener, F. J., Dereczynski, C., di Luca, A., Diongue Niang, A., Doblas-Reyes, F. J., Dosio, A., Douville, H., Engelbrecht, F., Eyring, V., Fischer, E., Forster, P., Fox-Kemper, B., Fuglestvedt, J. S., Fyfe, J. C., Gillett, N. P., Goldfarb, L., Gorodetskaya, I., Gutierrez, J. M., Hamdi, R., Hawkins, E., Hewitt, H. T., Hope, P., Islam, A. S., Jones, C., Kaufman, D. S., Kopp, R. E., Kosaka, Y., Kossin, J., Krakovska, S., Lee, J. Y., Li, J., Mauritsen, T., Maycock, T. K., Meinshausen, M., Min, S. K., Monteiro, P. M. S., Ngo-Duc, T., Otto, F., Pinto, I., Pirani, A., Raghavan, K., Ranasinghe, R., Ruane, A. C., Ruiz, L., Sallée, J. B., Samset, B. H., Sathyendranath, S., Seneviratne, S. I., Sörensson, A. A., Szopa, S., Takayabu, I., Tréguier, A. M., van den Hurk, B., Vautard, R., von Schuckmann, K., Zaehle, S., Zhang, X., Zickfeld, K., 2021 Technical summary. In: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R. & Zhou, B., eds). Cambridge University Press, Cambridge, UK; New York, NY, USA, pp. 33 – 144. https://doi.org/10.1017/9781009157896.002.

Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Schamm, K., Schneider, U. & Ziese, M. 2013 A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present. *Earth System Science Data* **5**, 71–99. https://doi.org/10.5194/essd-5-71-2013.

Cesana, G. V. & Del Genio, A. D. 2021 Observational constraint on cloud feedbacks suggests moderate climate sensitivity. *Nature Climate Change* **11**, 213–218. https://doi.org/10.1038/s41558-020-00970-y.

Dasgupta, S., Laplante, B., Meisner, C., Wheeler, D. & Yan, J. 2007 *The Impact of Sea Level Rise on Developing Countries: A Comparative Analysis*. The World Bank, Washington, DC.

Desmet, Q. & Ngo-Duc, T. 2022 A novel method for ranking CMIP6 global climate models over the Southeast Asian region. *International Journal of Climatology* **42** (1), 97–117. https://doi.org/10.1002/joc.7234.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J. & Taylor, K. E. 2016 Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development* **9** (5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016.

Gebresellase, S. H., Wu, Z., Xu, H. & Muhammad, W. I. 2022 Evaluation and selection of CMIP6 climate models in Upper Awash Basin (UBA), Ethiopia. *Theoretical and Applied Climatology* **149**, 1521–1547. https://doi.org/10.1007/s00704-022-04056-x.

Hoang-Cong, H., Ngo-Duc, T., Nguyen-Thi, T., Trinh-Tuan, L., Jing Xiang, C., Tangang, F., Jerasorn, S. & Phan-Van, T. 2022 A high-resolution climate experiment over part of Vietnam and the Lower Mekong Basin: performance evaluation and projection for rainfall. *Vietnam Journal of Earth Sciences* **44** (1), 92–108. https://doi.org/10.15625/2615-9783/16942.

Hong, J., Javan, K., Shin, Y. & Park, J.-S. 2021 Future projections and uncertainty assessment of precipitation extremes in Iran from the CMIP6 ensemble. *Atmosphere* **12** (8), 1052. https://doi.org/10.3390/atmos12081052.

IPCC 2021 Climate change 2021: the physical science basis. In: *Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R. & Zhou, B., eds). Cambridge University Press, Cambridge, UK and New York, NY, USA, p. 2391.

Katzfey, J., Nguyen, K., McGregor, J., Hoffmann, P., Ramasamy, S., Nguyen, H. V., Khiem, M. V., Nguyen, T. V., Truong, K. B., Vu, T. V., Nguyen, H. T., Thuc, T., Phong, D. H., Nguyen, B. T., Phan-Van, T., Nguyen-Quang, T., Ngo-Duc, T. & Trinh-Tuan, L. 2016 High-resolution simulations for Vietnam – methodology and evaluation of current climate. *Asia-Pacific Journal of Atmospheric Sciences* **52** (2), 91–106. https://doi.org/10.1007/s13143-016-0011-2.

