https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

A multi-criteria framework for CORDEX-CORE2 GCM selection

Moetasim Ashfaq[1], Erika Coppola[2], Chris Lennard[3], Claas Teichmann[4], Deeksha Rastogi[1], Elias Massoud[1], Erasmo Buonomo[5], Eun-Soon Im[6], George Zittis[7], Jason P. Evans[8], Jesus Fernandez[9], Katherine J. Evans[1], Maria Leidinice da Silva[2], Marianna Adinolfi[10], Melissa Bukovsky[11], Rosmeri Porfirio da Rocha[12], Shabeh ul Hasson[13], Silvina A. Solman[14], Stefan Sobolowski[15], Sushant Das[16], Swen Brands[9], Tereza Cavazos[17], Thanh Ngo-Duc[18], Xuejie Gao[19]

[1]Oak Ridge National Laboratory, Oak Ridge, TN, USA

[2]The Abdus Salam International Centre for Theoretical Physics, Trieste, Italy

[3]Climate System Analysis Group, University of Cape Town, South Africa

[4]Climate Service Center Germany (GERICS), Helmholtz-Zentrum Hereon, 20095 Hamburg, Germany

[5]Met Office Hadley Centre, Exeter, UK

[6]Department of Civil and Environmental Engineering, Division of Environment and Sustainability, The Hong Kong University of Science and Technology, Hong Kong, China

[7]Climate and Atmosphere Research Center (CARE-C), The Cyprus Institute, Nicosia, Cyprus

[8]ARC Centre of Excellence for Weather of the 21st Century and Climate Change Research Centre, University of New South Wales, Sydney, Australia

[9]Instituto de Física de Cantabria (IFCA), CSIC-Universidad de Cantabria, Santander, Spain

[10]CMCC Foundation - Euro-Mediterranean Center on Climate Change, Lecce, Italy

[11]Haub School of Environment and Natural Resources, University of Wyoming, USA

[12]Departamento de Ciências Atmosféricas, Universidade de São Paulo, São Paulo, Brazil

[13]HAREME Lab, Earth and Society Research Hub, University of Hamburg, Hamburg, Germany

[14]Departamento de Ciencias de la Atmósfera y los Océanos, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires; Centro de Investigaciones del Mar y la Atmósfera, Consejo Nacional de Investigaciones Científicas y Técnicas, Buenos Aires, Argentina

[15]Geophyisical Institute, University of Bergen and the Bjerknes Center for Climate Research, Bergen, Norway

[16]Department of Earth and Atmospheric Sciences, National Institute of Technology Rourkela, India

[17]Center for Scientific Research and Higher Education of Ensenada (CICESE), Baja California, Mexico

[18]Department of Space and Earth Sciences, University of Science and Technology of Hanoi, Vietnam Academy of Science and Technology, Hanoi, Vietnam

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

[19]Nansen-Zhu International Research Centre, Climate Change Research Center, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

*Corresponding Author

Moetasim Ashfaq

Email Address: mashfaq@ornl.gov

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

Abstract

We present a structured multi-criteria framework for the sub-selection of CMIP6 global climate models (GCMs) to support CORDEX-CORE2 dynamical downscaling. The framework integrates five key criteria: historical performance, model independence, regional temperature sensitivity, precipitation spread, and data availability, and is designed to identify a single, consistent subset of GCMs across all CORDEX domains to improve the comparability and interpretability of regional projections. A total of 45 GCMs are evaluated over the historical period (1981–2014), with 31 models further assessed for projected changes over 2015–2100. Application of the framework shows that model performance is systematically higher for large-scale circulation and thermodynamic fields than for precipitation seasonality and monsoon-related processes, which remain a dominant source of uncertainty across regions. Despite the diversity of climates represented across CORDEX domains, model rankings are broadly consistent, with top-performing models exhibiting stable performance across both tropical and extratropical regions, while lower-ranked models show more pervasive deficiencies rather than region-specific weaknesses. Sensitivity analyses demonstrate that rankings are largely insensitive to the choice of aggregation method but depend strongly on the breadth of evaluation metrics, with robust and reproducible rankings emerging only when a large fraction of the full metric suite is retained. Assessment of model independence reveals substantial clustering within the ensemble, indicating that many models share similar performance characteristics, while a smaller subset provides distinct and complementary information. Regional temperature sensitivity exhibits a coherent ordering across domains, suggesting that differences in projected warming are primarily governed by intrinsic model characteristics rather than region-specific effects. In contrast, precipitation spread shows strong regional variability, with both the magnitude and temporal structure of precipitation change differing widely across models. The relationship between precipitation and warming further highlights that, in some regions, precipitation responses scale with temperature, while in others they are dominated by circulation variability. By combining these criteria with data availability constraints, the framework identifies a reduced set of models that retains key aspects of performance, diversity, and projected change. This approach provides a transparent and reproducible basis for GCM selection within CORDEX-CORE2 and offers a generalizable strategy for coordinated regional climate modeling efforts.

