The CMIP6-downscaled CORDEX-Southeast Asia (SEA) ensemble: evaluation and benchmarking for megacities of SEA

Phuong Loan Nguyen[1,2], Lisa V. Alexander[1,3], Thanh Ngo-Duc[4], Faye Cruz[5], Jerasorn Santisirisomboon[6], Liew Juneng[7], Donadi S. Permana[8], Jing Xiang Chung[9], Julie Mae Dado[5], John L. McGregor[10], Grace Redmond[11], Tse Wai Po[12], Fredolin Tangang[7,13], Tan Phan-Van[14], Son C. H. Truong[10], Marcus Thatcher[10], Long Trinh-Tuan[15], Ummu Ma'rufah[8], Jennifer Tibay[5], Giovanni Di Virgilio[1,2], Stephen White[2]

[1] Climate Change Research Centre, UNSW Sydney, New South Wales, Australia
[2] Climate & Atmospheric Science, NSW Department of Climate Change, Energy, the Environment and Water, New South Wales, Australia
[3] ARC Centre of Excellence for the Weather of the 21st Century, UNSW Sydney, New South Wales, Australia
[4] Department of Space and Earth Sciences, University of Science and Technology of Hanoi, Vietnam Academy of Science and Technology, Hanoi, Vietnam
[5] Manila Observatory, Ateneo de Manila University campus, Quezon City, Philippines
[6] Ramkhamhaeng University Center of Regional Climate Change and Renewable Energy (RU-CORE)
[7] Department of Earth and Environmental Sciences, Faculty of Science and Technology, University Kebangsaan Malaysia, Bangi, Selangor, Malaysia
[8] Indonesia Agency of Meteorology, Climatology, and Geophysics (BMKG), Jakarta, Indonesia
[9] Faculty of Science and Marine Environment, University Malaysia Terengganu, Kuala Nerus, Terengganu, Malaysia
[10] The Commonwealth Scientific and Industrial Research Organisation, Victoria, Australia
[11] Met Office Hadley Centre, Exeter, United Kingdom (UK)
[12] Hong Kong Observatory, Hong Kong
[13] Faculty of Arts and Social Sciences, University Brunei Darussalam, Brunei Darussalam
[14] Faculty of Meteorology, Hydrology, and Oceanography, University of Science, Vietnam National University
[15] Vietnam Academy for Water Resources (VAWR), Hanoi, Vietnam

Correspondence to: Phuong Loan Nguyen (loan.nguyen@dcceew.nsw.gov.au)

Abstract. A 21-member ensemble of regional climate simulations has been produced for Southeast Asia (SEA) by dynamically downscaling Coupled Model Intercomparison Project Phase 6 (CMIP6) Global Climate Models (GCMs) under the World Climate Research Programme's Coordinated Regional Climate Downscaling Experiment (CORDEX). The ensemble was generated by several modelling institutes using three regional climate models (RCMs) with eight distinct model configurations, resulting in a total of 62 simulations spanning the historical period and multiple future emissions scenarios. Model performance for mean, daily maximum/minimum temperature, and precipitation was evaluated against multiple observations at annual, seasonal, and daily time scales over SEA and its two subregions: Mainland and Maritime Continent (MC). Despite large observational uncertainties in precipitation intensity, the CMIP6 CORDEX-SEA ensemble captures the spatial and seasonal rainfall distribution reasonably well but tends to substantially overestimate observed rainfall. Wet biases, evident in about two-thirds of the models, are regionally and seasonally heterogeneous and larger over monsoon-dominated regions and seasons (e.g., MC during November–April and the Mainland during May–October). All RCMs showed widespread, statistically significant cold biases in daily mean temperature, which were largest during boreal winter, over the Mainland, and in simulations that have significant wet biases. These cold biases primarily arise from the models' underestimation of daily maximum temperature. The MC remains a challenging region since models struggle to accurately capture the spatial variability of rainfall and the internal variability of temperature. A standardised benchmarking framework was applied to precipitation and temperature, which ultimately identified 15 historical simulations that met our a priori model performance expectations. Analysing the range of future projections and model independence shows that simulations from the same RCM family exhibit similar bias structures, highlighting the importance of RCM setup and the selection of statistically independent models. From this process, eight simulations spanning three RCM configurations were selected for further kilometre-scale dynamical downscaling over megacities of SEA.

1 Introduction

Despite climate change being a global phenomenon, its impacts vary significantly across regions. Current global climate models (GCMs) operated at a coarse spatial resolution (~100 - 250 km) limit their ability for direct use in many regional climate impact applications and contexts. GCMs often struggle to simulate sub-grid weather (e.g., local land use, complex topography, etc.) and therefore cannot accurately simulate mesoscale and local-scale processes and feedbacks (e.g., deep convection, land–atmosphere interactions, etc.) (Douville et al., 2021; Maraun and Widmann, 2018). As a result, climate information at finer scales (e.g., 50-4 km), from sub-continental to local, is essential for assessing impacts and risks (Doblas-Reyes et al., 2021; IPCC, 2021). To produce this information, regional climate modelling groups have been downscaling GCMs targeting specific regions under the World Climate Research Programme's (WCRP) Coordinated Regional Climate Downscaling Experiment (CORDEX) initiatives (Giorgi et al., 2008; Giorgi and Gutowski, 2016).

Since its launch in 2009, CORDEX has facilitated the generation of regional climate simulations across 14 continental-scale domains, moving in parallel with the development of the Coupled Model Intercomparison Project (CMIP) phases (e.g., CORDEX-CMIP5). To provide a more consistent modelling framework across regions, the CORDEX Coordinated Output for Regional Evaluation (CORDEX-CORE) further delivered a standardised set of projections for all CORDEX domains using three global climate models (GCMs) from CMIP5 (Giorgi et al., 2021).

Within this global framework, Southeast Asia (SEA), a home to more than 700 million people, is one of the key domains (Fig. 1). The Southeast Asia Regional Climate Downscaling (SEACLID)/CORDEX-SEA project, launched in November 2013, represents the most comprehensive regional climate modelling initiative for this region, aiming to provide high-resolution climate information to support mitigation and adaptation planning. In its first phase, fourteen dynamical downscaling experiments with a grid spacing of 25 km were conducted over the SEA domain (89.5°E–146.5°E, 14.8°S–27°N, Fig. 1) through a global collaboration among seven modelling centres (Tangang et al., 2020). Subsequently, the WCRP's CORDEX-CORE initiative contributed three additional CMIP5-driven simulations for the SEA domain. Evaluation of CMIP5-driven CORDEX-SEA simulations indicated that while their multi-model ensemble (MME) generally improves representation of climatological precipitation compared to their driving GCMs, significant systematic wet biases remain, particularly within specific regional climate models (Tangang et al., 2020). However, CMIP5-forced regional climate models (RCMs) exhibit larger inter-model variability than their driving counterparts. Furthermore, Nguyen et al. (2022) demonstrated that these RCMs are not necessarily closer to the observed precipitation distribution than their driving GCMs. This highlights the ongoing challenge of regional climate modelling with a demand for a new generation of regional climate projections that provide more reliable and actionable data for climate risk assessments.

Building on these legacy developments, CMIP Phase 6 introduced a new generation of GCMs integrated with the Shared Socioeconomic Pathways (SSPs) (O'Neill et al., 2016), with their key results presented in the Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6) (IPCC, 2021). These advances enabled more comprehensive scenario-based climate projections. In response, the CORDEX community released an updated experimental design tailored for the dynamical downscaling of CMIP6 models (CORDEX, 2021), and modelling groups worldwide have since contributed simulations to provide well-grounded climate information for regional climate resilience and adaptation planning.

Following the CMIP6 CORDEX guidance, several modelling centres from SEA, Hong Kong, Australia, and the United Kingdom have recently initiated new downscaling experiments using CMIP6 GCMs for an expanded CORDEX-SEA domain (17.3°S–31.9°N, 86.5°E–149.3°E; Fig. 1), which now also fully spans Myanmar. This coordinated effort establishes a valuable foundation for generating updated and improved regional climate projections for impact and adaptation studies (Table 1). It is worthy to note that a fundamental requirement for the reliable use of dynamically downscaled climate projections is that a model can adequately simulate the observed climate, including the extreme events that threaten vulnerable natural and socioeconomic systems.

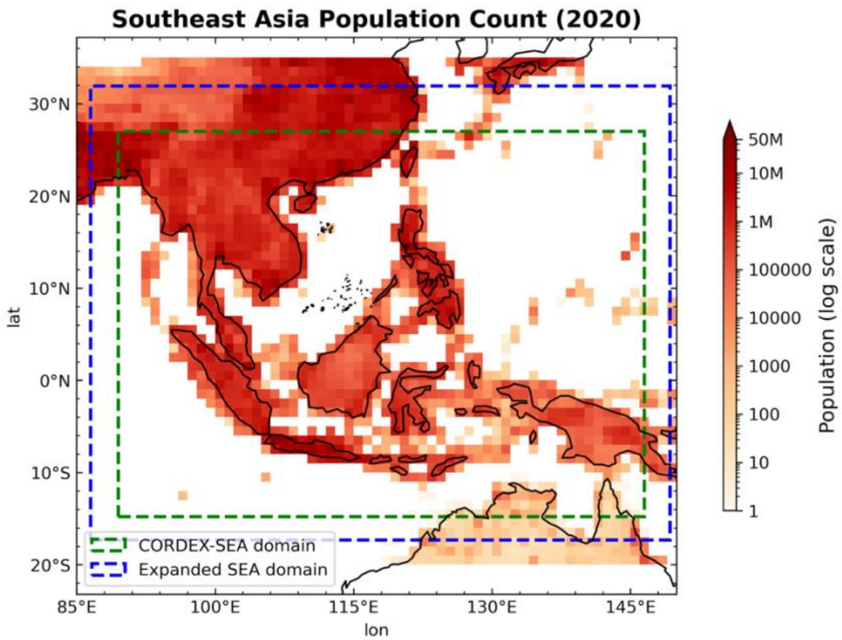

Figure 1. Southeast Asia's population in 2020, derived from the Gridded Population of the World (GPW), version 4, revision 11, provided by the Socioeconomic Data and Applications Center (SEDAC, 2018). The green dashed line indicates the CMIP5-downscaled CORDEX-SEA domain (89.5°E–146.5°E, 14.8°S–27°N), while the blue dashed line shows the expanded domain (86.5°E–149.3°E, 17.3°S–31.9°N) used for most CMIP6-downscaled simulations.

Cross-ensemble evaluations of regional climate models over SEA have been conducted using various approaches in the literature. The most common method, often applied by research groups targeting specific applications, is ranking models based on their ability to reproduce key climatological variables such as near-surface temperature and precipitation. This approach uses a set of statistical skill metrics, providing added value information of RCMs compared to their forcing GCMs (Tangang et al., 2020; Cruz et al., 2017; Juneng et al., 2016; Ngo-Duc et al., 2017; Nguyen et al., 2022), supporting the selection of suitable CMIP6 GCMs (Desmet and Ngo-Duc, 2022) or the optimization of Regional Climate Model version 4 (RegCM4; Giorgi et al., 2012) configurations (Ngo-Duc et al., 2024) for downscaling activities within the CORDEX-SEA domain. However, the results vary considerably across studies due to differences in the model ensembles and performance metrics employed. Consequently, a consistent observational benchmark and standardised evaluation framework are required to ensure comparability and robustness of assessments. To address this need, Nguyen et al. (2024) implemented a standardised benchmarking framework [BMF; Isphording et al. (2024)] to evaluate CMIP6 models against multiple observational datasets, aiming to identify a “fit-for-purpose” subset of GCMs for dynamical downscaling applications over Southeast Asia. Model selection within this framework follows a two-step process: first, models must meet minimum performance thresholds in reproducing key climatological characteristics of temperature and precipitation; and second, they are assessed for their ability to capture dominant precipitation drivers (e.g., the monsoon system) and teleconnections with major climate modes such as the Indian Ocean Dipole (IOD) and El Niño–Southern Oscillation (ENSO). The adoption of this standardised benchmarking approach represents a significant advancement, offering a transparent and objective foundation for model evaluation and strengthening confidence in future regional climate projections (Jiang et al., 2025). Furthermore, this framework could be directly applied to the CMIP6-downscaled CORDEX-SEA simulations for specific, target applications.

Climate extremes such as heatwaves, flooding, and drought have already caused substantial economic losses and damages, particularly in rapidly growing megacities such as Bangkok, Hanoi, Jakarta, Kuala Lumpur, and Manila (IPCC, 2022). Projections indicate that some climate extremes in SEA will become more frequent and intense with additional global warming (Tangang et al., 2020; Seneviratne et al., 2021). As urbanisation accelerates, it is critical that the development of these cities be guided by robust climate information to enhance resilience against future climate hazards. To address this need, the Climatic Hazard Assessment to Enhance Resilience against Climate Extremes for Southeast Asian Megacities (“CARE for SEA Megacities") project was established under the SEACLID/CORDEX-SEA framework with support from the Asia-Pacific Network for Global Change Research (APN, 2023). The CARE for SEA Megacities project aims to generate city-scale climate hazard information for these five SEA megacities under multiple SSPs by further dynamically downscaling RCM outputs to kilometre-scale resolution. Given the computational expense of such high-resolution experiments, it is not practical to downscale entire simulations but a subset of simulations. Consequently, implementing a standardised, fit-for-purpose benchmarking frameworks such as described above, is essential to select a subset of RCMs ensuring that downscaled urban projections are both robust and actionable for resilience planning.

To bridge those gaps and to support the robust application of dynamically downscaled datasets in climate impact studies, this study presents the first comprehensive cross-ensemble evaluation of CMIP6 CORDEX-SEA simulations. Specifically, we assess their capacity to reproduce historical climate characteristics, thereby benchmarking the ensemble for subsequent kilometre-scale dynamical downscaling. The objectives of this research are threefold:

(1) to document the experimental design for producing the updated CMIP6 dynamical downscaled climate projection over SEA;
(2) to present the first assessment of CMIP6 CORDEX-SEA simulations, which serves as a standard reference for CMIP6-downscaled CORDEX simulations over Southeast Asia, specially by providing a quantitative evaluation of how well each GCM–RCM combination reproduces different characteristics (e.g., climatological mean, spatial patterns, seasonal cycle and the whole distribution) of maximum, mean, and minimum temperature as well as precipitation for the entire SEA region and its two sub-regions: the Mainland and the Maritime Continent (MC);

(3) to benchmark precipitation and near-surface temperature from CMIP6 CORDEX-SEA ensembles using the BMF developed by Isphording et al. (2024) for subsetting models that can be considered for further kilometre-scale downscaling over the SEA megacities under the CARE for SEA Megacities project and to also consider other factors, including model dependence and spread in future response, given the fact that a balanced selection of models across the different model "families" spanning a reasonable range of future projections is desirable (Brunner et al., 2020; Di Virgilio et al., 2022; Nguyen et al., 2024; Grose et al., 2023; Sobolowski et al., 2025).

The remainder of this paper is structured as follows. Section 2 describes the RCM experimental design and analytical approach, considering observational uncertainties. Section 3 presents the model evaluation and results of the benchmarking framework to select a subset of RCMs for our purpose. Section 4 discusses the overall findings and concludes the paper.

2. Methods

2.1 Models and Experiment Design

To address the first objective of this study, this section documents the experiment design of the CMIP6 dynamically downscaled climate projection over SEA. Table 1 lists the CMIP6-downscaled simulations generated within the CORDEX-SEA framework. At the time of this analysis (September 2025), 62 simulations (denoted by cross) were available and are included in this study. As of July 2026, the CORDEX-SEA archive has expanded with additional simulations (e.g., denoted by ticks) completed and others still being produced (e.g., CMIP6 CORDEX-CORE) as part of this ongoing modelling effort.

These comprise three regional climate models (RCMs), eight model configurations, and eleven driving CMIP6 GCMs running across historical and future scenarios. A detailed description of the model configurations is provided in Table 2. The subset of GCMs was selected either based on their performance in reproducing key climatological features (Desmet and Ngo-Duc, 2022) or their ability to simulate dominant precipitation drivers (e.g., the monsoon) and teleconnections with major climate modes (e.g., IOD, ENSO), while considering model independence and the spread of future projections (Nguyen et al., 2024).

Most simulations cover the extended SEA domain (17.3°S–31.9°N, 86.5°E–149.3°E), which fully covers Myanmar, except for the CCAM-2021 simulations that span a slightly smaller region (14.8°S–27°N, 89.5°E–146.5°E). Simulation periods vary slightly across models but generally extend from 1960 to 2099, following the CMIP6 CORDEX Experiment Guidance (CORDEX, 2021), and include both SSP1-2.6 and SSP3-7.0 scenarios.

Eight RCM configurations are utilised in this study (Table 2). Non-hydrostatic RegCM [RegCM4-NH, Coppola et al. (2021)] simulations were conducted by several modelling institutions in SEA. Among the 44 physical parameterisation combinations conducted using RegCM4-NH in the sensitivity analysis phase with ERA5 forcing, two configurations—EXP16 and EXP28—

were selected based on their performance in reproducing near-surface temperature and precipitation (Ngo-Duc et al., 2024). Both use the University of Washington (UW) PBL scheme (Bretherton et al., 2004), but differ in their convective parameterisations: EXP16 employs the Tiedtke scheme (Tiedtke, 1989), while EXP28 uses the Kain–Fritsch scheme (Kain, 2004). Resolved-scale precipitation is represented using the SUBEX scheme (Pal et al., 2000). Other parameterisations follow the default RegCM4-NH setup, including the Community Climate Model version 3 (CCM3) radiative transfer scheme (Kiehl et al., 1998), the Community Land Surface Model version 4.5 [CLM4.5; Lawrence et al. (2019)] and the Zeng ocean flux scheme with roughness option 1 (Zeng et al., 1998). Recently, RegCM version 5 (RegCM5) has been released, featuring a new non-hydrostatic dynamical core derived from the MOLOCH weather prediction model (Giorgi et al., 2023). Accordingly, two additional experiments were performed using this configuration.

The Conformal Cubic Atmospheric Model (CCAM) is a global stretched-grid atmospheric model developed by the Commonwealth Scientific and Industrial Research Organisation (CSIRO), in which the target region is simulated at higher resolution (McGregor, 2005; McGregor and Dix, 2008; Nguyen et al., 2012; McGregor, 2015; Katzfey et al., 2016; Truong et al., 2025). CCAM employs a non-hydrostatic semi-Lagrangian dynamical core with multiple physical parameterisation options (Truong and Thatcher, 2025). Three CCAM configurations were applied by two institutions: the Climate Change Research Centre at the University of New South Wales (UNSW-CCRC) and CSIRO Australia.

The UNSW-CCRC CCAM configuration uses a C192 stretched grid, providing 25 km resolution over Southeast Asia. It employs 54 vertical atmospheric levels (from 20 m to 40 km) and 40 oceanic levels (to 5 km depth). Spectral nudging (Thatcher and McGregor, 2009; Truong et al., 2026) is applied to large-scale wind, temperature, and surface pressure fields from the host GCMs (e.g., ACCESS-CM2 and EC-Earth3-Veg) at a 6-hourly frequency and scales of ~3000 km, beginning around 850 hPa (~1.5 km above the surface). The UNSW-CCRC CCAM 2017 and 2021 versions differ in their convective and planetary boundary layer (PBL) schemes: the 2017 version uses a prognostic Turbulent Kinetic Energy (TKE)-based PBL scheme (Hurley, 2007), whereas the 2021 version applies a diagnostic Richardson-number-based scheme and an updated mass-flux convective closure including downdraft, entrainment, and detrainment processes (McGregor, 2003).

The CSIRO-CCAM configuration follows an AMIP-style setup driven by bias- and variance-corrected sea-surface temperature (SST) and sea-ice fields from CMIP6 GCMs (Hoffmann et al., 2016). It employs a C96 stretched grid and a modified version of the 2021 convective scheme (Mod2021), with 27 vertical levels extending from 20 m to 40 km.

Both UNSW-CCRC and CSIRO-CCAM share similar subsystems for the ocean and land surface [e.g., the Community Atmosphere Biosphere Land Exchange System—CABLE; Kowalczyk et al. (2006)] and a prognostic aerosol scheme implemented within the Simplified Edwards-Slingo four-band (SE4) (Freidenreich and Ramaswamy, 1999; Schwarzkopf and Ramaswamy, 1999) radiation scheme based on CSIRO-Mk3.6 aerosols.