Li, J., Miao, C., Wei, W., Zhang, G., Hua, L., Chen, Y. & Wang, X. 2021 Evaluation of CMIP6 global climate models for simulating land surface energy and water fluxes during 1979–2014. *Journal of Advances in Modeling Earth Systems* **13**, e2021MS002515. https://doi.org/10.1029/2021MS002515.

Li, Z., Liu, T., Huang, Y., Peng, J. & Ling, Y. 2022 Evaluation of the CMIP6 precipitation simulations over global land. *Earth's Future* **10**, e2021EF002500. https://doi.org/10.1029/2021EF002500.

Ministry of Natural Resources and Environment 2009 *Climate Changes and Sea Level Rise Scenarios for Viet Nam*. MONRE, Hanoi, p. 34 (in Vietnamese).

Ministry of Natural Resources and Environment 2012 *Climate Change, Sea Level Rise Scenarios for Viet Nam*. Viet Nam Natural Resources, Environment and Mapping Publishing House, Hanoi, p. 112 (in Vietnamese).

Ministry of Natural Resources and Environment 2016 *Climate Change, Sea Level Rise Scenarios for Viet Nam*. Viet Nam Natural Resources, Environment and Mapping Publishing House, Hanoi, p. 188 (in Vietnamese).

Ministry of Natural Resources and Environment 2020 *Climate Change, Sea Level Rise Scenarios for Viet Nam*. Viet Nam Natural Resources, Environment and Mapping Publishing House, Hanoi, p. 286 (in Vietnamese).

Ngo-Duc, T., 2014 Climate change in the coastal regions of Vietnam. Book chapter. In: *Coastal Disasters and Climate Change in Vietnam – Engineering and Planning Perspectives* (Thao, N. D., Takagi, H. & Esteban, M., eds). Elsevier Inc., pp. 175–198. https://doi.org/10.1016/B978-0-12-800007-6.00008-3.

Ngo-Duc, T., Kieu, C., Thatcher, M., Nguyen-Le, D. & Phan-Van, T. 2014 Climate projections for Vietnam based on regional climate models. *Climate Research* **60** (3), 199–213. https://doi.org/10.3354/cr01234.

Nguyen, D. N. & Nguyen, T. H. 2004 *Vietnamese Climate and Climatic Resources.* Hanoi Agriculture Press, Hanoi, p. 230 (in Vietnamese).

Nguyen-Le, D., Matsumoto, J. & Ngo-Duc, T. 2015 Onset of the rainy seasons in the eastern Indochina Peninsula. *Journal of Climate* **28** (14), 5645–5666. https://doi.org/10.1175/JCLI-D-14-00373.1.

Nguyen-Thuy, H., Ngo-Duc, T., Trinh-Tuan, L., Tangang, F., Cruz, F., Phan-Van, T., Juneng, L., Narisma, G. & Santisirisomboon, J. 2021 Time of emergence of climate signals over Vietnam detected from the CORDEX-SEA experiments. *International Journal of Climatology* **41** (3), 1599–1618. https://doi.org/10.1002/joc.6897.

Papalexiou, S. M., Rajulapati, C. R., Clark, M. P. & Lehner, F. 2020 Robustness of CMIP6 historical global mean temperature simulations: trends, long-term persistence, autocorrelation, and distributional shape. *Earth's Future* **8** (10), e2020EF001667. https://doi.org/10.1029/2020EF001667.

Phan, V. T., Ngo-Duc, T. & Ho, T. M. H. 2009 Seasonal and interannual variations of surface climate elements over Vietnam. *Climate Research* **40** (1), 49–60. https://doi.org/10.3354/cr00824.

Rajendran, K., Surendran, S., Varghese, S. J. & Sathyanath, A. 2022 Simulation of Indian summer monsoon rainfall, interannual variability and teleconnections: evaluation of CMIP6 models. *Climate Dynamics* **58**, 2693–2723. https://doi.org/10.1007/s00382-021-06027-w.

Richter, I. & Tokinaga, H. 2020 An overview of the performance of CMIP6 models in the tropical Atlantic: mean state, variability, and remote impacts. *Climate Dynamics* **55**, 2579–2601. https://doi.org/10.1007/s00382-020-05409-w.