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

1. Introduction

The growing diversity and complexity of global climate models (GCMs) present both opportunities and challenges for regional climate analysis and impact studies (Collins et al., 2018). Multi-model ensembles, such as those produced under the Coupled Model Intercomparison Project Phase 6 (CMIP6), provide a broad sampling of plausible future climates (Khan et al., 2025); however, their utility in downstream applications is constrained by both practical and scientific considerations (Brands, 2022a). At regional scales, many GCMs struggle to reproduce key features of the observed climate due to coarse spatial resolution and structural model biases (Hasson et al., 2018; Iles et al., 2020; Desmet and Ngo-Duc, 2022; Holtanova et al., 2024; Rampal et al., 2024; Arias et al., 2025; Toolan et al., 2025). Regional climate downscaling offers a pathway to address these shortcomings and improve representation of regionally relevant processes (Prein et al., 2016; Coppola et al., 2021). However, the substantial computational demands make it impractical to employ the full ensemble of available GCMs for dynamical downscaling (McSweeney et al., 2013; Longmate et al., 2023). Consequently, regional modeling efforts rely on a subset of GCMs selected from the broader ensemble. Despite its central role in shaping regional projections, this GCM selection process remains largely ad hoc, lacks standards of practice, and is often guided by subjective or pragmatic considerations (Raju and Kumar, 2020). At the same time, growing scientific evidence indicates that projection uncertainty can be reduced by identifying and prioritizing models that credibly reproduce key climate feedbacks evident in the present and past, which are physically linked to future climate responses. This "emergent constraints" approach has brought increased attention to process-based and targeted model evaluation as a means of producing more robust climate projections and actionable climate information (Hall et al., 2019; Brient, 2020; Shiogama et al., 2022).

The need for model selection is particularly pressing in the context of the CORDEX (Coordinated Regional Climate Downscaling Experiment) CORE-2 (COmmon Regional Experiment - CORE, Phase 2) experiments, in which modeling centers worldwide will produce a new generation of high-resolution regional simulations, at 12.5 and/or 25 km horizontal resolution, using CMIP6 GCMs as boundary conditions. Because each modeling group can only afford to downscale a limited number of GCMs, model choice becomes a critical step in the workflow. However, no consensus currently exists on how to select these subsets. In practice, decisions have often been based on informal factors and subjective criteria such as data availability, institutional preferences and model compatibility, or qualitative judgments of performance in specific regions (Evans et al., 2014; McSweeney et al., 2015; Parding et al., 2020; Di Virgilio et al., 2022; Nguyen et al., 2024; Sobolowski et al., 2025).

In this study, we present a structured, multi-criteria framework for GCM sub-selection aimed at informing downscaling choices for CORDEX CORE-2. However, the approach is broadly applicable to any coordinated regional modeling effort where GCM ensemble sub-selection is required. The framework balances physical credibility, ensemble diversity, and practical feasibility by evaluating models across five criteria: historical performance, model independence, regional temperature and precipitation spread, and data availability (Sobolowski et al., 2025).

Historical performance is assessed using quantitative metrics that measure each model's ability to simulate observed regional climate worldwide. To ensure consistency, a core set of metrics is applied uniformly across all CORDEX domains, while additional region-specific metrics capture climatic features unique to individual regions. Model independence is considered through inter-model similarity metrics, which reduce the risk of overrepresenting structurally similar models

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

and help to ensure that the selected ensemble reflects a set of relatively unique models, addressing the often overlooked issue of model interdependence in multi-model ensembles (Knutti et al., 2013; Sanderson et al., 2017, Merrifield et al., 2023). Regional temperature projections are incorporated to preserve the range of warming responses within each CORDEX domain, maintaining the ability to explore plausible futures that reflect different levels of system response. Regional precipitation spread is retained to capture the diversity of hydrological projections within each domain, particularly in regions where uncertainty in rainfall constitutes a dominant contributor to climate risk. Finally, data availability is evaluated to confirm that each model provides the required variables and output frequency to support dynamical downscaling in practice. Together, these criteria help produce GCM subsets that combine scientific credibility, structural diversity, and practical feasibility, while capturing a representative range of plausible regional future climates. As such, this work represents a practical and more comprehensive implementation of the principles suggested by Sobolowski et al. (2025).