The UK Met Office RCM HadREM3-GA7.05 (Tucker et al., 2022) is a limited-area configuration of the global GA7.05 model (Walters et al., 2019) with some modifications described in Buonomo et al. (2024). The RCM has a grid with 0.11° × 0.11° horizontal resolution (~12 km), a 4 minute timestep, and 63 levels (with an upper lid at ~39 km). It is driven at its lateral boundaries using a one-way nesting approach every 6 hours using variables from the driving GCMs (UKESM1-0-LL, EC-Earth3-Veg and NorESM2-MM). The RCM adjusts to the surface and lateral boundary forcing across a 13-point external rim. No nudging or parameter perturbations were used in either model configuration. In HadREM3-GA7.05, convection is parameterised using a mass-flux approach based on Gregory and Rowntree (1990). Large scale microphysics is a single-moment scheme derived from that of Wilson and Ballard (1999), and large-scale cloud is based on the prognostic cloud fraction and prognostic condensate PC2 scheme. The land surface model used is the Joint UK Land Environment Simulator (JULES) (Best et al., 2011; Clark et al., 2011). Aerosol radiation and cloud effects are derived from the MACv2-SP data set for the historical (Stevens et al., 2017) and SSP3-7.0 (O'Neill et al., 2016) scenarios. Monthly 3D fields of shortwave and longwave optical properties (absorption, extinction, scattering, and asymmetry) and cloud droplet number concentration on the 12 km model grid were calculated, and then these were prescribed as a time series in the RCM simulation for use in its calculation of time-varying radiative forcing.

Table 1. List of simulations carried out in CORDEX-SEA (as of July 2026)

No	RCM	Simulations					RCM (Institution contribution)
		Historical	SSP1-2.6	SSP2-4.5	SSP3-7.0	SSP5-8.5	
1	CCAM-2017-ACCESS-CM2	x	x		x		CCAM-2017 (UNSW-CCRC, Australia)
2	CCAM-2017-EC-Earth3-Veg	x	x		x		
3	CCAM-2021-ACCESS-CM2	x	x		x		CCAM-2021 (UNSW-CCRC, Australia)
4	CCAM-2021-EC-Earth3-Veg	x	x		x		
5	CCAM-Mod2021-ACCESS-CM2	x			✓		CCAM-Mod2021 (CSIRO, Australia)
6	CCAM-Mod2021-GFDL-CM4	x			✓		
7	HadGEM3-RA7-EC-Earth3-Veg	x			x		HadGEM3-RA7 (Met Office, UK)
8	HadGEM3-RA7-NorESM2-MM	x			x		
9	HadGEM3-RA7-UKESM1-0-LL	x			x		
10	RegCM4-CESM2-exp28	x	x		x		RegCM4 (Manila Observatory, Philippines, HKO, Hong Kong, China and USTH, Vietnam)
11	RegCM4-CNRM-ESM2-1-exp28	x	x	x	x	x	RegCM4 (USTH, Vietnam)
12	RegCM4-CNRM-ESM2-1-exp16	x	x	x	x	✓	
13	RegCM4-NorESM2-MM-exp28	x	x	x	x	x	
14	RegCM4-CanESM5-exp28	x			x		RegCM4 (UMT, Malaysia)
15	RegCM4-EC-Earth3-Veg-exp28	x	✓	✓	x	✓	RegCM4 (UKM, Malaysia; UMT, Malaysia and HKO, Hong Kong,

							China
16	RegCM4-MPI-ESM1-2-HR-exp28	x			x		RegCM4 (UKM, Malaysia and UMT, Malaysia)
17	RegCM4-MIROC-ES2L-exp28	x	x	x	x	x	RegCM4 (BMKG, Indonesia)
18	RegCM5-MIROC6-exp16	x		x		x	RegCM5 (RU-CORE, Thailand)
19	RegCM5-CMCC-ESM2-exp28	x		x		x	
20	RegCM5-EC-Earth3-Veg-exp28	x		x		x	
21	RegCM4-NorESM2-MM-exp16	x	x	x	x	x	RegCM4 (HKO, Hong Kong, China)
22	RegCM4-EC-Earth3-Veg-exp16	✓	✓	✓	✓	✓	

“x”: simulations available at the time of the analysis (September 2025) and included in this study.

“✓”: updated simulations available in July 2026 that are not included in this study

Table 2. Description of participating regional climate models

RCM	Planetary boundary layer physics	Cumulus physics	Radiation scheme	Land surface	Cloud microphysics schemes	Vertical Level	Nudging
CCAM-2017	Turbulent Kinetic Energy (TKE) (Hurley, 2007) closure scheme.	Updated version of mass-flux closure McGregor (McGregor, 2003) version 2017	Simplified Edwards-Slingo 4-band radiation scheme (more spectral bands for SW and LW) (SE4) (Freidenreich and Ramaswamy, 1999; Schwarzkopf and Ramaswamy, 1999)	CABLE land surface model (Kowalczyk et al., 2006)	a full mixed-phase, multi-category bulk microphysics scheme including liquid, ice, rain, snow, and graupel. (Rotstayn, 1997)	54	GCM nudging at 3000 km of u, v, t (Thatcher and McGregor, 2009)
CCAM-2021	Richardson-number based scheme (Ri)(McGregor, 1993)	Updated version of mass-flux closure (McGregor, 2003) version 2021					
CCAM-Mod2021	Richardson-number based scheme (Ri) (McGregor, 1993)	Updated version of mass-flux closure (McGregor, 2003), modified version 2021				27	SST and sea ice bias correction
RegCM4-NH-exp16	University of Washington (UW) (Bretherton et al., 2004)	Tiedtke (Tiedtke, 1989)	The community climate model version 3 (CCM3) radiative transfer scheme (Kiehl et al., 1998)	The community land surface model version 4.5 (CLM4.5) (Lawrence et al., 2019)	Sub-grid explicit moisture (SUBEX) (Pal et al., 2000)	18	NA
RegCM5-NH-exp16							NA
RegCM4-NH-exp28		Kain-Fritsch (KF) (Kain, 2004)					NA
RegCM5-NH-exp28							NA
HadGEM3-RA7	Wilson et al. (1999)	Mass-flux scheme (Gregory and Rowntree, 1990)	MACv2-SP (Stevens et al., 2017; O'Neill et al., 2016)	Joint UK Land Environment Simulator (JULES), (Best et al., 2011; Clark et al., 2011)	Prognostic cloud fraction and prognostic condensate PC2 scheme.	63	NA

2.2 Observations and data

Table 3. List of observational datasets of precipitation and temperature used in this study. Datasets marked with an asterisk (*) include daily maximum and minimum temperature data, while those marked with a double asterisk (**) include monthly data only.

Variables	Name	Short name	Coverage		Resolution		References
			Temporal	Spatial	Temporal	Spatial (°)	
Precipitation	APHRODITE_MA V1101 and V1101 XR	APHRODITE	1951–2015	60°E–15°E 15°S–55°N	Daily	0.25 x 0.25	Yatagai et al. (2012)
	REGEN_ALL_2019	REGEN_ALL	1950–2016	90°S–90°N	Daily	1 x 1	Contractor et al. (2020)
	CHIRPS_v2.0	CHIRPS	1981–2023	50°S–50°N	Daily	0.05 x 0.05	Funk et al. (2015)
	GPCC_FDD_2022	GPCC_FDD	1982–2020	90°S–90°N	Daily	1 x 1	Ziese (2022)
	SACA&D	SACAD	1981–2017	20°S–25°N 80°E–180°E	Daily	0.25 x 0.25	Van Den Besselaar et al. (2017)
Temperature	APHRO_MA V1808	APHRODITE	1961–2015	60°E–150°E 15°S–55°N	Daily	0.25 x 0.25	Yasutomi et al. (2011)
	ERA5*	ERA5	1940–2024	90°S–90°N	Daily	0.25 x 0.25	Hersbach et al. (2020)
	BEST*	BEST	1960–2019	90°S–90°N	Daily	1 x 1	Rohde and Hausfather (2020)
	CRU TS v4.08**	CRU TS	1901–2014	90°S–90°N	Monthly	0.5 x 0.5	Harris et al. (2020)
	SACA&D*	SACAD	1981–2017	20°S–25°N 80°E–180°E	Daily	0.25 x 0.25	Van Den Besselaar et al. (2017)

To address second and third objectives of this study on the evaluation and benchmarking of the performance of the CMIP6 CORDEX-SEA ensemble, we use multiple global and regional gridded observational datasets to account for observational uncertainties in model evaluation (Nguyen et al., 2020). It is worth noting that we focus on precipitation and temperature over land, as the temporal coverage of oceanic observational products is limited. The selected datasets are summarised in Table 3. Precipitation datasets include Asian Precipitation-Highly-Resolved Observational Data Integration Towards Evaluation of water resources (APHRODITE) (V1101 and V1101EX) at 0.25° resolution (Yatagai et al., 2012); Rainfall Estimates on a Gridded Network with 1° × 1° resolution [REGEN version Allstns V1 2019 (Contractor et al., 2020)] and Climate Hazards Group InfraRed Precipitation with Station at 0.25° × 0.25° resolution [CHIRPS version 2.0 (Funk et al., 2015)] They were chosen based on their demonstrated consistency in capturing daily precipitation and extremes over SEA (Nguyen et al., 2020; Nguyen et al., 2023). For daily near-surface temperature, we also use APHRODITE_MA V1808 at 0.25° resolution (Yasutomi et al., 2011); Berkeley Earth Surface Temperatures [BEST; Rohde and Hausfather (2020)]; the Climate Research Unit (CRU) TS v4.08 dataset at 0.5° resolution (Harris et al., 2020); and the ECMWF Reanalysis v5 (ERA5; Hersbach et al., 2020) at 0.25° resolution. Recognizing the sparse observational network over the Maritime Continent in both global datasets and the regional high-resolution dataset of APHRODITE (Yatagai et al., 2012), we additionally include the Southeast Asia Climate Assessment and Dataset (SACAD) version 2.0 at 0.25° resolution. Providing precipitation, mean, maximum, and minimum temperature at a daily scale, this dataset enhances the station network density and coverage, particularly over Indonesia, through contributions from the Indonesian meteorological institutes (Van Den Besselaar et al., 2017).

Hereafter, APHRODITE is selected as the primary baseline for all main precipitation figures, as it utilises the largest number of rain gauges among global datasets and compensates for the limited spatial coverage of SACAD. Note that for temperature, APHRODITE and CRUTS do not provide maximum, minimum, and mean temperature data at a daily time scale, while SACAD has limited spatial coverage (Fig. S1). Therefore, BEST is chosen as the reference dataset because it is an observational dataset that provides daily mean, maximum, and minimum temperature data across our Southeast Asia domain (89.26°E–127.28°E, 15.14°S–27.26°N). It should also be noted that inhomogeneities in the raw observations are not fully corrected in the BEST algorithm, which may introduce systematic temperature biases (e.g., the underestimation of temperature trend and magnitude), as previously reported in Canada (Way et al., 2017). Results from all other observational datasets are included in the Supplement (Figures S2–S11), with detailed explanations provided within the main text for intercomparison purposes.

All simulations are analysed over the climatological period 1982–2014, with overlapping coverage of multiple observations and simulations. For trend analysis, the period is extended to 1960–2014 to maximise the common period across all RCM simulations. Prior to assessment, all observational datasets and model outputs are regridded to the common 0.22° models' grid using conservative remapping for precipitation and bilinear interpolation for temperature for a fair comparison.

2.3. Methodology

2.3.1 Model evaluation

This section describes our approach to address the second objective of this study, which is to evaluate the ability of the CMIP6 CORDEX-SEA ensemble in simulating climatology of four model core variables: mean, maximum and minimum temperature and precipitation. Note that SEA exhibits clear seasonal and spatial contrasts in precipitation distribution. Higher rainfall is observed over the Mainland and less precipitation over MC during May-October (MJJASO). Meanwhile, the Mainland receives less precipitation while MC receives more precipitation during November-April (NJDFMA) (Juneng et al., 2016; Nguyen et al, 2022). Therefore, analyses of monthly totals of wet-day precipitation ($\geq$ 1mm, prcptot) are conducted at the seasonal scale, focusing on two distinct seasons: NDJFMA and MJJASO. Meanwhile, annual and two seasonal means of the boreal winter (December-January-February, DJF) and summer (June-July-August; JJA) are calculated using daily data for mean, maximum, and minimum temperatures. Analysis is conducted for the whole SEA domain (89.26°E–147.28°E, 15.14°S–27.26°N) and its two sub-regions: Mainland (5°N–27.5°N; 89.26°E–147.28°E) and MC (15.14°S–5°N; 89.26°E–147.28°E). The performance of each RCM in reproducing observed climate over these timescales is assessed using several metrics: (i) model bias, defined as the regionally averaged difference between simulated and observed values; (ii) mean absolute percentage error (MAPE); and (iii) spatial correlation with observations, which is adopted from the standardised benchmarking framework (Isphording et al., 2024). To assess the statistical significance of model biases, the Mann–Whitney U test ($\alpha = 0.05$) is applied for precipitation due to its non-normal distribution, while a student's t-test assuming equal variance ($\alpha = 0.05$) is applied for temperature.

The RCM's ability to simulate observed climate variables at daily time scales was also assessed by comparing each quantile in the daily mean observational distribution with those of the RCMs through quantile-quantile (Q-Q) plots. Then the distributions of RCMs and observations are compared using the area score metric (ASM) (Nguyen et al., 2022). In particular, the ASM calculates the absolute geometric area between the RCM simulation curve and the corresponding observational reference lines. By applying numerical integration via the trapezoidal rule, the metric quantifies cumulative deviations across the evaluated distribution (e.g., across percentiles). A lower ASM value indicates higher model fidelity, representing a tighter agreement between the downscaled output and historical observations.

2.3.2 Benchmarking

This section details the methodology to meet the third objective of this study, which is to identify a subset of CMIP6-downscaled RCMs that meet predefined performance criteria so as to be considered for further high-resolution dynamical downscaling at kilometre-scale over SEA megacities as part of the "CARE for SEA megacities" project. Scientists from the CORDEX–SEA community, in consultation with local policymakers and stakeholders from each city, have identified priority hazards, which include heat and rainfall extremes (https://cordex.org/wp-content/uploads/2024/01/CARE_for_SEA_megacities_workshop_summary.pdf). This prioritization underscores the need for models that can accurately represent the seasonal and spatial variability of temperature and precipitation across SEA. In addition to performance, the final subset of selected models should maintain diversity and minimise redundancy, ensuring that the ensemble captures a broad range of plausible future outcomes while remaining relevant and useful for decision-making.

Therefore, model selection is conducted in three stages. First, we benchmark all RCMs against multiple observational datasets for precipitation and near-surface temperature to assess their ability to reproduce key climatological characteristics using minimum standard metrics (MSMs) from the standardised benchmarking framework proposed by Isphording et al. (2024) with several modifications and additional considerations. In particular, MSMs include (i) mean absolute percentage error (MAPE) for precipitation and mean absolute error (MAE) for temperature; (ii) spatial correlation (Scor); (iii) seasonal cycle (Scyc); and (iv) significant temporal changes (Isphording et al., 2024). These metrics are designed to evaluate model skill in simulating the spatial and temporal variability, timing, and quantity of rainfall and temperature. In this research, we use multiple in situ and satellite-based observational datasets for evaluation. Figure S1 presents the spatial distribution of seasonal total precipitation and annual daily mean, maximum, and minimum temperature across multiple observational products. Given the substantial observational uncertainties associated with temperature and precipitation (Nguyen et al., 2020), assessment against a single reference dataset is not considered sufficient justification for model selection; instead, a model is included only if it meets our model performance expectations (e.g., predefined thresholds) for at least half (e.g., three out of five in this study) of the reference datasets.

In the second stage, we consider the projected CMIP6 CORDEX–SEA climate change signals for 2070–2099 relative to 1981–2010 during the MJJASO and NDJFMA seasons to ensure that the expected subset spans a range of plausible future outcomes. In the final stage, we examine inter-model dependencies using hierarchical clustering based on the historical climatology of temperature and precipitation, following Nguyen et al. (2024) and Gibson et al. (2024) This step ensures that at least two simulations are retained from each independent modelling group to minimise redundancy.

Note that due to the limited availability of reliable city-scale observational datasets, the benchmarking is conducted over the land of the entire SEA domain, rather than at city scale.

3. Results

3.1 Near-surface mean, maximum and minimum temperature

We evaluated the performance of the CMIP6-driven CORDEX-SEA simulations by comparing simulated and observed annual mean daily temperature (*tas)* from BEST over 1982–2014 (Fig. 2). Overall, the simulations reproduce the spatial patterns of observed temperature reasonably well, with Pearson's correlations (Scor) exceeding 0.8 in all cases. However, despite this consistency, all simulations systematically underestimate near-surface temperature. The area-averaged bias across terrestrial SEA ranges from −0.58°C to −3.04°C, which is substantially larger than the spread among reference datasets (−0.45°C to 0.45°C; Fig. S1). Seasonally, the cold bias is more pronounced during boreal winter (DJF) than boreal summer (JJA) (Fig. S2a). Most RCMs show statistically significant cold biases between approximately −5°C and −1.5°C across most grid points of the domain. The spatial distribution of these biases is strongly dependent on the RCM rather than the driving GCM. For example, RegCM simulations show cold biases over the mainland and western Indonesia but warm biases over parts of northwestern Mainland SEA (e.g., Myanmar), whereas CCAM and HadGEM3-RA7 simulations exhibit widespread cold biases across nearly the entire region. CCAM simulations, in particular, stand out with substantial cold biases exceeding 3°C, likely linked to excessive simulated precipitation (see Section 3.4). It is worth noting that some driving GCMs (e.g., EC-Earth3-Veg, NorESM2-MM, MPI-ESM1-2-HR) also exhibit cold biases, although these are less pronounced (Nguyen et al., 2024), whereas others (e.g., CMCC-ESM2, ACCESS-CM2, UKESM1-0-LL) exhibit warm biases. These results suggest that systematic errors in daily mean temperature are strongly influenced by the RCM rather than by the driving GCMs. In terms of mean absolute errors (MAE) to observation, CCAM simulations tend to show the biggest errors compared with other simulations, with the biggest difference over the Mainland and during the boreal winter (DJF) (Fig. S3).

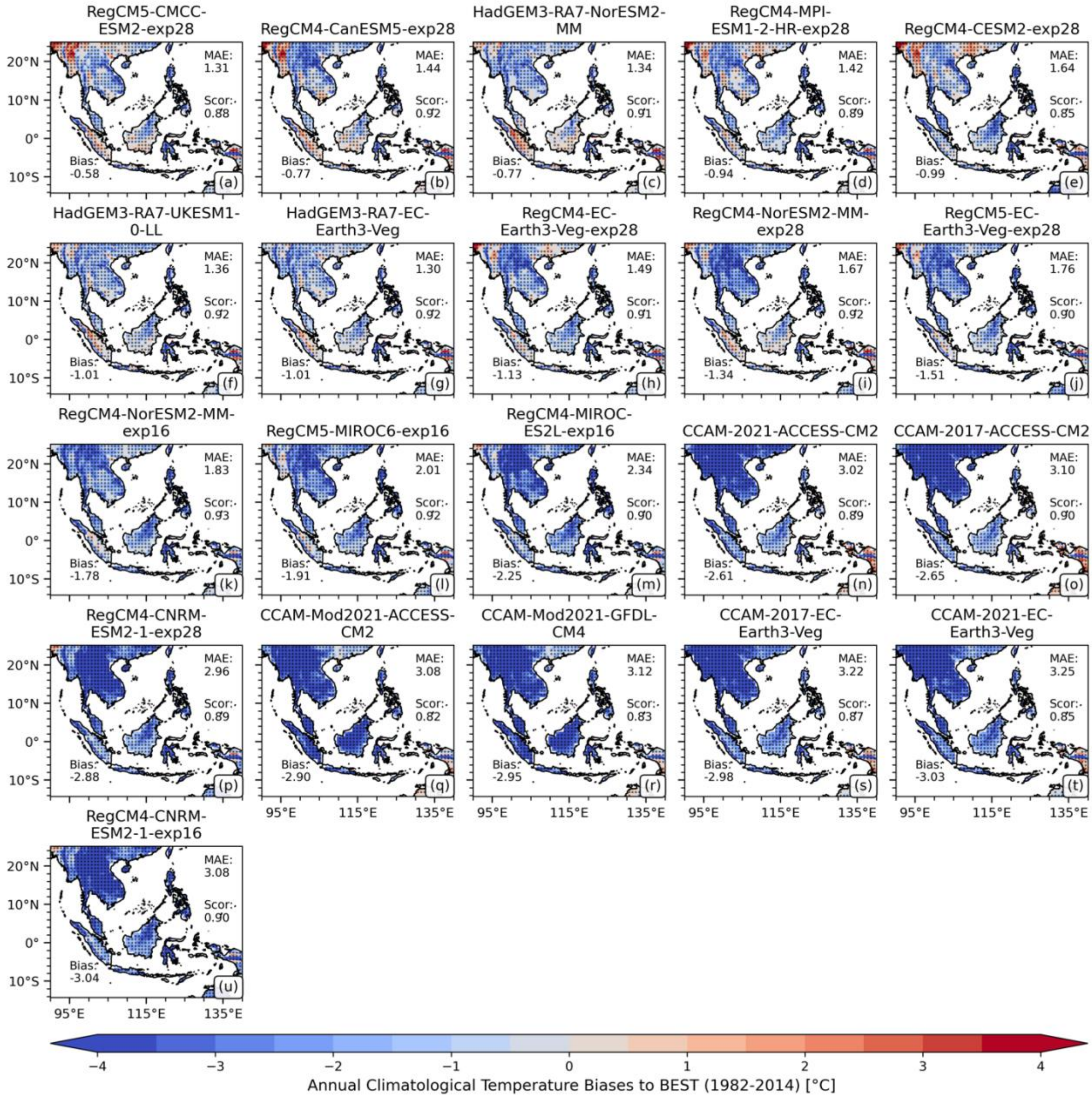

Figure 2. Annual climatological (1982–2014) mean near-surface temperature biases (in °C) for each model, ranked from warmest to coolest based on regionally averaged bias, against observational product BEST. The mean absolute errors (MAEs) and spatial correlations (Scor) calculated against BEST are shown in the upper-right corner. All analyses are considered at the resolution of the CMIP6-downscaled CORDEX-SEA simulation (i.e., ∼ 25 km). Stippled areas indicate locations where an RCM shows statistically significant bias ($p < 0.05$) according to the student two-sided t-test.