Seneviratne, S. I. & Hauser, M. 2020 Regional climate sensitivity of climate extremes in CMIP6 versus CMIP5 multimodel ensembles. *Earth's Future* **8** (9), e2019EF001474. https://doi.org/10.1029/2019EF001474.

Srivastava, A., Grotjahn, R. & Ullrich, P. A. 2020 Evaluation of historical CMIP6 model simulations of extreme precipitation over contiguous US regions. *Weather and Climate Extremes* **29**, 100268. https://doi.org/10.1016/j.wace.2020.100268.

Tangang, F., Chung, J. X., Juneng, L., Supari, Salimun, E., Ngai, S. T., Jamaluddin, A. F., Mohd, M. S. F., Cruz, F., Narisma, G., Santisirisomboon, J., Ngo-Duc, T., van Tan, P., Singhruck, P., Gunawan, D., Aldrian, E., Sopaheluwakan, A., Grigory, N., Remedio, A. R. C., Sein, D. V., Hein-Griggs, D., McGregor, J. L., Yang, H., Sasaki, H. & Kumar, P. 2020 Projected future changes in rainfall in Southeast Asia based on CORDEX–SEA multi-model simulations. *Climate Dynamics* **55** (5–6), 1247–1267. https://doi.org/10.1007/s00382-020-05322-2.

Taylor, K. E. 2001 Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research* **106** (D7), 7183–7192. https://doi.org/10.1029/2000JD900719.

Tong, M., Zheng, Z. & Fu, Q. 2022 Evaluation of East Asian Meiyu from CMIP6/AMIP simulations. *Climate Dynamics* **59**, 2429–2444. https://doi.org/10.1007/s00382-022-06218-z.

Tran-Anh, Q., Ngo-Duc, T., Espagne, E. & Tuan-Trinh, L. 2022 A high-resolution projected climate dataset for Vietnam: construction and preliminary application in assessing future change. *Journal of Water and Climate Change*, jwc2022144. https://doi.org/10.2166/wcc.2022.144.

Trinh-Tuan, L., Matsumoto, J., Tangang, F. T., Juneng, L., Cruz, F., Narisma, G., Santisirisomboon, J., Phan-Van, T., Gunawan, D., Aldrian, E. & Ngo-Duc, T. 2019 Application of quantile mapping bias correction for mid-future precipitation projections over Vietnam. *Scientific Online Letters on the Atmosphere* **15**, 1–6. https://doi.org/10.2151/SOLA.2019-001.

Tuyet, N. T., Ngo-Duc, T. & Phan-Van, T. 2019 Performance of SEACLID/CORDEX-SEA multi-model experiments in simulating temperature and rainfall in Vietnam. *Vietnam Journal of Earth Sciences* **41** (4), 374–387. https://doi.org/10.15625/0866-7187/41/4/14259.

Vu Dinh, Q., Doan, Q.-V., Ngo-Duc, T., Dinh, V. N. & Nguyen, D. D. 2022 Offshore wind resource in the context of global climate change over a tropical area. *Applied Energy* **308**, 118369. https://doi.org/10.1016/j.apenergy.2021.118369.

Wang, X. Y., Zhu, J., Chang, C. H., Johnson, N. C., Liu, H., Li, Y., Song, C., Xin, M., Zhou, Y. & Li, X. 2021 Underestimated responses of Walker circulation to ENSO-related SST anomaly in atmospheric and coupled models. *Geoscience Letters* **8**, 17. https://doi.org/10.1186/s40562-021-00186-8.

Xin, X., Wu, T., Zhang, J., Yao, J. & Fang, Y. 2020 Comparison of CMIP6 and CMIP5 simulations of precipitation in China and the East Asian summer monsoon. *International Journal of Climatology* **40** (15), 6423–6440. https://doi.org/10.1002/joc.6590.

Yasutomi, N., Hamada, A. & Yatagai, A. 2011 Development of a long-term daily gridded temperature dataset and its application to rain/snow discrimination of daily precipitation. *Global Environmental Research* **V15N2**, 165–172. https://www.chikyu.ac.jp/precip/data/Yasutomi2011GER.pdf.