Figure S4 shows the seasonal cycle of near-surface temperature over two subregions: Mainland and MC, highlighting the systematic underestimation by CMIP6 CORDEX-SEA simulations. Overall, the RCMs capture the phase of the seasonal cycle well but tend to underestimate temperature magnitude throughout most months of the year, particularly during the boreal winter (DJF) over the Mainland. While most RCMs realistically simulate both the lower (DJF) and higher (MJJ) peaks over the Mainland (Fig. S4a), several simulations (e.g., RegCM5–MIROC6-exp16 and RegCM4–MIROC–ESM2L-exp28) fail to reproduce these features over MC (Fig. S4b). The mainland exhibits a clearly unimodal seasonal cycle, with a wide amplitude (e.g., 12–26°C) between lower and higher peaks, whereas the MC displays a bimodal cycle with a narrower range (e.g., 23–26.5°C).

To better understand the biases in mean temperature, we further examined model performance in daily maximum temperature (*tasmax*) and minimum temperature (*tasmin*) (Fig. 3 and Fig. 4, respectively). In general, the spatial coherence of RCMs (e.g., Scor) with observations in *tasmax* is weaker than that for *tas*, although Scor values exceed 0.7 across models. In addition, many RCMs substantially underestimate *tasmax* both annually (Fig. 3) and seasonally (Fig. S5), where regionally averaged *tasmax* biases vary widely across simulations (e.g., ranging from -4.92°C to −0.38°C at the annual scale; Fig. 3). These are the major contributions to the cold biases in *tas* found in Fig. 2. Notably, there are significant differences between the Mainland and MC. In particular, while all RCMs show a significant cold bias over MC, most RCMs show a mix of warm biases over the northwest and cold biases over the remainder of the Mainland. Interestingly, simulations from the same RCM setup show very similar spatial biases but differ in magnitude. For example, CCAM-Mod2021 simulations exhibit considerable warm biases north of 15° N, while RegCM simulations show strong warm biases exceeding 4°C over Myanmar and pronounced cold biases over the MC. Further seasonal and sub-regional analyses reveal that *tasmax* cold biases intensify (Fig. S5) and expand spatially during boreal summer (figures not shown), whereas during winter these cold biases weaken, often disappearing or even reversing to warm biases over parts of the Mainland. Consistent with these patterns, MAE values for *tasmax* peak during boreal summer in most simulations (Fig. S6).

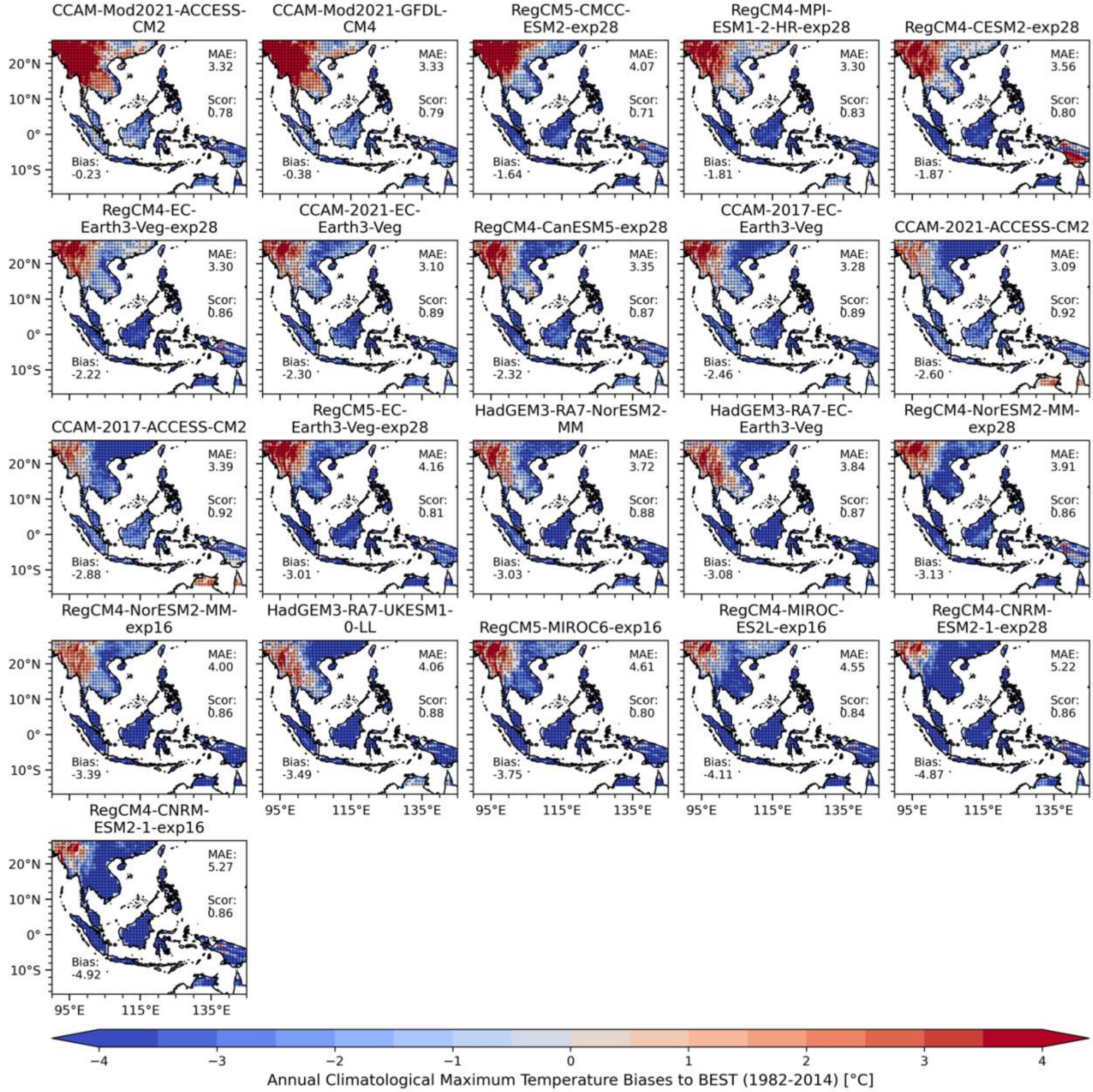

Figure 3. Annual climatological (1982–2014) daily maximum temperature biases (in °C) for each model, ranked from warmest to coolest based on regionally averaged bias, against observational product BEST. The mean absolute error (MAE) and spatial correlation (Scor) calculated against BEST are shown in the upper-right corner. All analyses are considered at the resolution of the CMIP6-downscaled CORDEX-SEA simulation (i.e., ~ 25 km). Stippled areas indicate locations where an RCM shows statistically significant bias ($p < 0.05$) according to the student t-test.

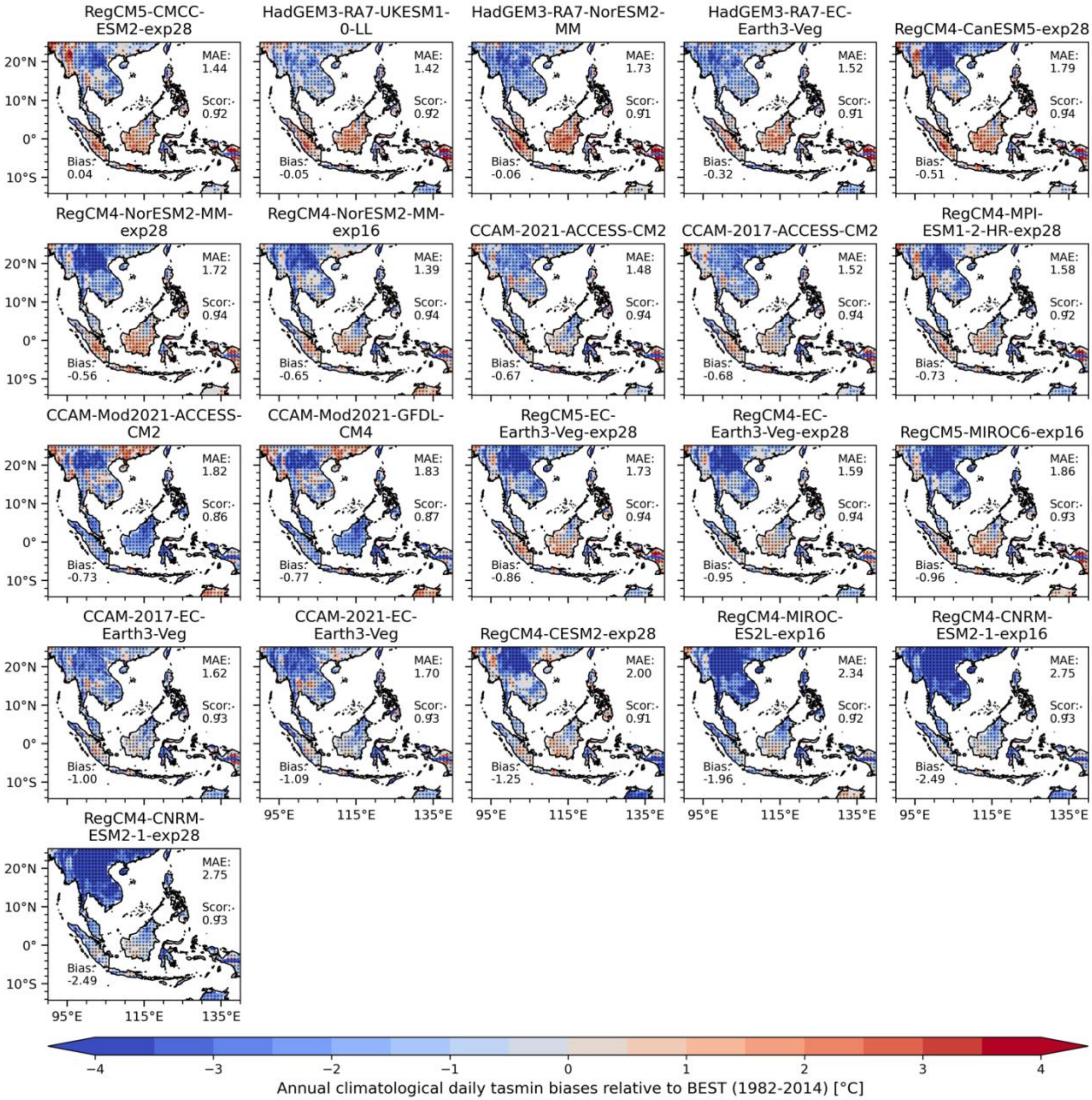

Figure 4. Annual climatological (1982–2014) daily minimum temperature biases (tasmin, in °C) for each model, ranked from warmest to coolest based on regionally averaged bias, against observational product BEST. The mean absolute error (MAE) and spatial correlation (Scor) calculated against BEST are shown in the upper-right corner. All analyses are considered at the resolution of the CMIP6-downscaled CORDEX-SEA simulation (i.e., ∼ 25 km). Stippled areas indicate locations where an RCM shows statistically significant bias ($p < 0.05$) according to the Student t-test.

In contrast, biases in daily minimum temperature (*tasmin*) are generally much smaller in magnitude but exhibit greater spatial heterogeneity across models and grid points than those in *tas* and *tasmax* (Fig. 4). CMIP6-downscaled simulations typically underestimate *tasmin* annually, with regionally averaged biases ranging from −0.04°C to −2.52°C. Spatially, cold biases in *tasmin* are more pronounced over the northern and eastern Mainland. Meanwhile, over the MC, the signal is more heterogeneous, with weaker cold biases or even warm biases in several simulations (e.g., HadGEM3-RA7, RegCM5-CMCC-ESM-exp28, RegCM4-CanESM5-exp28, and RegCM4-NorESM2-MM-exp28). Seasonally, *tasmin* cold biases are generally stronger during DJF (Fig. S7), with MAE values also peaking in the boreal winter (DJF) for most simulations (Fig. S8). In contrast, during boreal summer (JJA), nearly half of the simulations (10 out of 22) exhibit warm biases in *tasmin* (Fig. S7).

The ability of the CMIP6-driven CORDEX-SEA simulations to reproduce the whole observed temperature distribution at the daily time scale was further evaluated using quantile–quantile (Q–Q) analysis and ASM (Sect. 2.3.1) (Fig. 5). The higher the ASM, the further away it is from the observations. ASM was calculated for the entire SEA domain in boreal summer and winter (Fig. S9 and S10, respectively) and for *tas*, *tasmax* and *tasmin*. To provide a detailed assessment across the full distribution, temperature was evaluated across three intervals: the lower half (0–50th percentile), the central-to-upper range (50–99th percentile), and the extreme upper tail (>99th percentile).

Consistent with the mean bias results, the Q–Q plots for daily mean temperature (*tas,* Fig. 5a-c) show that most RCM simulations lie below the observational reference line, confirming a systematic cold bias across much of the distribution. These discrepancies are most pronounced at the lower end of the distribution (below the 50th percentile), indicating that models particularly underestimate cooler-than-average conditions. This deficiency is strongest over Mainland SEA and during boreal winter (Fig. S10d-f), consistent with the seasonal and regional bias patterns identified earlier. In contrast, model performance improves at higher percentiles, where simulated temperatures are closer to observations, and in some cases even exceed observed values at the extreme upper tail (>99th percentile). Model skill varies substantially among simulations. For example, RegCM5-CMCC-CESM2-exp28, HadGEM3-RA7-NorESM2-MM, and RegCM4-MPI-ESM1-2-HR-exp28 consistently rank among the best-performing simulations, with the lowest ASM values across regions and seasons. In contrast, RegCM4-CNRM-ESM2-1-exp16 and CCAM-Mod2021 simulations driven by ACCESS-CM2 and GFDL-CM4 exhibit the largest ASM values, reflecting pronounced cold biases and broader distributional errors (Fig. S3).

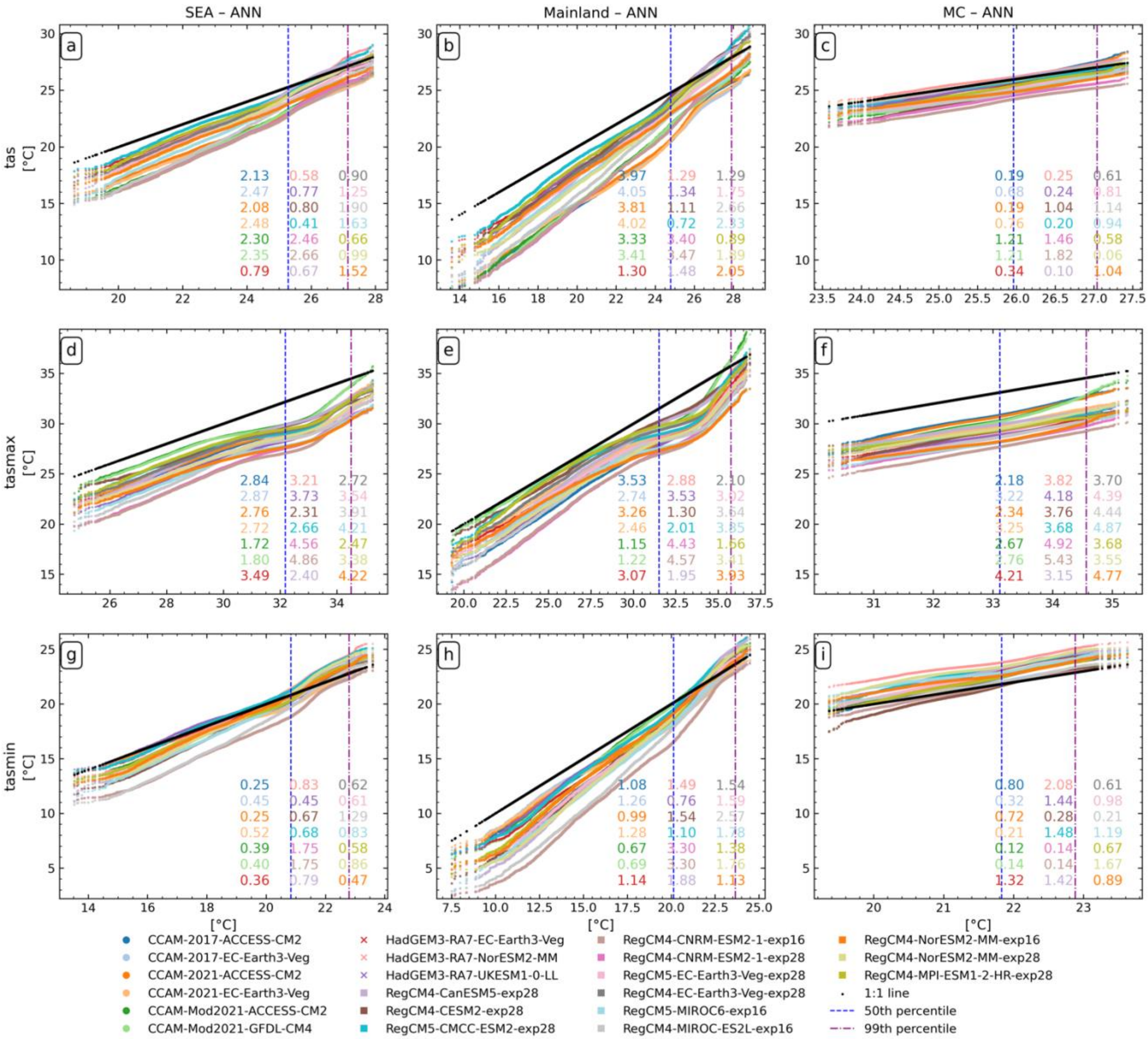

Figure 5. Quantile-quantile plots for regionally averaged daily near-surface temperature (tas, in °C; a-c), maximum temperature (tasmax; d-f), and minimum temperature (tasmin; g-i) across Southeast Asia and its two sub-regions: Mainland (5°N–27.5°N, 89°E–147°E) and the Maritime Continent (15°S–5°N, 89°E–147°E) during the climatological period of 1982–2014. The inserted number indicates the values of the area score metric (ASM; Nguyen et al., 2022), which measures the proximity between the two distributions: the model and the reference (BEST) (section 2.3.1). The vertical lines indicate the 50th percentile (blue) and 99th percentile (purple), respectively.

These distributional biases in *tas* are closely linked to deficiencies in simulating *tasmax*, which shows larger and more systematic departures from observations (Fig. 5d-f). Most RCM simulations fall well below the observational reference line across much of the distribution over SEA and its two subregions, confirming widespread cold biases. These biases are particularly pronounced in the mid-range of the distribution (50th–90th percentiles, corresponding to approximately 32–34°C), indicating that models struggle to reproduce typical daytime temperature conditions, particularly over the mainland (Fig. 5e). Interestingly, CCAM-Mod2021 simulations exhibit distinct behaviour in simulating the distribution of *tasmax*, with warm biases emerging at the extreme upper tail (>99th percentile), particularly over the Mainland. Among CMIP6-downscaled RCMs, CCAM-Mod2021-ACCESS-CM2 generally shows better agreement with observations than most other simulations. Meanwhile, RegCM4-CNRM-ESM2-1-exp16 consistently ranks among the five poorest-performing simulations based on ASM across both annual and seasonal timescales (Figs. S6, S9, and S10). More broadly, RegCM simulations tend to reproduce the *tasmax* distribution less accurately than other RCM families. These results highlight the earlier finding that cold biases in *tasmax* are a primary driver of errors in *tas* particularly due to the models' inability to accurately represent the central and upper portions of the temperature distribution.

In contrast, the RCMs generally demonstrate better skill in reproducing the distribution of *tasmin*, although systematic biases remain (Fig. 5g-i). Unlike *tas* and *tasmax*, *tasmin* biases exhibit a distinct pattern, with models tending to underestimate cooler-than-average conditions (below the 50th percentile) and overestimate warmer-than-average conditions (above the 50th percentile). This indicates an overall compression of the temperature distribution, with insufficient representation of the full observed variability. Despite this, several simulations (e.g., CCAM-Mod2021-ACCESS-CM2, CCAM-2017-ACCESS-CM2, CCAM-2017-EC-Earth3-Veg, and HadGEM3-RA7-EC-Earth3-Veg) show relatively good agreement with observations across the full distribution and across multiple reference datasets (Figs. S8-10). In contrast, some RegCM simulations, particularly RegCM4-CNRM-ESM2-1-exp28 and RegCM5-MIROC6-exp16, exhibit substantially poorer performance, with large ASM values reflecting systematic underestimation across much of the distribution, especially on annual and winter timescales.

Regional differences are also evident in *tasmin* performance. Most simulations show better agreement with observations over the MC. In contrast, larger errors over the Mainland suggest greater challenges in representing land-surface processes, boundary-layer dynamics, and regional circulation influences that strongly affect night time temperature.

Since there are large observational uncertainties in temperature variables over Southeast Asia (Fig. S1), the additional results for multiple reference products are presented in the supplementary material (Fig. S2-S10). It is noted that the temperature from BEST is higher than any other product. In general,, the main conclusion remains the same, regardless of the choice of reference. In summary, all simulations actually capture well the spatial distribution of temperature across SEA. However, cold biases of

CMIP6-downscaled RCMs in *tas* primarily originate from deficiencies in simulating *tasmax*, which exhibits larger magnitude errors and stronger regional and seasonal variability than those in *tas* and *tasmin*. While *tasmin* biases are generally smaller and more spatially heterogeneous, *tasmax* biases show more systematic and pronounced errors, particularly over the MC and during JJA. Importantly, across all temperature variables (*tas, tasmax, and tasmin)*, the spatial structure of biases is strongly determined by the RCM, highlighting the dominant role of regional model physics and parameterisations in shaping near-surface temperature performance in CORDEX-SEA simulations.

3.2 Precipitation

We then compared the seasonal mean total precipitation on wet days from RCM simulations (1982–2014) against APHRODITE (Figs. 6-7). Precipitation biases exhibit larger spatial and inter-RCM variability than temperature biases. Most models capture the spatial distribution of precipitation reasonably well, with 16 out of 21 simulations showing spatial correlations exceeding 0.4. In terms of regional averages, almost all RCMs exhibit wet biases for MJJASO total precipitation, ranging from 3.18 mm yr^{-1} to 158.95 mm yr^{-1}. The only exception is RegCM4-CESM2-exp28, which shows a dry bias of -16.84 mm yr^{-1} and a low spatial correlation score of 0.08.

Despite generally reasonable spatial correlations, precipitation bias patterns are highly spatially heterogeneous, characterised by a combination of wet and dry regions. Distinct groups of RCM-type-related biases can be identified. For instance, CCAM-2021 and CCAM-2017 simulations exhibit wet biases over topographically complex regions, particularly the MC, where numerous islands of varying sizes are present. Similarly, biases in HadGEM3-RA7 and CCAM-Mod2021 simulations remain consistent regardless of the driving GCM. For CCAM-Mod2021, this similar behaviour is likely due to the use of bias- and variance-corrected SST forcing, as previously noted by Evans et al. (2020). The underlying cause of the comparable behaviour in HadGEM3-RA7, however, remains unclear. In contrast, RegCM's simulations show pronounced wet biases over high wind-speed regions such as Thailand, along with dry biases over Borneo. Comparison of RegCM5 and RegCM4 simulations forced by EC-Earth3-Veg (Figs. 6d and 6g, respectively) shows an improvement in presentation of precipitation with substantial reductions of wet biases over Thailand. Interestingly, although RegCM-MIROC6 (Fig. 6m) and RegCM4-CanESM5-exp28 (Fig. 6n) exhibit similar domain-mean biases (approximately 83 mm yr^{-1}, Fig. 6), their spatial bias structures differ markedly. RegCM4-CanESM5-exp28 shows a predominantly wet bias across most of the domain, except over the north-western part of SEA, whereas RegCM5-MIROC6-exp16 exhibits dry biases over much of the domain, with wet biases confined primarily to north-western Myanmar and southern Thailand. Using the same GCM nudging technique with different convective parameterisations, CCAM-2017 simulations (Fig. 6c, k) show significant improvement compared with CCAM-2021 (Fig. 6p, u). These findings highlight the critical role of RCM setup, particularly the convective parameterisation in shaping precipitation biases over the SEA domain.

The rainfall pattern of the NDJFMA reveals some interesting features (Fig. 7). RCMs show better skill (e.g., higher spatial correlation with APHRODITE) in capturing the spatial distribution of precipitation than that during the MJJASO season, with most models exhibit dry biases over the mainland and wet biases over MC. Interestingly, the higher biases compared with that during the MJJASO season are more prevalent in MC in all simulations. Specifically, the RegCM simulations cannot capture the spatial distribution of precipitation well over MC, with all models showing a low correlation coefficient of less than 0.2 (figure not shown).

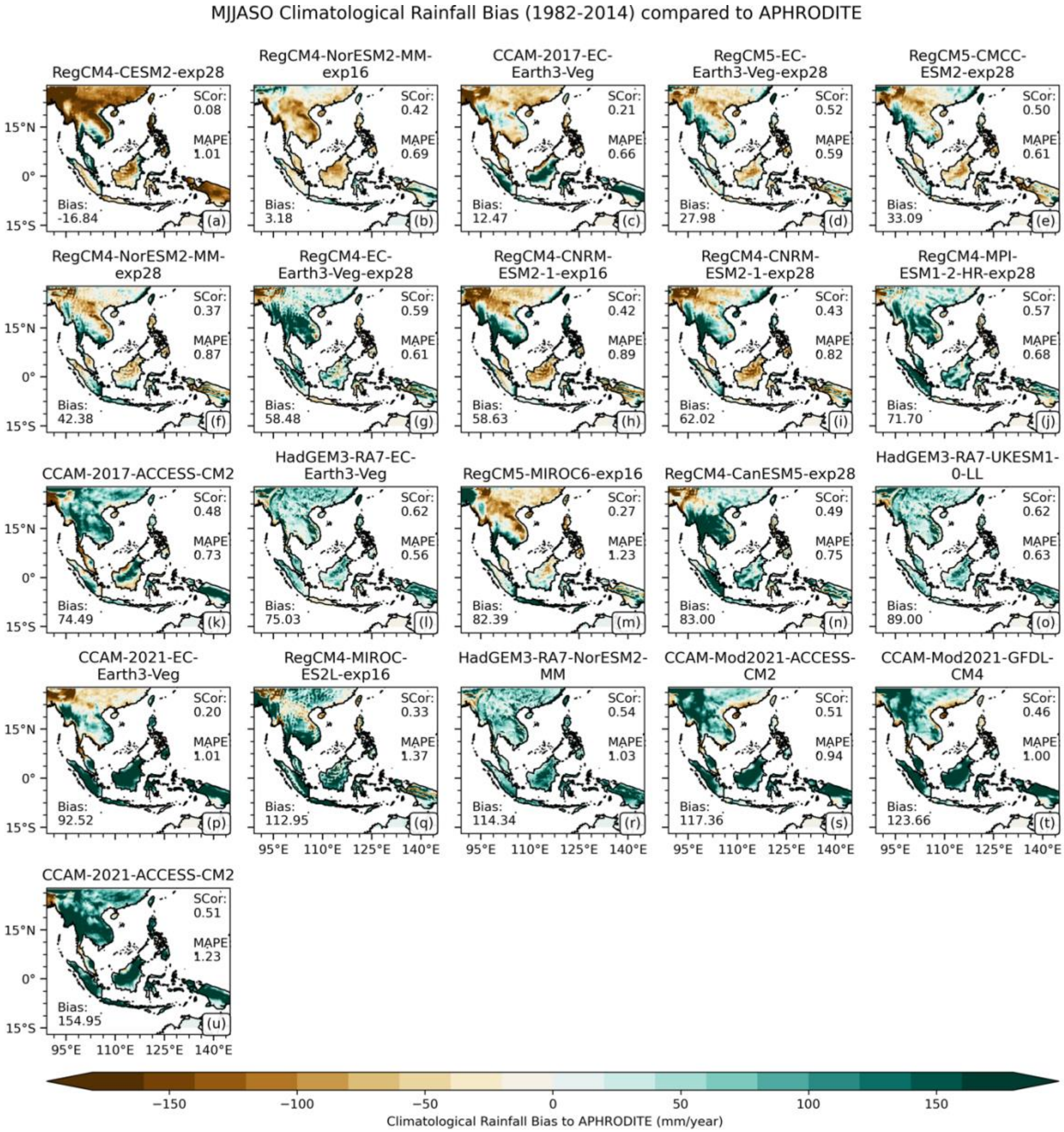

Figure 6. Seasonal climatological (1982–2014) biases (in mm yr^{-1}) for each model against the APHRODITE observational product during the May-October (MJJASO) season, ranked from driest to wettest based on regionally averaged bias. The mean absolute percentage errors (MAPEs) and spatial correlations (Scor) calculated against APHRODITE are shown in the upper-right corner. All analyses are considered at the resolution of the CORDEX-SEA simulation (i.e., ∼ 25 km). The stippled area indicates locations where RCM shows statistically significant bias ($p < 0.05$) according to the Mann-Whitney U test.

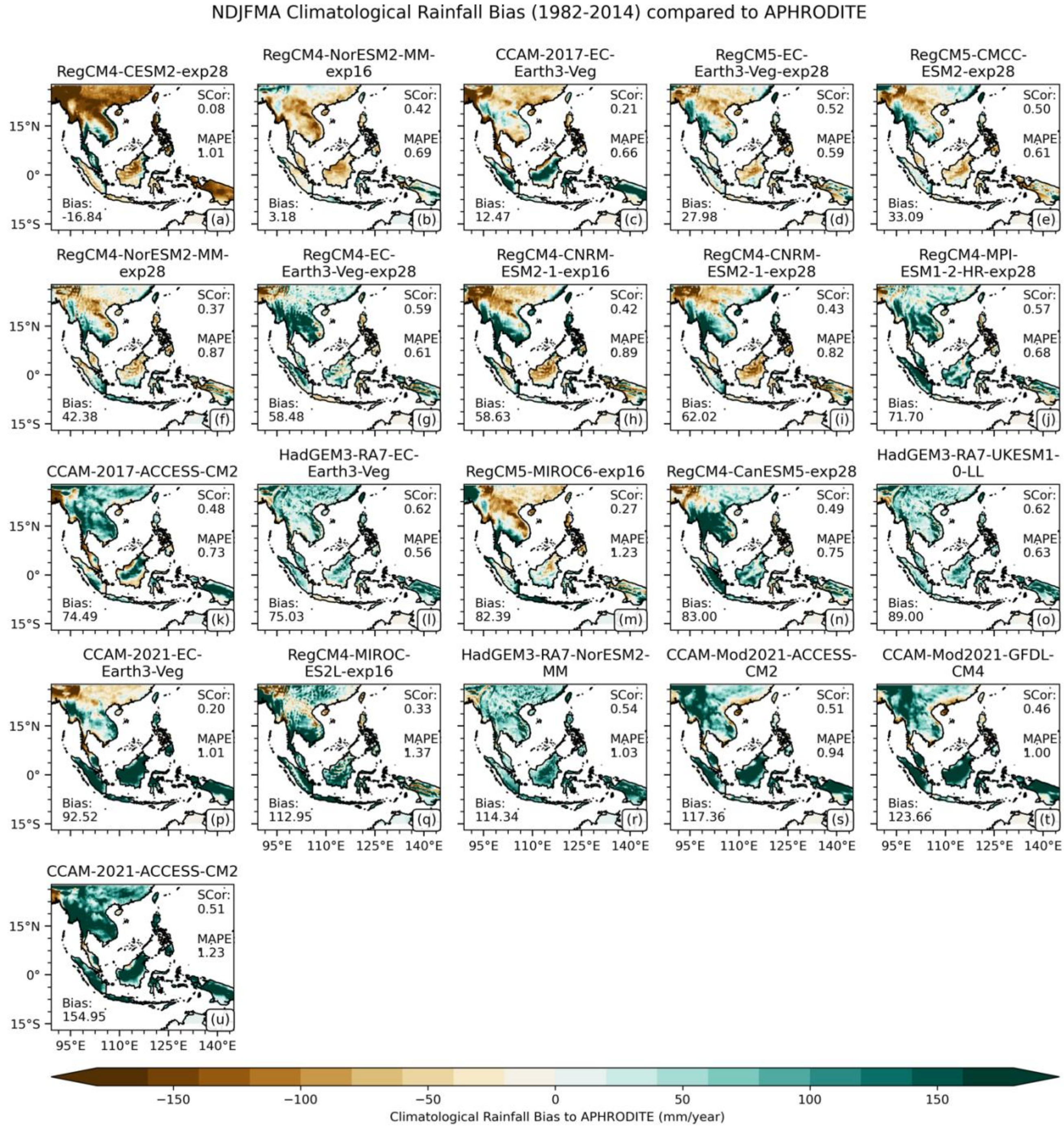

Figure 7. Same as Fig. 6 but for the November-April (NDJFMA) season.

Due to large observational uncertainties in precipitation during both seasons (Fig. S1), the similar analyses of mean bias are applied for other different observational datasets (Fig. S11). Interestingly, there are at least 14 out of 21 RCM simulations exhibiting the wet biases in terms of regionally averaged total wet-day precipitation, regardless of the choice of reference data.

The seasonal cycle of monthly mean total rainfall over two sub-regions of SEA is presented in Fig. S12, confirming the substantial overestimation of precipitation intensity in both regions as discussed previously. In term of timing, observed rainfall peaks during May–October over the Mainland and minimal rainfall are observed during the boreal winter (December-February). RCMs generally capture well the peak time of rainfall over the Mainland. However, most RegCM simulations fail to adequately capture the precipitation peak over the MC. It is important to note that different sub-regions of the Maritime Continent (MC) exhibit multiple distinct climate regimes, including monsoonal regions (e.g., Java) with a dominant rainfall peak during the austral summer (December–February), semi-monsoonal regions (e.g., parts of Borneo and Sumatra) with equatorial bimodal peaks during the transitional seasons (March–May and September–November), and anti-monsoonal regions (e.g., eastern part of the Maritime Continent) characterised by an opposite seasonal phase that peaks during the boreal summer (June–August) due to its local topography, specific mountain ranges, and ocean-atmosphere interactions (Tangang et al., 2020; Aldrian et al., 2004) These regimes are further modulated by synoptic processes such as cold surges and the Borneo vortex over Peninsular Malaysia and surrounding areas, leading to substantial spatial variability in rainfall seasonality. As a result, rainfall peaks occur at different times across sub-regions, and no single unimodal seasonal-cycle representation is appropriate for the entire Maritime Continent. This deficiency highlights the models' inability to correctly simulate the complex circulations and precipitation dynamics over MC, thereby warranting further sub-regional analyses of atmospheric circulation patterns across localised zone within MC.

Figure 8 compares regionally-averaged daily precipitation quantiles from CORDEX-SEA RCMs with APHRODITE observations over the whole of SEA, the Mainland, and the MC for the MJJASO and NDJFMA seasons during 1982–2014. Precipitation is also considered in three percentile ranges: 0–50th percentile (left of the blue dashed line), 50–99th percentile (between the dashed and dot-dashed lines), and 99th percentile (right of the dot-dashed line), allowing a detailed assessment of model performance across the distribution.

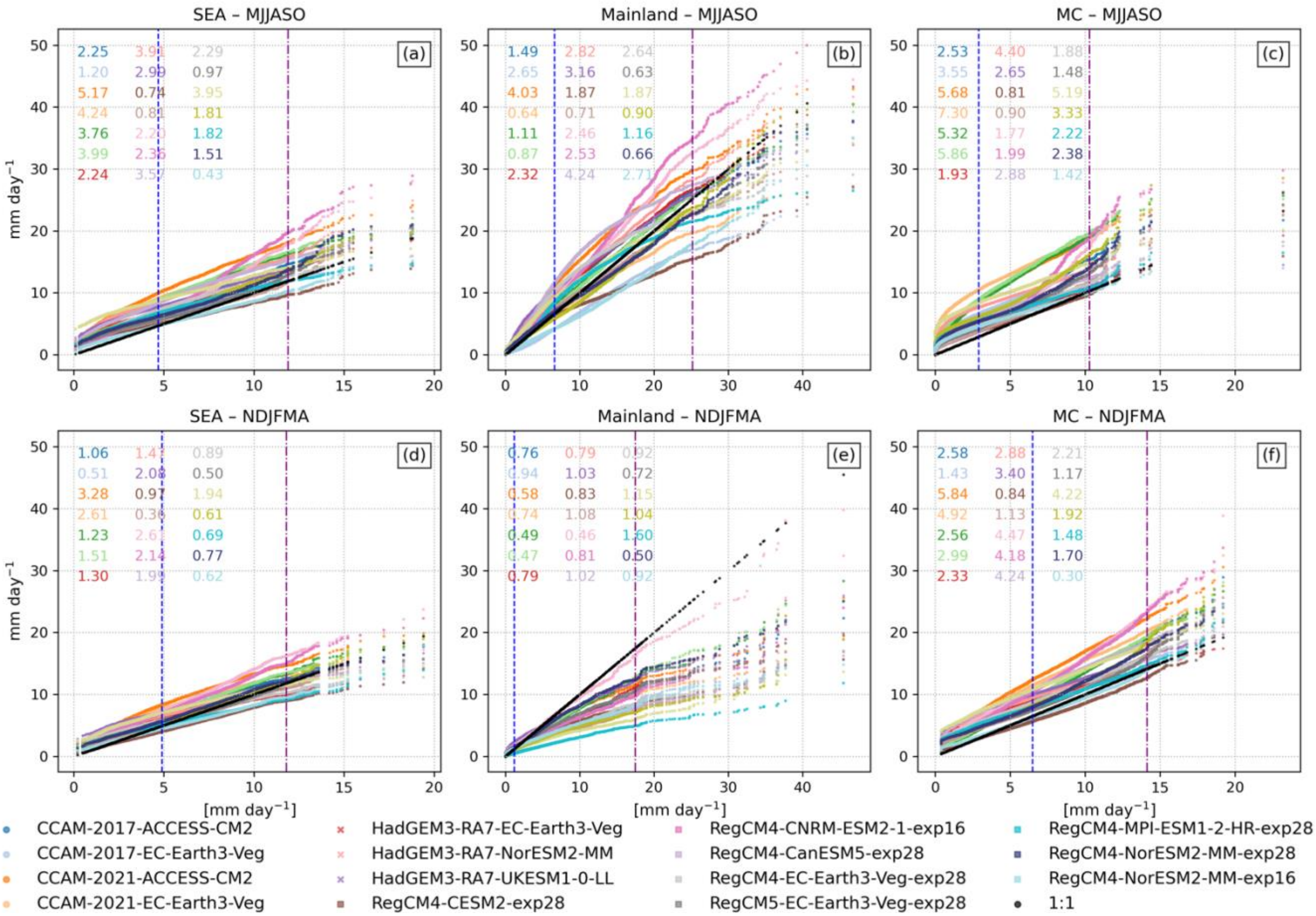

Figure 8. Quantile-quantile plots for regionally averaged daily precipitation (pr, in mm d^{-1}) across Southeast Asia for the MJJASO and NDJFMA seasons during the climatological period of 1982-2014. The inserted number indicated values of the area score metric (ASM), which measures the proximity between the two distributions: model and observation (APHRODITE). The vertical lines indicate the 50th percentile (blue) and 99th percentile (purple), respectively.

Overall, almost all RCM (coloured) lines are above the observational line (black), indicating more intense precipitation in simulations except over the Mainland during the NDJFMA season. This systematic shift toward higher precipitation values in RCMs compared to APHRODITE is more evident after the 50th percentile. At the highest percentiles, RCMs diverge markedly from the reference dataset, highlighting the ongoing challenges for RCMs to capture the intensity of extreme precipitation. Note that some RCMs (CCAM-2017-ACCESS-CM2, CCAM-2021-EC-Earth3-Veg) more closely match observed extremes during dry seasons, while others remain biased, either overestimating (e.g., RegCM4-EC-Earth3-Veg -exp28 and RegCM4-MIROC-ES2L-exp28) or underestimating (RegCM4-CESM2-exp28) the heaviest rainfall during the dry seasons. An exception is RegCM-CESM2, which has persistent dry biases in the RCM in both seasons. Interestingly, all simulations are below the observational line for the Mainland during DNJFMA, indicating the dry biases as mentioned above (Fig. 8e). Quantitative evaluation using the area score shows that CORDEX-SEA simulation aligns more closely with observations during the dry season compared to the wet season (17 out of 21 simulations exhibit smaller area scores, regardless of the reference dataset, Fig. S13). Among all simulations, CCAM-2021-ACCESS-CM2 displays the largest deviations from observations across both seasons, as reflected in its area scores, which are higher than those of other RCMs.

There is an interesting difference between the two sub-regions. While the RCMs generally capture precipitation below the 50th percentile well over the Mainland, the distribution over the MC is more divergent, even at percentiles below the 50th. This highlights the challenges in simulating precipitation over the MC, which consists of numerous islands of varying sizes and complex atmospheric-oceanic interactions. Moreover, due to the dominant role of convective rainfall over the MC, this region is highly sensitive to the cumulus parameterisations schemes used in RCMs (Ngo-Duc et al., 2017)

3.3 Benchmarking CMIP6-CORDEX SEA simulations for further downscaling at kilometre-scale

In this section, we move to the next step of sub-setting the 21 RCM simulations from the CMIP6 CORDEX-SEA ensemble for further dynamical downscaling at the kilometre-scale over SEA megacities, under the CARE for SEA megacities project using the minimum standard metrics (MSMs) from the benchmarking framework (BMF) (Isphording et al., 2024) as mentioned in Sect. 2.2.

3.3.1 Spatial benchmark

This benchmarking evaluates the ability of the RCM ensemble to reproduce the magnitude and spatial distribution of observed precipitation and temperature by comparing model outputs against five reference datasets. The expectation is that models should not exhibit substantial wet or dry biases in precipitation nor strong cold or warm biases in temperature and should adequately capture the spatial variability of these core climate variables. Figure 9 presents the mean absolute percentage error (MAPE) for precipitation during the MJJASO and NDJFMA seasons and the mean absolute error (MAE) for annual temperature, while Fig. 10 shows the corresponding spatial correlations with multiple reference datasets. Bias metrics are calculated using absolute percentage errors for precipitation and absolute temperature differences for temperature, ensuring that positive and negative deviations across grid points do not cancel out. Given the considerable observational uncertainties in rainfall and temperature (Fig. S1), caution is required when interpreting these bias benchmarks.

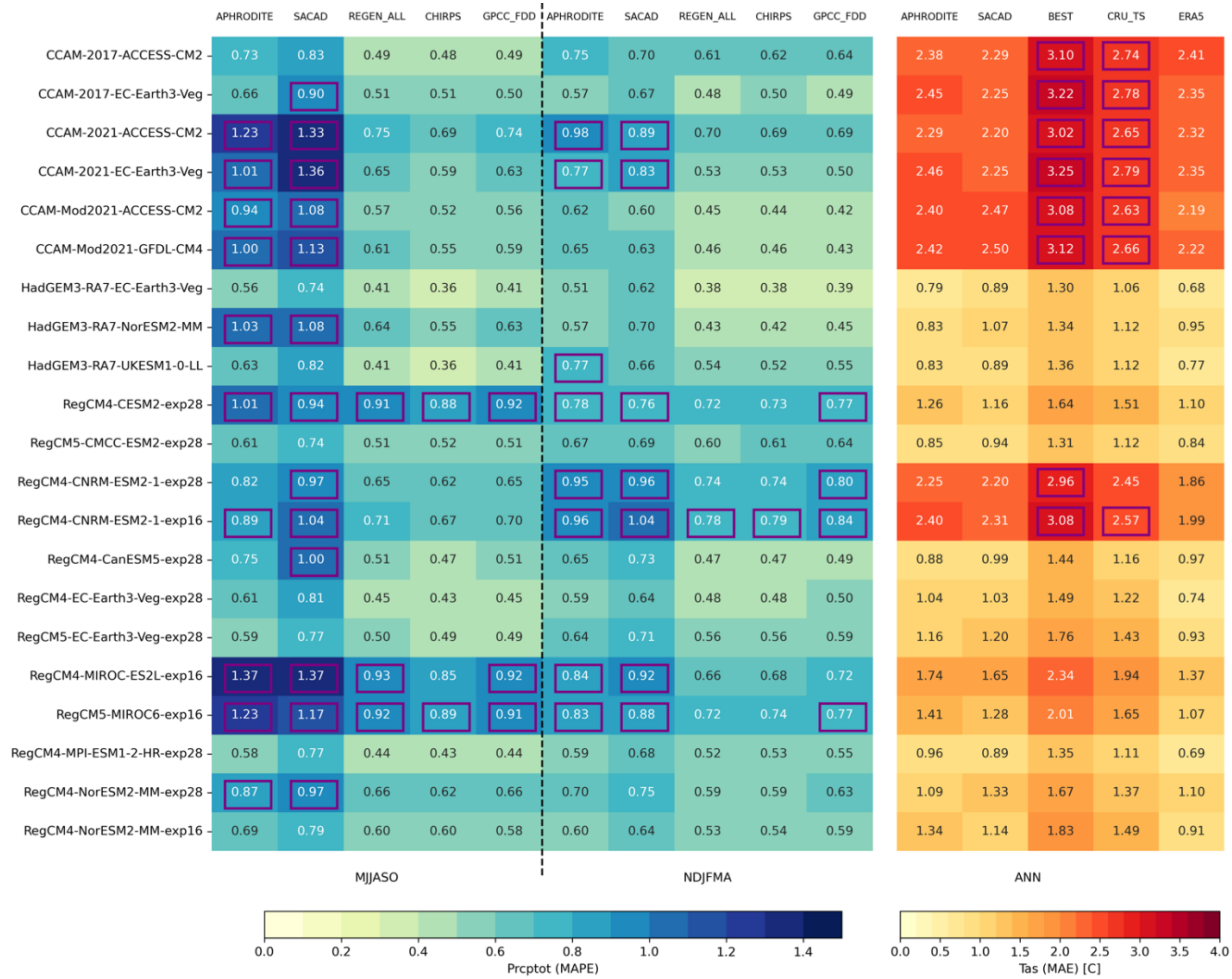

Figure 9. Score benchmarks for biases in seasonal total precipitation (MAPE: Mean Absolute Percentage Errors) and annual average of daily mean near-surface temperature (MAE: Mean Absolute Errors) referenced to multiple observational datasets. Biases larger than the benchmark are marked by purple squares. Lighter colours typically indicate better performance, whereas darker colours often indicate worse performance.

At this MSM stage, we would like to retain as many models as possible to maintain a sufficiently large ensemble for further steps (e.g., future response spread, model dependence) in the RCM selection process. Therefore, relaxed thresholds were purposefully applied to traditionally challenging metrics. To account for seasonal variations in model performance, different fixed thresholds were prescribed for wet and dry seasons. In particular, thresholds of 0.85 (MAPE ≤ 0.85) and 0.75 (MAPE ≤ 0.75) were applied to the precipitation benchmark for the MJJASO and NDJFMA seasons, respectively while a threshold of 2.5 (MAE ≤ 2.5) was used for the temperature benchmark. These specific thresholds were informed by a combination of observational uncertainty and expert scientific judgment. Rather than applying statistical constraints, models were given the benefit of the doubt to maximise model retention and preserve a wide range of plausible climate futures. These are similar to the selection strategies of Nguyen et al. (2024), who balanced model performance and ensemble representativeness in selecting CMIP6 model performance for dynamical downscaling purposes over SEA. It is worth noting that more relaxed thresholds were applied here due to the smaller number of RCMs considered compared with the number of GCMs. Similarly, for the spatial correlation metric (Scor), thresholds of 0.3 and 0.5 were applied for precipitation during the MJJASO and NDJFMA seasons, respectively, while 0.8 is applied for annual mean temperature.

Figure 9a illustrates the sensitivity of the intensity benchmark (MAPE) to different precipitation reference datasets. For instance, when regional datasets such as APHRODITE and SACAD are used as references, at least ten models exceed the benchmark thresholds. These include the CCAM-2021 and CCAM-Mod2021 simulations, HadGEM3-RA7-NorESM2-MM, RegCM4-CESM2-exp28, RegCM4-CNRM-ESM2-1-exp16, RegCM5-MIROC-ES2L-exp16, RegCM4-MIROC6-exp28, and RegCM-NorESM2-MM. It is worth noting that although regional datasets incorporate a greater number of gauge stations than global products, they tend to be drier in both mean and extreme precipitation (Alexander et al., 2025; Nguyen et al., 2020; Yatagai et al., 2012) compared with global precipitation dataset due to topographic wind-undercatch (particularly across complex mountainous terrains) and station sparsity and the smoothing effects of spatial interpolation (Yatagai et al., 2012 and Nguyen et al., 2020). A subset of models consistently fails the intensity benchmark (e.g., exceeding thresholds for at least three references datasets) regardless of the reference used. These include RegCM4-CESM2-exp28, RegCM5-MIROC-ES2L-exp16, and RegCM4-MIROC6-exp28 in both seasons, as well as RegCM4-CNRM-ESM2-1-exp16 and RegCM4-CNRM-ESM2-1-exp28 during the dry season. In contrast, temperature exhibits less sensitivity across reference datasets. Several models, including the six CCAM simulations, RegCM4-CNRM-ESM2-1-exp28, and RegCM4-CNRM-ESM2-1-exp16, still show MAE exceeding 2°C irrespective of the reference dataset. This might result from the consistent cold biases across all grids in these models.

Figure 10 shows that nearly all models pass the spatial variability benchmark for temperature, with spatial correlations exceeding 0.8. Note that CCAM's simulations have a lower spatial correlation with SACAD (e.g., less than 0.8) than other references. Only two models, RegCM-CESM2 and RegCM-MIROC6, fail to meet our expectation in capturing the spatial distribution of precipitation in either or both seasons, regardless of the reference dataset used.

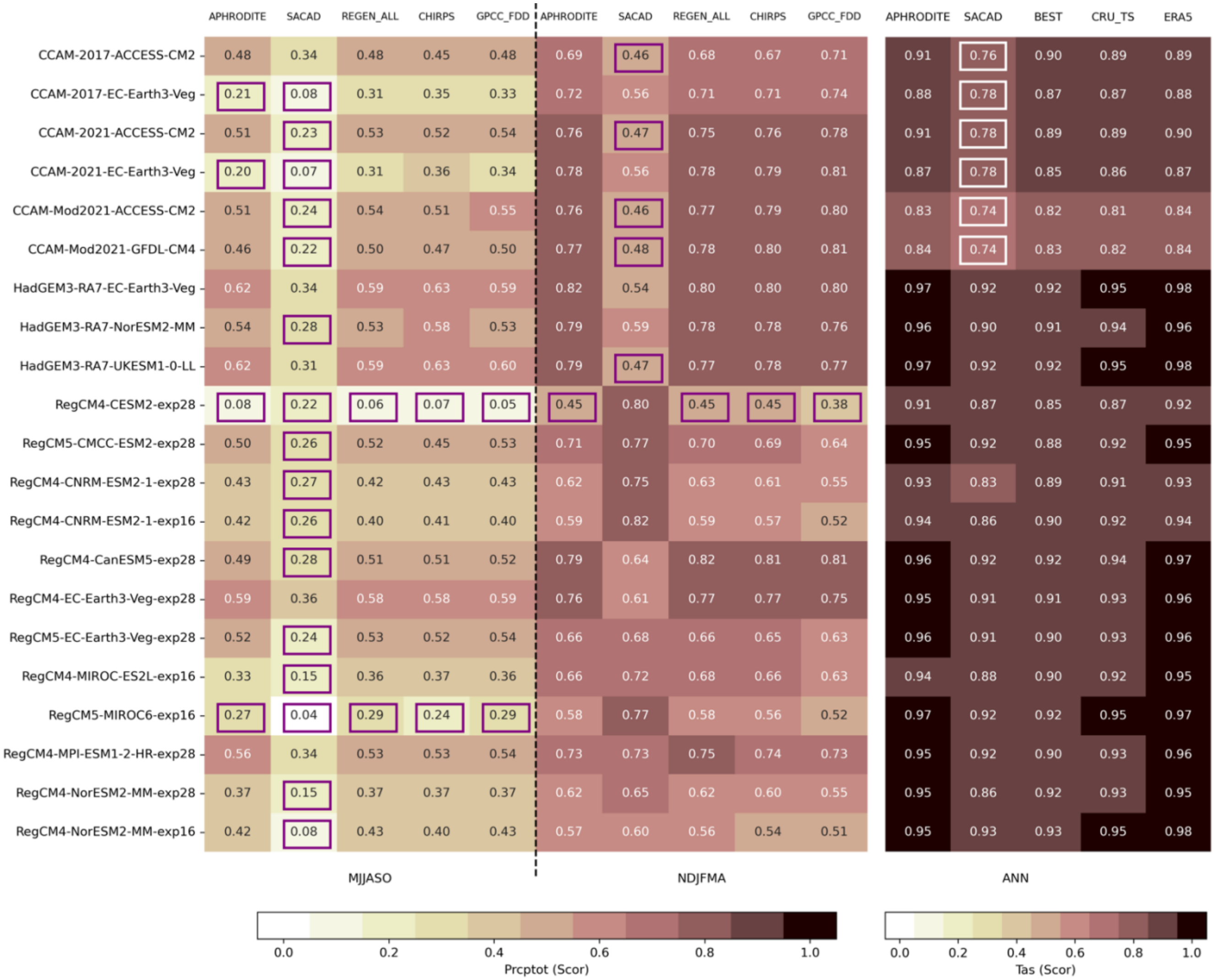

Figure 10. Score benchmarks for spatial correlation of seasonal total precipitation (prcptot) and annual daily mean near-surface temperature (*tas*) with multiple observational datasets. Scores less than the benchmark are marked by squares. Darker colours typically indicate better performance, whereas lighter colours typically indicate worse performance.

3.3.2 Seasonal cycle benchmark

The seasonal cycle benchmark follows the recommendation of Isphording et al. (2024) for unimodal seasonal patterns. This criterion requires that the three-month observed low and high peaks occur within the modelled lowest and highest six months of the year, respectively. The benchmark was applied to both precipitation and temperature and evaluated separately for two sub-regions to reflect differences in the timing of seasonal peaks. Figure 11 presents the months satisfying this benchmark for each variable. Colours and numbers denote the ranking of months from driest/coldest (1) to wettest/hottest (12), based on observed seasonal cycles derived from all reference datasets.

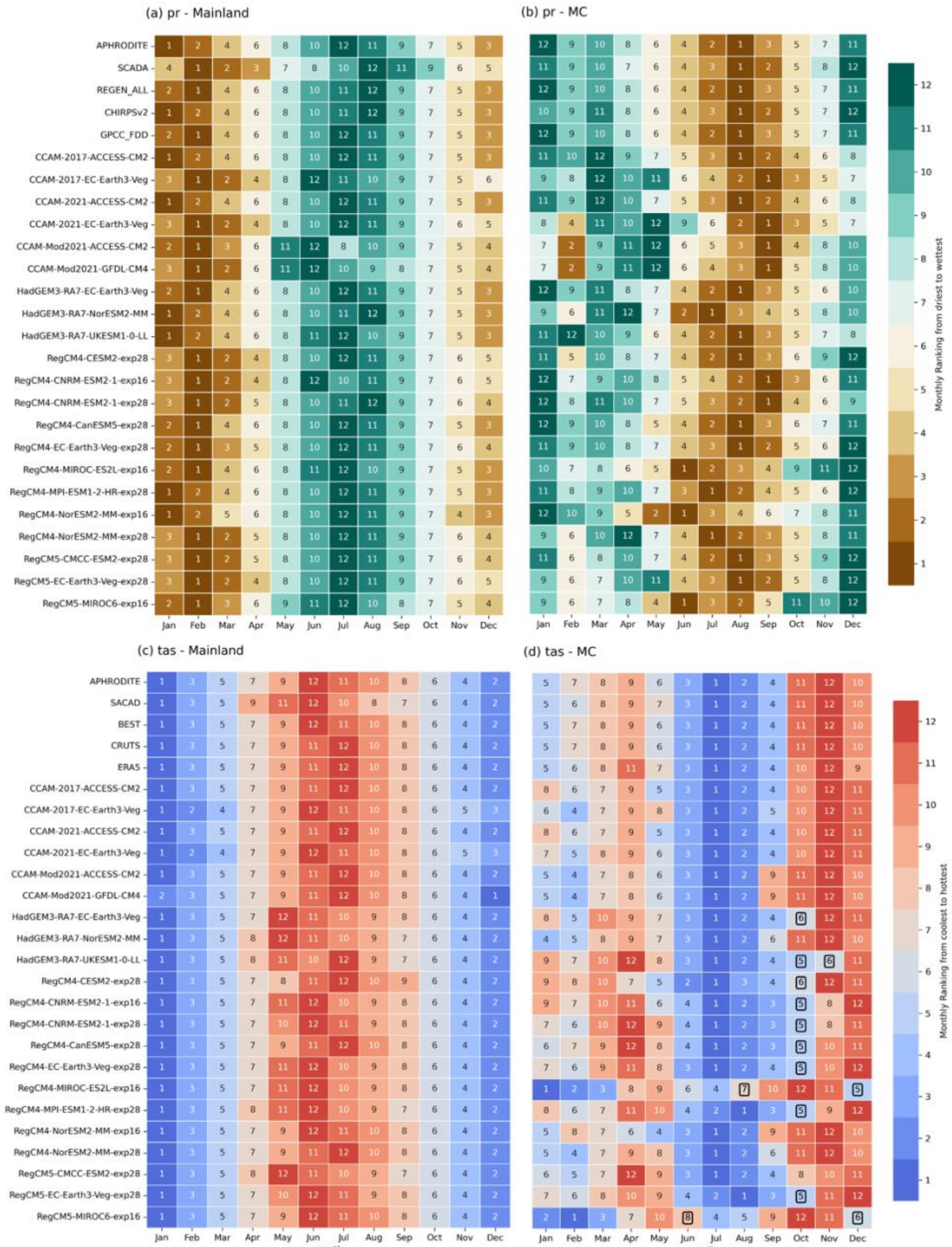

Figure 11. Seasonal cycle benchmark following Isphording et al. (2024) for the unimodal seasonal cycle of (a-b) precipitation (pr) and (c-d) near-surface mean temperature (*tas*) in the Mainland and MC. Colours and numbers indicate the ranking for latitude-weighted monthly total precipitation/daily temperature, with 1 being the driest/coolest and 12 being the wettest/hottest. The values highlighted in the black box indicate that the simulation does not meet our benchmarking expectations.

Observations reveal consistent timing of seasonal peaks: the wettest period occurs from June to August over the Mainland and from December to April over the MC, corresponding to the prevailing monsoon regimes. For temperature, the hottest months occur in June–August over the mainland and in October–December over the MC.

Overall, all models meet the precipitation benchmark across both sub-regions. However, some RegCM simulations display a weaker representation over the MC, with anomalously dry months during the observed wet season. As shown in Fig. S12, this discrepancy arises from shifts in the simulated low and high rainfall peaks. For temperature, all models satisfy the benchmark over the Mainland, which exhibits a clear unimodal cycle. Note that the MC spans the equator, where the semi-annual overhead passage of the sun naturally drives a bimodal surface temperature regime (Li et al., 2018). This regime is characterised by two periods of rapid warming in March–April and October–December, respectively, and a period of rapid cooling in June–August. In addition, changes in oceanic dynamics (including ocean heat content change) and the interaction with localised monsoonal wind shifts and regional cloud-convection feedbacks complicate this bimodal cycle (Aldrian et al., 2004; Chang et al., 2005). Given that the original benchmark assumes a strict unimodal annual cycle, this benchmark is not strictly applicable there. Thus, all models are considered to pass for the BMF of the temperature seasonal cycle over the MC, although 11 of 21 simulations fail to capture the observed solar-irradiance peak in Oct–December (ranks 10-11–12) over the MC, instead simulating cooler conditions around ranks 5–6.

3.3.3 Long-term trend benchmark

This benchmark is based on the expectation that simulating observed large-scale climate variability and change is crucial for climate change adaptation. Trends were considered in terms of time series of regional-averaged total precipitation and temperature, rather than at a single grid, given the substantial variability arising from small-scale processes or model parameterisations, which may not be robust across models or observational datasets.

To reduce the impact of decadal variability, we have extended the full historical period, from 1960 to 2014. The trend is estimated using a Theil-Sen slope estimator (Theil, 1950; Sen, 1968). Confidence intervals are computed with a significant value of $p<0.05$. Confidence intervals are marked as error bars on Fig. 12 for reference-based values and for all simulations. We test the significance of the trend using the Mann-Kendall significance test at the 5% level (Hussain and Mahmud, 2019).

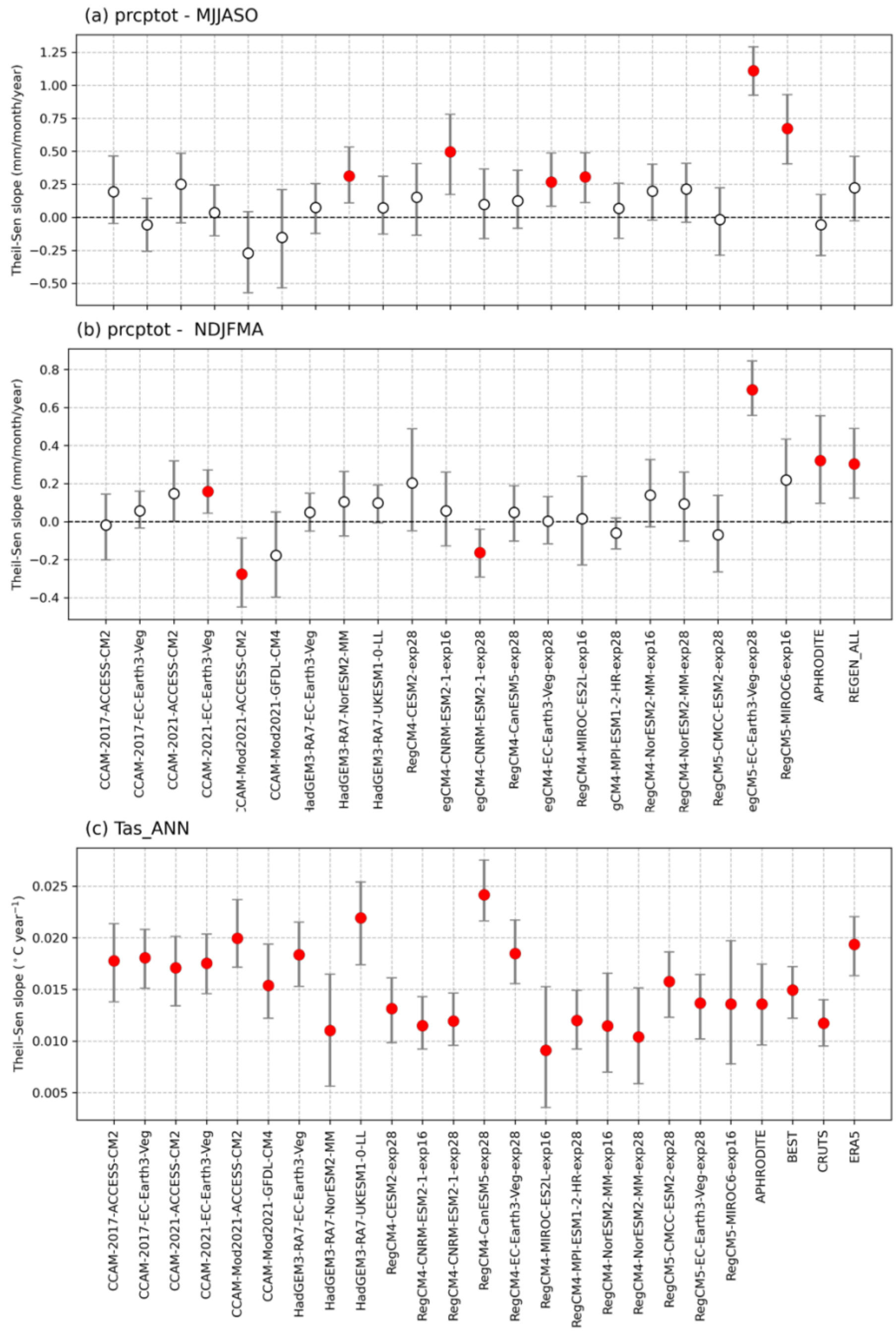

Figure 12. Trend estimated based on the Theil-Sen slope estimator of the time series of seasonal total wet-day precipitation (a–b) and annual near-surface mean temperature (c) among models and observations during 1960-2014. The red dot indicates a significant trend, as determined by the Mann-Kendall test at a 5% level of confidence. The bar indicates 90% confidence intervals according to Theil-Sen slope estimator.

Figures 12a–b show the long-term trend for the annual time series of seasonal average total precipitation, while Fig. 12c shows the trend for the annual time series of annual average daily mean near-surface temperature. In terms of total wet-day precipitation, APHRODITE and REGEN_ALL show different directions in the long-term MJJASO trend, although neither trend is significant (Fig. 12a). It is acknowledged that APHRODITE was merged from two versions (see Sect. 2.2), which have slight differences in data sources and algorithms for periods before and after 2007; thus, trends in APHRODITE should be considered with care. Our benchmarking threshold is therefore "no trend" (since neither one of the observed datasets has a significant trend), and all RCM simulations pass this benchmark during the MJJASO season. Meanwhile, during the NDJFMA season, most RCM simulations underestimate the magnitude of the observed positive trends found in APHRODITE and REGEN_ALL (e.g., 0.36 and 0.32 mm/year, respectively, Fig. 12b). All RCMs pass the trend benchmark for the dry season except RegCM4-CNRM-ESM-1-2-exp28 and CCAM-Mod2021-ACCESS-CM2, which have significant negative trends. Note that in both seasons, the confidence interval in observations and models is quite wide, reflecting an uncertain change signal. For temperature, all models and references show a significant increasing trend, coincident with global warming. Therefore, only two models: RegCM4-CNRM-ESM2-1-exp28 and CCAM-Mod2021-ACCESS-CM2, fail this benchmark.

Note that IPCC AR6 [Chapter 11, Seneviratne et al. (2021)] documented a positive precipitation trend over the MC after 1980 with high confidence. Therefore, we also conducted additional trend tests over a shorter 33-year period (1982 to 2014), coinciding with the availability of all reference products (Figs. S14–S15). Although the trends in observations vary from non-significant to significant positive for precipitation, no model fails this benchmark. Although these results somehow contrast the trend analysis over a longer period, RegCM4-CNRM-ESM2-1-exp28 and CCAM-Mod2021-ACCESS-CM2 fail another benchmark, so they are still removed from further analysis.

Table 4 summarises the results of 21 CMIP6 CORDEX-SEA simulations using the MSM from BMF. At the point of applying BMF, we find 15 simulations (highlighted by blue in Table 4) that meet our expectations in simulating the fundamental characteristics of precipitation and temperature over Southeast Asia.

Table 4. Summary of the BMF results. An "x" denotes models that pass the benchmark, while "–" indicates models that do not pass the benchmark. The bold number in the "total" column highlights the models that pass all benchmarks.

Model	precipitation							temperature				Total /11
	MJJASO			NDJFMA			Scyl	ANN				
	MAPE	Scor	Trend	MAPE	Scor	Trend		MAE	Scor	Scyl	Trend	
CCAM-2017-ACCESS-CM2	x	x	x	x	x	x	x	x	x	x	x	**11**
CCAM-2017-EC-Earth3-Veg	x	x	x	x	x	x	x	x	x	x	x	**11**
CCAM-2021-ACCESS-CM2	x	x	x	x	x	-	x	x	x	x	x	10
CCAM-2021-EC-Earth3-Veg	x	x	x	x	x	x	x	x	x	x	x	**11**
CCAM-Mod2021-ACCESS-CM2	x	x	x	x	x	x	x	x	x	x	x	**11**
CCAM-Mod2021-GFDL-CM4	x	x	x	x	x	x	x	x	x	x	x	**11**
HadGEM3-RA7-EC-Earth3-Veg	x	x	x	x	x	x	x	x	x	x	x	**11**
HadGEM3-RA7-NorESM2-MM	x	x	x	x	x	x	x	x	x	x	x	**11**
HadGEM3-RA7-UKESM1-0-LL	x	x	x	x	x	x	x	x	x	x	x	**11**
RegCM4-CESM2-exp28	-	-	x	-	-	x	x	x	x	x	x	7
RegCM5-CMCC-ESM2-exp28	x	x	x	x	x	x	x	x	x	x	x	**11**
RegCM4-CNRM-ESM2-1-exp28	x	x	x	-	x	-	x	x	x	x	x	9
RegCM4-CNRM-ESM2-1-exp16	x	x	x	-	x	x	x	x	x	x	x	10
RegCM4-CanESM5-exp28	x	x	x	x	x	x	x	x	x	x	x	**11**
RegCM4-EC-Earth3-Veg-exp28	x	x	x	x	x	x	x	x	x	x	x	**11**
RegCM5-EC-Earth3-Veg-exp28	x	x	x	x	x	x	x	x	x	x	x	**11**
RegCM4-MIROC-ES2L-exp28	-	x	x	x	x	x	x	x	x	-	x	9
RegCM5-MIROC6-exp28	-	x	x	-	x	x	x	x	x	-	x	8
RegCM4-MPI-ESM1-2-HR-exp28	x	x	x	x	x	x	x	x	x	x	x	**11**
RegCM4-NorESM2-MM-exp28	x	x	x	x	x	x	x	x	x	x	x	**11**
RegCM4-NorESM2-MM-exp16	x	x	x	x	x	x	x	x	x	x	x	**11**

3.4 Future climate change signal and model dependence.

In this section, we examine the projected future climate change spread from the CMIP6 CORDEX-SEA ensemble, which provides simulations under the SSP3-7.0 scenario (Table 1). Fig. 13 shows the distribution of projected changes in mean precipitation (y-axis) versus mean temperature (x-axis) during the wet and dry seasons. In this study, we classify regional temperature change into low, medium, and high warming categories. Specifically, low, medium, and high regional warming are defined as regional mean temperature increases of approximately +1.5°C, +2°C, and > +3°C relative to the baseline period

(1981-2010), respectively. Meanwhile, precipitation thresholds are set at dry (below 5%), neutral (5–15%), and wet (above 15%), respectively.

In general, both variables exhibit an increasing tendency, with temperature changes sitting in the mid to high ranges (e.g., changes from 2 to 5°C) and precipitation changes from 1.5 to 35%. However, several RegCM experiments (RegCM4-MPI-ESM1-2-HR-exp28, RegCM4-NorESM2-MM-exp28, and RegCM4-CanESM5-exp28) indicate decreasing trends in daily precipitation during NDJFMA, despite lying within the mid- to low-range precipitation projections during MJJASO.

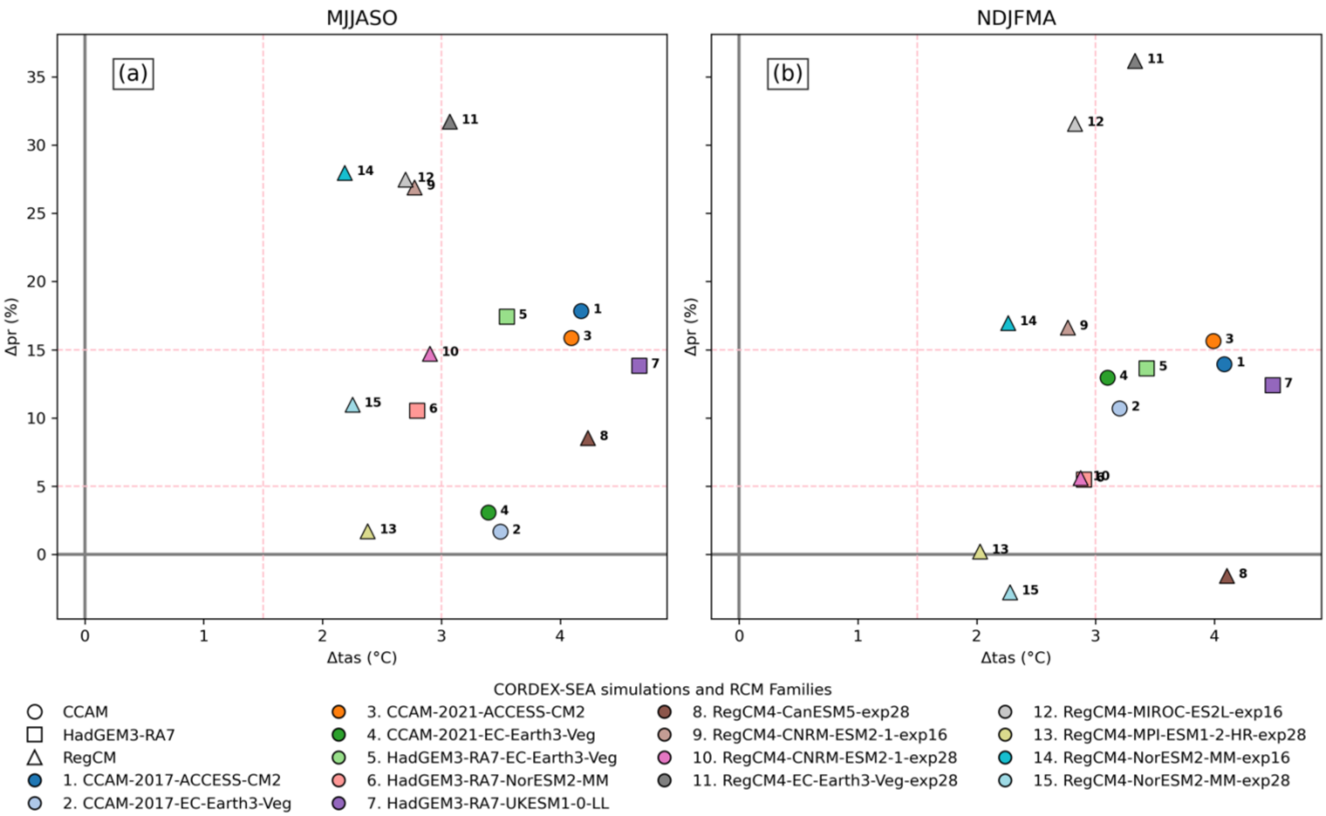

Figure 13. CMIP6 CORDEX-SEA ensemble future spread (2070–2099 relative to 1981–2010) over the land of Southeast Asia during (a) the MJJASO and (b) the NDJFMA seasons. The analysis is conducted for simulations that are available for SSP3-7.0 scenarios (Table 1). Vertical and horizontal dashed red lines indicate threshold values used to separate low, medium, and high ranges of projected temperature and precipitation changes. Temperature thresholds are set at +1.5°C, +2°C, and >+3°C, and precipitation thresholds are set at dry: <5%, neutral: 5–15%, and wet: >15%, respectively.

Among the 15 models passing the BMF criteria, 11 experiments provide SSP3-7.0 simulations. These experiments cluster around the mid-range of temperature projections but span a wide range of precipitation responses. Wet models projecting precipitation increases include RegCM4-EC-Earth3-Veg-exp28, RegCM4-NorESM2-MM-exp16 (e.g. above 15%), CCAM-

2017-ACCESS-CM2, CCAM-2021-ACCESS-CM2, HadGEM3-RA7-EC-Earth3-Veg, HadGEM3-RA7-NorESM2-MM, HadGEM3-RA7-UKESM1-0-LL, and RegCM4-NorESM2-MM-exp28 (e.g., 5–15%); while no or dry changes (< 5%) and are represented by CCAM-2017-EC-Earth3-Veg, RegCM4-MPI-ESM1-2-HR-exp28, and RegCM4-CanESM5-exp28.

In the next step, the seasonal spatial bias patterns (e.g., Figs. 2, 6 and 7) were subjected to hierarchical clustering using Pearson's correlation matrix as the distance measure between models. This aims to group similar spatial patterns of bias. Clusters shown in the dendrogram (Fig. 14) are designated by different colours. The horizontal axis on the dendrogram is a measure of similarity between individual models and clusters, where models positioned on the same branches share spatial patterns similarities. Hierarchical clustering of historical NDJFMA total wet-day precipitation and annual mean temperature reveals three dominant spatial clusters (Figs. 14b-c). These cluster, distinguished by three distinct colours, broadly corresponding to the RCM families: CCAM, HadGEM3-RA7, and RegCM (Figs. 14b, c). In contrast, the MJJASO precipitation pattern (Fig. 14a) exhibits a strong seasonal dependence, forming three distinct clusters that separate CCAM, specific RegCM simulations (e.g., RegCM4 forced by NorESM2-MM using experiment No.16, see Ngo-Duc et al., 2024; MICROC-ES2L, CESM2, and RegCM5 forced by MIROC6), and HadGEM3-RA7 simulations grouped alongside the majority of RegCM. This seasonal variation in clustering behaviour aligns with the findings of Nguyen et al. (2024) and Gibson et al. (2023).

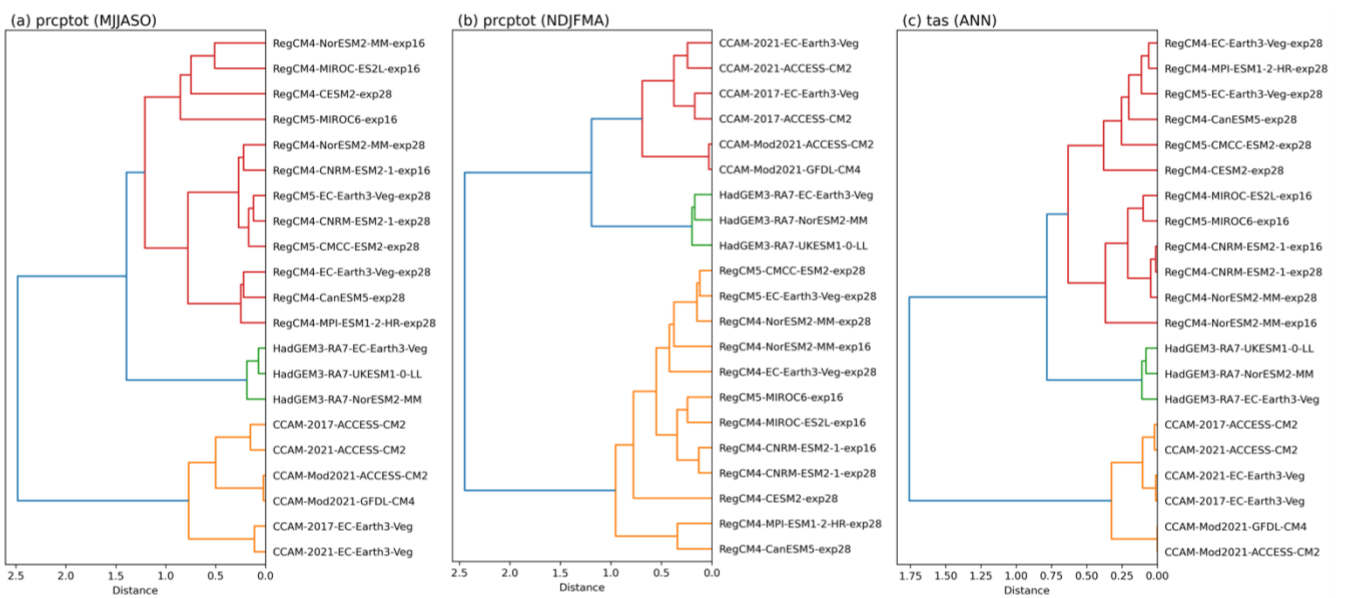

Figure 14. Dendrogram with hierarchical clustering of spatial patterns of historical seasonal total wet-day precipitation (a-b) and annual daily mean temperature (c). Colours indicate the different cluster assignments. The horizontal axis on the dendrogram is a measure of similarity between individual models and clusters, where models aligned on the same branches have similar spatial patterns.

Overall, based on model performance, future spread under the SSP3-7.0 scenario, model dependency, and data availability, we select at least two experiments from three independent model groups to be considered for further km-scale dynamical downscaling over SEA megacities. These include:

- RegCM4: EC-Earth3-Veg-exp28 – high range; NorESM2-MM-exp28 – mid range; MPI-ESM1-2-HR-exp28 and CanESM5-exp28 – low range
- HadGEM3-GA7: EC-Earth3-Veg and UKESM1-0-LL—mid-range; and
- CCAM-2017: ACCESS-CM2 – mid-range; EC-Earth 3-Veg – low range

4. Discussion and Conclusions

The overall aims of this study were to (1) document the experimental design for producing the updated CMIP6 dynamical downscaled climate simulations for Southeast Asia (SEA); (2) comprehensively assess model performance over the historical period across climatological fields of model core variables: daily precipitation, near-surface temperature (*tas*), daily maximum temperature (*tasmax*), and daily minimum temperature (*tasmin*); and (3) select a subset of RCMs for further dynamical downscaling at km-scale over five megacities in SEA based on benchmarking model performance, model independence, and a range of future model responses.

Observational uncertainties are an important aspect in the context of model evaluation and benchmarking. Over SEA, observations are sparse and have large uncertainties (Nguyen et al., 2020; Alexander et al., 2020), complicating model assessment and benchmarking (Nguyen et al., 2024). Note that the observational uncertainties are quite large compared with benchmarks, reaching greater than 50% of benchmarking errors (MAPE and MAE) in the wet season for total precipitation and for temperature at annual scales, particularly over the MC (Table S1). Therefore, benchmarking based on MJJASO seasonal rainfall and annual mean temperature should be interpreted with caution. Uncertainty in spatial correlation is low for both temperature and rainfall, ranging from 2% to 10% of the benchmarking values. On the other hand, all observational products show a similar trend (Fig. 12 and Figs. S14-S15) and seasonal cycle (Figs. S4 and S12 for temperature and precipitation, respectively). To deal with observational uncertainty, model biases are reassessed with multiple observations from different sources (e.g., in situ, blended satellite, and reanalysis datasets; global and high-resolution regional gridded datasets). The following are key findings of this evaluation:

(1) Despite the large observational uncertainties in precipitation intensity, the CMIP6-downscaled CORDEX-SEA simulations capture the spatial distribution and seasonality of precipitation over SEA reasonably well. However, they tend to substantially overestimate observed precipitation (e.g., at least 14 out of 21 models regardless of the choice of observed reference). The magnitude of biases is both regionally and seasonally heterogeneous, with larger wet biases over regions and seasons

dominated by the prevailing monsoon (e.g., over the MC during NDJFMA and over the Mainland during MJJASO). These wet biases are particularly pronounced in the CCAM and HadGEM3-RA7 simulations. This requires further investigation using process-based metrics, such as 850 hPa wind and teleconnection of precipitation with the IOD and ENSO which assess the ability of the RCM ensemble to present the key physical processes driving precipitation variability and extremes over SEA (Nguyen et al., 2024).

(2) The CMIP6 CORDEX-SEA ensemble reproduces the observed spatial distribution of temperature reasonably well but generally exhibits cold biases, particularly over the Mainland and during the boreal winter (DJF) season, as well as in simulations that overestimate precipitation (e.g., CCAM and HadGEM3-RA7 simulations). Comparisons with the BEST, SACAD and CRUTS datasets indicate that these cold biases are primarily linked to an underestimation of *tasmax* rather than *tasmin*. Previous studies also reported persistent cold biases in CORDEX-Australasia simulations of *tasmax*, most notably over western, northern and eastern Australia during the wet season (Schroeter et al., 2024; Chapman et al., 2023; Di Virgilio et al., 2019), which strongly correlate with wet biases in precipitation.

These biases might be associated with an overestimation of high-level cloud cover inherent to the uncoupled, atmosphere-only configuration of RCM used (Van der Linden et al., 2019). Unlike coupled systems, where intense rainfall and heavy cloud cover induce a negative feedback loop by cooling the underlying ocean mixed layer, the prescribed SST boundaries in these simulations act as an infinite thermodynamic reservoir of heat and moisture. Lacking interactive ocean-cooling feedback, the uncoupled RCMs continuously simulate excessive daytime maritime convection. This generates persistent regional cloud decks that steadily block daytime shortwave solar radiation over land, driving down *tasmax.* Di Virgilio et al. (2019) further suggested that excessive soil moisture might contribute to this cold bias through enhanced evaporative cooling. Another complementary hypothesis points to convective parameterisations that trigger premature, massive moisture dumps. As suggested by Howard et al. (2024), this process leads to a persistent cold bias in *tasmax* and a compressed diurnal temperature range due to overestimated daytime cloud block.

While these physical processes in RCMs could help to explain the wet biases, part of the apparent model bias may also reflect the uncertainty in the reference dataset used for evaluation. Notably, observational coverage over Southeast Asia is relatively sparse, particularly over MC, and both precipitation and temperature are subject to uncertainties (Fig. S1) associated with gauge density and quality (Nguyen et al., 2020), interpolation method and retrieval technique. For example, using ERA5 as a reference displays an opposite pattern, with biases mainly arising from the underestimation of *tasmin*, rather than *tasmax* (Figs. S2 and S7). Reanalysis products assimilate a range of observations within a modelling framework and therefore contain their own structural uncertainties. Taken together, those differences highlight that part of diagnosed bias is sensitive to the choice of reference dataset. Further investigation is therefore warranted to better disentangle model-related deficiencies from observational uncertainties and to fully explore the physical mechanisms behind the cold biases in RCMs and their potential link with wet biases in precipitation over SEA.

Overall, the regional models demonstrate higher skill in simulating temperature characteristics than precipitation over SEA. This is consistent with findings from CMIP6 performance assessments over SEA (Nguyen et al., 2024) and from studies in other regions, such as over Australia (Evans et al., 2021) and Africa (Dosio et al., 2015), where the climate models generally reproduce temperature more reliably than precipitation, which remains more challenging due to its strong dependency on convection, local processes, and topography.

(3) Comparisons among simulations driven by the same GCM or originating from the same RCM family reveal that runs from a given RCM family often exhibit similar spatial patterns and bias characteristics. This indicates that RCM configuration exerts a stronger influence on model biases than the choice of driving GCM. This is inline with previous studies on CMIP5-downscaled simulations over SEA (Nguyen et al., 2022) or over other CORDEX domains, e.g., Australia (Di Virgilio et al., 2025), Europe (Kotlarski et al., 2014; Vautard et al., 2021). Within the CMIP6 CORDEX-SEA ensemble, the CCAM-2017 simulations outperform the CCAM-2021 versions. Meanwhile, the RegCM5 simulation forced by EC-Earth3-Veg shows notable improvements in reducing wet biases compared with RegCM4. These results are based on analyses of a 33-year climatological period (1982–2014), which differs from Ngo-Duc et al. (2024), who reported no clear improvement in RegCM5 relative to RegCM4-NH based on 5-year ERA5-forced simulations. The discrepancies likely arise from the length of the evaluation period. As only one EC-Earth3-Veg–forced RegCM5 simulation is currently available, additional RegCM5 simulations with the same forcing GCMs could be further conducted to confirm these improvements compared to RegCM4 and enable more robust intercomparison.

(4) The MC remains a particularly challenging region due to its complex geography, consisting of numerous small islands and steep topography. Consequently, models struggle to accurately capture the spatial variability of rainfall and the internal variability of key climate variables (e.g., the seasonal cycle of near-surface temperature).

Given the goal of selecting RCM simulations to be considered for km-scale dynamical downscaling over SEA megacities, we applied a novel benchmarking framework (BMF)—a systematic approach designed to identify a subset of fit-for-purpose models that meet predefined performance expectations. We acknowledge that whether a model passes the BMF depends on how these performance expectations and reference datasets are defined. Therefore, the benchmarking thresholds were developed in consultation with model developers and regional stakeholders within CORDEX-SEA, using multiple observational references. Importantly, assessment against a single reference dataset was not considered sufficient for model inclusion or exclusion. Instead, a model was excluded only if its skill scores exceeded predefined thresholds for at least half of the reference datasets (i.e., three out of five in this study).

It is worth noting that Nguyen et al. (2024) also established performance thresholds for CMIP6 GCMs to identify models that meet regional performance expectations over SEA, particularly in simulating precipitation, its key drivers, and associated teleconnections. However, since the number of RCM simulations in this study is smaller than the GCM ensemble used by Nguyen et al. (2024) (e.g., 32 CMIP6 GCMs), slightly more relaxed performance thresholds were adopted. In addition, the objective here is to select suitable RCMs for km-scale downscaling rather than to assess whether RCMs outperform their driving GCMs.

All RCMs were evaluated using Minimum Standard Metrics (MSMs) for *tas* and precipitation, resulting in 15 of 21 simulations meeting the BMF. Subsequent assessments of model independence revealed strong similarities among simulations from the same RCM family. Analysis of the projected future response spread led to the identification of three independent model groups suitable for further downscaling: RegCM4 (EC-Earth3-Veg, MPI-ESM1-2-HR, CanESM5, and NorESM2-MM with exp28); HadREM3-GA7 (EC-Earth3-Veg and UKESM1-0-LL); and CCAM-2017 (EC-Earth3-Veg and ACCESS-CM2). It is worth noting that among these simulations, a few, namely CCAM-2017-ACCESS393 CM2, CCAM-2017-EC-Earth3-Veg, and HadGEM3-RA7-EC-Earth3-Veg, exhibit distribution consistently close to multiple observed references. These selected models best align with our performance-based expectations and provide a balanced representation of regional processes for future high-resolution downscaling across the five megacities of Southeast Asia. We do note that benchmarking was done at a regional scale (e.g., over the Mainland, MC, and SEA) rather than at a city scale, considering the limited availability of reliable observational datasets in the SEA megacities. Recognizing the high climate variability in SEA, supplementary analysis can also be done for subdomains or specific areas of interest in SEA.

This study presents the first assessment of the CMIP6 CORDEX-SEA ensemble, with a focus on individual ensemble member performance rather than ensemble mean skill in simulating temperature and precipitation. This will improve the utility of CMIP6 CORDEX-SEA projections for climate services by clarifying the reliability and robustness of different RCM families. By identifying the strengths and weaknesses of each model, we contribute to the ongoing efforts of regional climate modelling, highlighting the key processes, variables, and sub-regions that require improvement. These insights inform the future model development while supporting more informed decision-making at the regional scale. We also introduce a benchmarking approach in which we utilise the assessment results to select models for further downscaling over SEA megacities, supporting climate risk assessment for climate adaptation and mitigation over these urban areas. Notably, Singapore's Third National Climate Change Study [V3; CCRS (2024)] has recently provided high-resolution climate change projections for Singapore and the broader SEA region through dynamic downscaling of coarse-resolution CMIP6 simulations. The framework developed in this study can be readily applied as additional ensemble members become available, particularly considering the recent inclusion of V3 data in CORDEX-SEA.

Code and Data Availability

Code for evaluation and benchmarking the CMIP6-downscaled CORDEX-Southeast Asia (SEA) performance is available at https://doi.org/10.5281/zenodo.8365065 (Isphording, 2023).

Model source codes used in this study are available as follows:

- RegCM4 is available at https://zenodo.org/records/4603556 (Coppola et al., 2021)
- RegCM5 is available at https://zenodo.org/records/17348623 (Giorgi et al., 2024)
- CCAM-2307 is available at https://doi.org/10.5281/zenodo.19303856 (Nguyen et al., 2026a)
- HadGEM3-RA07 is available at https://doi.org/10.5194/gmd-16-1713-2023 (Bush et al., 2023).

Data used in the production of this study are archived in repositories as follows:

- Simulation data: https://doi.org/10.5281/zenodo.19334179 (Nguyen et al., 2026b)
- Reference dataset: https://doi.org/10.5281/zenodo.19334323 (Nguyen et al., 2026c)

Author contributions

PLN and LVA designed the study, carried out the analysis, and wrote the initial manuscript draft. Other co-authors provided supervision and contributed to manuscript review and revisions.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Acknowledgment

UNSW-CCRC CCAM experiments were performed on the computational facility of the National Computational Infrastructure (NCI) through the UNSW HPC Scheme (DOI: 10.26190/PMN5-7J50). We would also like to thank the NCI HPC (Gadi) for the computational resources.

Financial support

PLN and LVA are supported by Australian Research Council (ARC) grant FT210100459. LVA is also supported by ARC grant CE230100012. TND is supported by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under Grant 105.06-2021.14. FC, JMD and JT are supported by the "High-Definition Clean Energy, Climate, and Weather Forecasts for the Philippines" project of the Manila Observatory. The downscaled CMIP6 data were obtained from the SEACLID/CORDEX-SEA project funded by the Asia Pacific Network for Global Change Research (CRRP2023-08MY-Cruz).

References

Aldrian, E., Dümenil-Gates, L., Jacob, D., Podzun, R., and Gunawan, D.: Long-term simulation of Indonesian rainfall with the MPI regional model, Climate Dynamics, 22, 795-814, 10.1007/s00382-004-0418-9, 2004.

Alexander, L. V., Bador, M., Roca, R., Contractor, S., Donat, M. G., and Nguyen, P. L.: Intercomparison of annual precipitation indices and extremes over global land areas from in situ, space-based and reanalysis products, Environmental Research Letters, 15, 055002, 10.1088/1748-9326/ab79e2, 2020.

Alexander, L. V., Nguyen, P. L., Donat, M. G., Dunn, R. J. H., Tett, S., Zhang, X., Alves, L. M., Bador, M., Deng, X., Gibson, P. B., King, A., Lennard, C., Min, S.-K., Roca, R., and Trewin, B.: Less Intense Daily Precipitation Maxima in Regional Compared to Global Gridded Products, Journal of Climate, 38, 7669-7693, https://doi.org/10.1175/JCLI-D-25-0222.1, 2025.

APN: Climatic hazard assessment to enhance resilience against climate extremes for Southeast Asian megacities (CARE for SEA megacities), https://doi.org/10.30852/p.26687, 2023.

Best, M. J., Pryor, M., Clark, D., Rooney, G. G., Essery, R. L., Ménard, C., Edwards, J., Hendry, M., Porson, A., and Gedney, N.: The Joint UK Land Environment Simulator (JULES), model description–Part 1: energy and water fluxes, Geoscientific Model Development, 4, 677-699, 2011.

Bretherton, C. S., Uttal, T., Fairall, C. W., Yuter, S., Weller, R., Baumgardner, D., Comstock, K., and Wood, R.: The EPIC 2001 stratocumulus study, 2004.

Brunner, L., Pendergrass, A. G., Lehner, F., Merrifield, A. L., Lorenz, R., and Knutti, R.: Reduced global warming from CMIP6 projections when weighting models by performance and independence, Earth Syst. Dynam., 11, 995-1012, 10.5194/esd-11-995-2020, 2020.

Buonomo, E., Savage, N., Dong, G., Becker, B., Jones, R. G., Tian, Z., and Sun, L.: Tropical cyclone changes in convection-permitting regional climate projections: A study over the Shanghai region, Journal of Geophysical Research: Atmospheres, 129, e2023JD038508, 2024.

CCRS: Singapore's Third National Climate Change Study (V3): Climate Change Projections to 2100 — Science (Technical) Report, https://www.mss-int.sg/docs/default-source/v3_reports/v3_science_report/v3-science-report-full.pdf. , 2024.

Chang, C. P., Wang, Z., McBride, J., and Liu, C.-H.: Annual Cycle of Southeast Asia—Maritime Continent Rainfall and the Asymmetric Monsoon Transition, Journal of Climate, 18, 287-301, https://doi.org/10.1175/JCLI-3257.1, 2005.

Chapman, S., Syktus, J., Trancoso, R., Thatcher, M., Toombs, N., Wong, K. K.-H., and Takbash, A.: Evaluation of Dynamically Downscaled CMIP6-CCAM Models Over Australia, Earth's Future, 11, e2023EF003548, https://doi.org/10.1029/2023EF003548, 2023.

Clark, D., Mercado, L., Sitch, S., Jones, C., Gedney, N., Best, M., Pryor, M., Rooney, G., Essery, R., and Blyth, E.: The Joint UK Land Environment Simulator (JULES), model description–Part 2: carbon fluxes and vegetation dynamics, Geoscientific Model Development, 4, 701-722, 2011.

Contractor, S., Donat, M. G., Alexander, L. V., Ziese, M., Meyer-Christoffer, A., Schneider, U., Rustemeier, E., Becker, A., Durre, I., and Vose, R. S.: Rainfall Estimates on a Gridded Network (REGEN) – a global land-based gridded dataset of daily precipitation from 1950 to 2016, Hydrol. Earth Syst. Sci., 24, 919-943, https://doi.org/10.5194/hess-24-919-2020, 2020.

Coppola, E., Stocchi, P., Pichelli, E., Torres Alavez, J. A., Glazer, R., Giuliani, G., Di Sante, F., Nogherotto, R., and Giorgi, F.: Non-Hydrostatic RegCM4 (RegCM4-NH): model description and case studies over multiple domains, Geosci. Model Dev., 14, 7705-7723, 10.5194/gmd-14-7705-2021, 2021.

CORDEX: CORDEX experiment design for dynamical downscaling of CMIP6, https://cordex.org/wp-content/uploads/2021/05/CORDEX-CMIP6_exp_design_RCM.pdf, 2021.

Cruz, F. T., Narisma, G. T., Dado, J. B., Singhruck, P., Tangang, F., Linarka, U. A., Wati, T., Juneng, L., Phan-Van, T., Ngo-Duc, T., Santisirisomboon, J., Gunawan, D., and Aldrian, E.: Sensitivity of temperature to physical parameterization schemes of RegCM4 over the CORDEX-Southeast Asia region, International Journal of Climatology, 37, 5139-5153, https://doi.org/10.1002/joc.5151, 2017.

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

Di Virgilio, G., Evans, J. P., Di Luca, A., Olson, R., Argüeso, D., Kala, J., Andrys, J., Hoffmann, P., Katzfey, J. J., and Rockel, B.: Evaluating reanalysis-driven CORDEX regional climate models over Australia: model performance and errors, Climate Dynamics, 53, 2985-3005, 10.1007/s00382-019-04672-w, 2019.

Di Virgilio, G., Ji, F., Tam, E., Nishant, N., Evans, J. P., Thomas, C., Riley, M. L., Beyer, K., Grose, M. R., Narsey, S., and Delage, F.: Selecting CMIP6 GCMs for CORDEX Dynamical Downscaling: Model Performance, Independence, and Climate Change Signals, Earth's Future, 10, e2021EF002625, https://doi.org/10.1029/2021EF002625, 2022.

Di Virgilio, G., Evans, J. P., Ji, F., Tam, E., Kala, J., Andrys, J., Thomas, C., Choudhury, D., Rocha, C., White, S., Li, Y., El Rafei, M., Goyal, R., Riley, M. L., and Lingala, J.: Design, evaluation, and future projections of the NARCliM2.0 CORDEX-CMIP6 Australasia regional climate ensemble, Geosci. Model Dev., 18, 671-702, 10.5194/gmd-18-671-2025, 2025.

Doblas-Reyes, F. J., Sörensson, A. A., Almazroui, M., Dosio, A., Gutowski, W. J., Haarsma, R., Hamdi, R., Hewitson, B., Kwon, W. T., Lamptey, B. L., Maraun, D., Stephenson, T. S., Takayabu, I., Terray, L., Turner, A., and Zuo, Z.: Linking Global to Regional Climate Change, 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, edited by: 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., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1363–1512, https://doi.org/10.1017/9781009157896.012, 2021.

Dosio, A., Panitz, H.-J., Schubert-Frisius, M., and Lüthi, D.: Dynamical downscaling of CMIP5 global circulation models over CORDEX-Africa with COSMO-CLM: evaluation over the present climate and analysis of the added value, Climate Dynamics, 44, 2637-2661, 10.1007/s00382-014-2262-x, 2015.

Douville, H., Raghavan, K., Renwick, J., Allan, R. P., Arias, P. A., Barlow, M., Cerezo-Mota, R., Cherchi, A., Gan, T. Y., Gergis, J., Jiang, D., Khan, A., Pokam Mba, W., Rosenfeld, D., Tierney, J., and Zolina, O.: Water Cycle Changes, 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, edited by: 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., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1055–1210, https://doi.org/10.1017/9781009157896.010, 2021.

Evans, J. P., Di Virgilio, G., Hirsch, A. L., Hoffmann, P., Remedio, A. R., Ji, F., Rockel, B., and Coppola, E.: The CORDEX-Australasia ensemble: evaluation and future projections, Climate Dynamics, 57, 1385-1401, 10.1007/s00382-020-05459-0, 2021.

Freidenreich, S. and Ramaswamy, V.: A new multiple-band solar radiative parameterization for general circulation models, Journal of Geophysical Research: Atmospheres, 104, 31389-31409, 1999.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., and Michaelsen, J.: The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes, Scientific Data, 2, 150066, https://doi.org/10.1038/sdata.2015.66, 2015.

Gibson, P. B., Rampal, N., Dean, S. M., and Morgenstern, O.: Storylines for Future Projections of Precipitation Over New Zealand in CMIP6 Models, Journal of Geophysical Research: Atmospheres, 129, e2023JD039664, https://doi.org/10.1029/2023JD039664, 2024.

Giorgi, F. and Gutowski, W. J.: Coordinated Experiments for Projections of Regional Climate Change, Current Climate Change Reports, 2, 202-210, https://doi.org/10.1007/s40641-016-0046-6, 2016.

Giorgi, F., Jones, C., and Asrar, G.: Addressing climate information needs at the regional level: The CORDEX framework, WMO Bull, 53, 2008.

Giorgi, F., Coppola, E., Teichmann, C., and Jacob, D.: Editorial for the CORDEX-CORE Experiment I Special Issue, Climate Dynamics, 57, 1265-1268, https://doi.org/10.1007/s00382-021-05902-w, 2021.

Giorgi, F., Coppola, E., Giuliani, G., Ciarlo`, J. M., Pichelli, E., Nogherotto, R., Raffaele, F., Malguzzi, P., Davolio, S., Stocchi, P., and Drofa, O.: The Fifth Generation Regional Climate Modeling System, RegCM5: Description and Illustrative Examples at Parameterized Convection and Convection-Permitting Resolutions, Journal of Geophysical Research: Atmospheres, 128, e2022JD038199, https://doi.org/10.1029/2022JD038199, 2023.

Gregory, D. and Rowntree, P.: A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependent closure, Monthly Weather Review, 118, 1483-1506, 1990.

Grose, M. R., Narsey, S., Trancoso, R., Mackallah, C., Delage, F., Dowdy, A., Di Virgilio, G., Watterson, I., Dobrohotoff, P., Rashid, H. A., Rauniyar, S., Henley, B., Thatcher, M., Syktus, J., Abramowitz, G., Evans, J. P., Su, C.-H., and Takbash, A.: A CMIP6-based multi-model downscaling ensemble to underpin climate change services in Australia, Climate Services, 30, 100368, https://doi.org/10.1016/j.cliser.2023.100368, 2023.

Harris, I., Osborn, T. J., Jones, P., and Lister, D.: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset, Scientific Data, 7, 109, 10.1038/s41597-020-0453-3, 2020.

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., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999-2049, https://doi.org/10.1002/qj.3803, 2020.

Hoffmann, P., Katzfey, J. J., McGregor, J. L., and Thatcher, M.: Bias and variance correction of sea surface temperatures used for dynamical downscaling, Journal of Geophysical Research: Atmospheres, 121, 12,877-812,890, https://doi.org/10.1002/2016JD025383, 2016.

Hurley, P.: Modelling Mean and Turbulence Fields in the Dry Convective Boundary Layer with the Eddy-Diffusivity/Mass-Flux Approach, Boundary-Layer Meteorology, 125, 525-536, 10.1007/s10546-007-9203-8, 2007.

Hussain, M. and Mahmud, I.: pyMannKendall: a python package for non parametric Mann Kendall family of trend tests, Journal of open source software, 4, 1556, 2019.

IPCC: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 10.1017/9781009157896, 2021.

IPCC: Climate Change 2022: Impacts, Adaptation and Vulnerability Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge University Press, Cambridge, UK and New York, NY, USA, 3056 pp., 10.1017/9781009325844., 2022.

Isphording, R. N., Alexander, L. V., Bador, M., Green, D., Evans, J. P., and Wales, S.: A Standardized Benchmarking Framework to Assess Downscaled Precipitation Simulations, Journal of Climate, 37, 1089-1110, https://doi.org/10.1175/JCLI-D-23-0317.1, 2024.

Jiang, X., Howard, E., Su, C.-H., Isphording, R. N., Ng, B., Chapman, S., Ji, F., Grose, M., Syktus, J., Trancoso, R., Thatcher, M., Narsey, S., Di Virgilio, G., and Kala, J.: Towards benchmarking the dynamically downscaled CMIP6 CORDEX-Australasia ensemble over Australia, Journal of Southern Hemisphere Earth Systems Science, 75, 10.1071/es24050, 2025.

Juneng, L., Tangang, F., Chung, J. X., Ngai, S. T., Teh, T. W., Narisma, G., Cruz, F., Phan-Van, T., Ngo-Duc, T., Santisirisomboon, J., Singhruck, P., Gunawan, D., and Aldrian, E.: Sensitivity of Southeast Asia rainfall simulations to cumulus and air–sea flux parameterizations in RegCM4, Climate Research, 69, https://doi.org/10.3354/cr01386, 2016.

Kain, J. S.: The Kain–Fritsch Convective Parameterization: An Update, Journal of Applied Meteorology, 43, 170-181, https://doi.org/10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2, 2004.

Katzfey, J., Nguyen, K., McGregor, J., Hoffmann, P., Ramasamy, S., Nguyen, H. V., Khiem, M. V., Nguyen, T. V., Truong, K. B., and Vu, T. V.: High-resolution simulations for Vietnam-methodology and evaluation of current climate, Asia-Pacific Journal of Atmospheric Sciences, 52, 91-106, 2016.

Kiehl, J. T., Hack, J. J., Bonan, G. B., Boville, B. A., Williamson, D. L., and Rasch, P. J.: The National Center for Atmospheric Research Community Climate Model: CCM3, Journal of Climate, 11, 1131-1149, https://doi.org/10.1175/1520-0442(1998)011<1131:TNCFAR>2.0.CO;2, 1998.

Kotlarski, S., Keuler, K., Christensen, O. B., Colette, A., Déqué, M., Gobiet, A., Goergen, K., Jacob, D., Lüthi, D., van Meijgaard, E., Nikulin, G., Schär, C., Teichmann, C., Vautard, R., Warrach-Sagi, K., and Wulfmeyer, V.: Regional climate modeling on European scales: a joint standard evaluation of the EURO-CORDEX RCM ensemble, Geosci. Model Dev., 7, 1297-1333, 10.5194/gmd-7-1297-2014, 2014.

Kowalczyk, E., Wang, Y., Law, R., Davies, H., McGregor, J., and Abramowitz, G.: The CSIRO Atmosphere Biosphere Land Exchange (CABLE) model for use in climate models and as an offline model, 1615, 2006.

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., Collier, N., Ghimire, B., van Kampenhout, L., Kennedy, D., Kluzek, E., Lawrence, P. J., Li, F., Li, H., Lombardozzi, D., Riley, W. J., Sacks, W. J., Shi, M., Vertenstein, M., Wieder, W. R., Xu, C., Ali, A. A., Badger, A. M., Bisht, G., van den Broeke, M., Brunke, M. A., Burns, S. P., Buzan, J., Clark, M., Craig, A., Dahlin, K., Drewniak, B., Fisher, J. B., Flanner, M., Fox, A. M., Gentine, P., Hoffman, F., Keppel-Aleks, G., Knox, R., Kumar, S., Lenaerts, J., Leung, L. R., Lipscomb, W. H., Lu, Y., Pandey, A., Pelletier, J. D., Perket, J., Randerson, J. T., Ricciuto, D. M., Sanderson, B. M., Slater, A., Subin, Z. M., Tang, J., Thomas, R. Q., Val Martin, M., and Zeng, X.: The Community Land Model Version 5: Description of New Features, Benchmarking, and Impact of Forcing Uncertainty, Journal of Advances in Modeling Earth Systems, 11, 4245-4287, https://doi.org/10.1029/2018MS001583, 2019.

Li, Y., Yang, S., Deng, Y., Hu, X., and Cai, M.: A process-level attribution of the annual cycle of surface temperature over the Maritime Continent, Climate Dynamics, 51, 2759-2772, 10.1007/s00382-017-4043-9, 2018.

Maraun, D. and Widmann, M.: Cross-validation of bias-corrected climate simulations is misleading, Hydrol. Earth Syst. Sci., 22, 4867-4873, 10.5194/hess-22-4867-2018, 2018.

McGregor, J. L.: Economical determination of departure points for semi-Lagrangian models, Monthly Weather Review, 121, 221-230, 1993.

McGregor, J. L.: A new convection scheme using a simple closure, Current issues in the parameterization of convection, BMRC, Melbourne, Australia, Research Report, 93, 33-36, 2003.

McGregor, J. L.: C-CAM: Geometric aspects and dynamical formulation, CSIRO Atmospheric Research Dickson ACT, Australia, 2005.

McGregor, J. L.: Recent developments in variable-resolution global climate modelling, Climatic Change, 129, 369-380, 2015.

McGregor, J. L. and Dix, M. R.: An updated description of the conformal-cubic atmospheric model, in: High resolution numerical modelling of the atmosphere and ocean, Springer, 51-75, 2008.

Ngo-Duc, T., Tangang, F. T., Santisirisomboon, J., Cruz, F., Trinh-Tuan, L., Nguyen-Xuan, T., Phan-Van, T., Juneng, L., Narisma, G., Singhruck, P., Gunawan, D., and Aldrian, E.: Performance evaluation of RegCM4 in simulating extreme rainfall and temperature indices over the CORDEX-Southeast Asia region, International Journal of Climatology, 37, 1634-1647, https://doi.org/10.1002/joc.4803, 2017.

Ngo-Duc, T., Nguyen-Duy, T., Desmet, Q., Trinh-Tuan, L., Ramu, L., Cruz, F., Dado, J. M., Chung, J. X., Phan-Van, T., Pham-Thanh, H., Truong-Ba, K., Tangang, F. T., Juneng, L., Santisirisomboon, J., Srisawadwong, R., Permana, D., Linarka, U. A., and Gunawan, D.: Performance ranking of multiple CORDEX-SEA sensitivity experiments: towards an optimum choice of physical schemes for RegCM over Southeast Asia, Climate Dynamics, 62, 8659-8673, 10.1007/s00382-024-07353-5, 2024.

Nguyen, K. C., Katzfey, J. J., and McGregor, J. L.: Global 60 km simulations with CCAM: evaluation over the tropics, Climate dynamics, 39, 637-654, 2012.

Nguyen, P.-L., Bador, M., Alexander, L. V., and Lane, T. P.: Selecting regional climate models based on their skill could give more credible precipitation projections over the complex Southeast Asia region, Climate Dynamics, 61, 3431-3452, 10.1007/s00382-023-06751-5, 2023.

Nguyen, P.-L., Bador, M., Alexander, L. V., Lane, T. P., and Funk, C. C.: On the Robustness of Annual Daily Precipitation Maxima Estimates Over Monsoon Asia, Frontiers in Climate, 2, https://doi.org/10.3389/fclim.2020.578785, 2020.

Nguyen, P.-L., Bador, M., Alexander, L. V., Lane, T. P., and Ngo-Duc, T.: More intense daily precipitation in CORDEX-SEA regional climate models than their forcing global climate models over Southeast Asia, International Journal of Climatology, n/a, https://doi.org/10.1002/joc.7619, 2022.

Nguyen, P. L., Alexander, L. V., Thatcher, M. J., Truong, S. C. H., Isphording, R. N., and McGregor, J. L.: Selecting CMIP6 global climate models (GCMs) for Coordinated Regional Climate Downscaling Experiment (CORDEX) dynamical downscaling over Southeast Asia using a standardised benchmarking framework, Geosci. Model Dev., 17, 7285-7315, 10.5194/gmd-17-7285-2024, 2024.

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J. F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., and Sanderson, B. M.: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6, Geosci. Model Dev., 9, 3461-3482, 10.5194/gmd-9-3461-2016, 2016.

Pal, J. S., Small, E. E., and Eltahir, E. A. B.: Simulation of regional-scale water and energy budgets: Representation of subgrid cloud and precipitation processes within RegCM, J. Geophys. Res., 105, 29 579-529 594, 2000.

Rohde, R. A. and Hausfather, Z.: The Berkeley Earth Land/Ocean Temperature Record, Earth Syst. Sci. Data, 12, 3469-3479, 10.5194/essd-12-3469-2020, 2020.

Rotstayn, L. D.: A physically based scheme for the treatment of stratiform clouds and precipitation in large-scale models. I: Description and evaluation of the microphysical processes, Quarterly Journal of the Royal Meteorological Society, 123, 1227-1282, 1997.

Schroeter, B. J. E., Ng, B., Takbash, A., Rafter, T., and Thatcher, M.: A Comprehensive Evaluation of Mean and Extreme Climate for the Conformal Cubic Atmospheric Model (CCAM), Journal of Applied Meteorology and Climatology, 63, 997-1018, https://doi.org/10.1175/JAMC-D-24-0004.1, 2024.

Schwarzkopf, M. D. and Ramaswamy, V.: Radiative effects of CH4, N2O, halocarbons and the foreign-broadened H2O continuum: A GCM experiment, Journal of Geophysical Research: Atmospheres, 104, 9467-9488, 1999.

Sen, P. K.: Estimates of the regression coefficient based on Kendall's tau, Journal of the American statistical association, 63, 1379-1389, 1968.

Seneviratne, S. I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., Ghosh, S., Iskandar, I., Kossin, J., Lewis, S., Otto, F., Pinto, I., Satoh, M., Vicente-Serrano, S. M., Wehner, M., and Zhou, B.: Weather and Climate Extreme Events in a Changing Climate, 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, edited by: 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., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1513–1766, https://doi.org/10.1017/9781009157896.013, 2021.

Sobolowski, S., Somot, S., Fernandez, J., Evin, G., Brands, S., Maraun, D., Kotlarski, S., Jury, M., Benestad, R. E., Teichmann, C., Christensen, O. B., Bülow, K., Buonomo, E., Katragkou, E., Steger, C., Sørland, S., Nikulin, G., McSweeney, C., Dobler, A., Palmer, T., Wilcke, R., Boé, J., Brunner, L., Ribes, A., Qasmi, S., Nabat, P., Sevault, F., and Oudar, T.: GCM Selection and Ensemble Design: Best Practices and Recommendations from the EURO-CORDEX Community, Bulletin of the American Meteorological Society, 106, E1834-E1850, https://doi.org/10.1175/BAMS-D-23-0189.1, 2025.

Stevens, B., Fiedler, S., Kinne, S., Peters, K., Rast, S., Müsse, J., Smith, S. J., and Mauritsen, T.: MACv2-SP: A parameterization of anthropogenic aerosol optical properties and an associated Twomey effect for use in CMIP6, Geoscientific Model Development, 10, 433-452, 2017.

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., and Kumar, P.: Projected future changes in rainfall in Southeast Asia based on CORDEX–SEA multi-model simulations, Climate Dynamics, 55, 1247-1267, https://doi.org/10.1007/s00382-020-05322-2, 2020.

Thatcher, M. and McGregor, J. L.: Using a Scale-Selective Filter for Dynamical Downscaling with the Conformal Cubic Atmospheric Model, Monthly Weather Review, 137, 1742-1752, https://doi.org/10.1175/2008MWR2599.1, 2009.

Theil, H.: A rank-invariant method of linear and polynomial regression analysis, Indagationes mathematicae, 12, 173, 1950.

Tiedtke, M.: A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models, Monthly Weather Review, 117, 1779, https://doi.org/10.1175/1520-0493(1989)117<1779:Acmfsf>2.0.Co;2, 1989.

Truong, S. and Thatcher, M.: Evaluation of clouds in the conformal cubic atmospheric model using the CFMIP observation simulator package, International Journal of Climatology, 45, e8846, 2025.

Truong, S. C., Ramsay, H. A., Rafter, T., and Thatcher, M. J.: Simulation of an intense tropical cyclone in the conformal cubic atmospheric model and its sensitivity to horizontal resolution, Weather and Climate Extremes, 47, 100744, 2025.

Truong, S. C. H., Thatcher, M. J., Nguyen, P. L., Alexander, L. V., and McGregor, J. L.: Spectral Nudging Impacts on Precipitation Downscaling in the Conformal Cubic Atmospheric Model, version CCAM-2504: Insights from Summer 2011, EGUsphere, 2026, 1-31, 10.5194/egusphere-2025-5847, 2026.

Tucker, S. O., Kendon, E. J., Bellouin, N., Buonomo, E., Johnson, B., and Murphy, J. M.: Evaluation of a new 12 km regional perturbed parameter ensemble over Europe, Climate Dynamics, 58, 879-903, 2022.

Van Den Besselaar, E. J. M., Van Der Schrier, G., Cornes, R. C., Iqbal, A. S., and Klein Tank, A. M. G.: SA-OBS: A Daily Gridded Surface Temperature and Precipitation Dataset for Southeast Asia, Journal of Climate, 30, 5151-5165, https://doi.org/10.1175/jcli-d-16-0575.1, 2017.

Vautard, R., Kadygrov, N., Iles, C., Boberg, F., Buonomo, E., Bülow, K., Coppola, E., Corre, L., van Meijgaard, E., Nogherotto, R., Sandstad, M., Schwingshackl, C., Somot, S., Aalbers, E., Christensen, O. B., Ciarlo, J. M., Demory, M.-E., Giorgi, F., Jacob, D., Jones, R. G., Keuler, K., Kjellström, E., Lenderink, G., Levavasseur, G., Nikulin, G., Sillmann, J., Solidoro, C., Sørland, S. L., Steger, C., Teichmann, C., Warrach-Sagi, K., and Wulfmeyer, V.: Evaluation of the Large EURO-CORDEX Regional Climate Model Ensemble, Journal of Geophysical Research: Atmospheres, 126, e2019JD032344, https://doi.org/10.1029/2019JD032344, 2021.

Way, R. G., Oliva, F., and Viau, A. E.: Underestimated warming of northern Canada in the Berkeley Earth temperature product, International Journal of Climatology, 37, 1746-1757, https://doi.org/10.1002/joc.4808, 2017.

Wilson, D. R. and Ballard, S. P.: A microphysically based precipitation scheme for the UK Meteorological Office Unified Model, Quarterly Journal of the Royal Meteorological Society, 125, 1607-1636, 1999.

Yasutomi, N., Hamada, A., and Yatagai, A.: Development of a long-term daily gridded temperature dataset and its application to rain/snow discrimination of daily precipitation, Global Environmental Research, 15, 165-172, 2011.

Yatagai, A., Kamiguchi, K., Arakawa, O., Hamada, A., Yasutomi, N., and Kitoh, A.: APHRODITE: Constructing a Long-Term Daily Gridded Precipitation Dataset for Asia Based on a Dense Network of Rain Gauges, Bulletin of the American Meteorological Society, 93, 1401-1415, https://doi.org/10.1175/BAMS-D-11-00122.1, 2012.

Zeng, X., Zhao, M., and Dickinson, R. E.: Intercomparison of Bulk Aerodynamic Algorithms for the Computation of Sea Surface Fluxes Using TOGA COARE and TAO Data, Journal of Climate, 11, 2628-2644, https://doi.org/10.1175/1520-0442(1998)011<2628:IOBAAF>2.0.CO;2, 1998.

Ziese, M. R.-S., Armin; Becker, Andreas; Finger, Peter; Rustemeier, Elke; Hänsel, Stephanie; Schneider, Udo: GPCC Full Data Daily Version 2022 at 1.0°: Daily Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data., 10.5676/DWD_GPCC/FD_D_V2022_100, 2022.