An additional design constraint in our framework is the requirement to identify a single, shared subset of GCMs that can be used across all CORDEX domains. This cross-domain consistency enhances the interpretability and intercomparison of CORDEX CORE-2 downscaling experiments by minimizing differences in ensemble composition across domains and ensuring that regional differences in projections are not confounded by variations in the driving model ensemble. A different set of GCMs downscaled for each domain makes it difficult to disentangle regional climate signals from model-specific effects, as variations in projected change may arise not only from genuine regional climate responses but also from structural differences among the selected models, including differences in climate sensitivity, feedback representation, and large-scale circulation biases. Such heterogeneity in driving ensembles complicates physically meaningful comparisons across domains and limits the robustness of coordinated impact assessments. By using the same GCMs across all domains, we ensure that inter-regional differences in downscaled projections reflect genuine climatic contrasts rather than inconsistencies in the driving models, while providing a consistent large-scale forcing framework for regional downscaling and cross-domain comparison.

This framework does not prescribe a universal set of "best" models; rather it offers a transparent, reproducible process for evaluating and selecting GCMs based on scientific and practical considerations. In addition to facilitating the sub-selection of CMIP6 GCMs to be downscaled within CORDEX CORE-2, it is designed to support modeling centers and climate service organizations across regions worldwide in identifying model subsets that are both credible and usable for regional downscaling and climate impact applications.

2. Data and Methods

The implementation of the GCMs evaluation framework requires different data periods for its key components. Model performance and independence are assessed using simulations from the historical period, while model climate sensitivity and precipitation spread are analyzed using projections from the future period. A total of 45 GCMs from the Coupled Model Intercomparison Project Phase 6 (CMIP6; Eyring et al., 2016) archive are evaluated for the historical period spanning 1981–2014 (see Table 1), with ERA5 reanalysis (Hersbach et al., 2020) serving as reference. The historical evaluation is based on metrics derived from monthly zonal and meridional winds, atmospheric specific humidity, sea level pressure, sea surface temperature, precipitation, and near-surface air temperature.

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

Future climate projections are analyzed under the SSP3-7.0 forcing scenario for the period 2015–2100 (O'Neil et al., 2016). SSP3-7.0 is the forcing scenario of choice for CORDEX-CORE2. These simulations are used to assess regional climate sensitivity, derived from projected changes in near-surface air temperature, and precipitation spread, which represents inter-model variability in projected changes in total precipitation. Among the 45 GCMs, only 31 provide the required variables for the SSP3-7.0 scenario (see Table 1); therefore, the assessment of the spread of future outcomes is restricted to this subset. All ERA5 and model outputs are regridded to a common 1° × 1° spatial grid to ensure consistency across datasets.

The analysis covers the nine CORDEX-CORE domains: Africa (AFR), Australasia (AUS), Central America (CAM), East Asia (EAS), Europe (EUR), North America (NAM), South America (SAM), South Asia (WAS), and Southeast Asia (SEA). Each CORDEX domain is subdivided into IPCC AR6 regions following the IPCC AR6 regional framework. Model evaluations are conducted within each AR6 region to ensure regionally consistent assessment. Figure 1 shows the geographical coverage of the nine CORDEX domains and the AR6 regions.

Model	Institution	Variant Label	Model References
ACCESS-CM2	CSIRO (Australia) and Bureau of Meteorology (Australia)	r1i1p1f1	Dix et al. (2019)
ACCESS-ESM1-5	CSIRO (Australia)	r1i1p1f1	Ziehn et al. (2020)
AWI-CM-1-1-MR	Alfred Wegener Institute (Germany)	r1i1p1f1	Semmler et al. (2020)
AWI-ESM-1-1-LR	Alfred Wegener Institute (Germany)	r1i1p1f1	Sidorenko et al. (2015)
BCC-CSM2-MR	Beijing Climate Center (China)	r1i1p1f1	Wu et al. (2019)
BCC-ESM1	Beijing Climate Center (China)	r1i1p1f1	Wu et al. (2020)
CAMS-CSM1-0	Chinese Academy of Meteorological Sciences (China)	r1i1p1f1	Rong et al. (2019)
CanESM5	Canadian Centre for Climate Modelling and Analysis (Canada)	r1i1p1f1	Swart et al. (2019)
CESM2	National Center for Atmospheric Research (NCAR, USA)	r1i1p1f1	Danabasoglu et al. (2020)
CESM2-FV2	NCAR (USA)	r1i1p1f1	Danabasoglu et al. (2020)
CESM2-WACCM	NCAR (USA)	r1i1p1f1	Danabasoglu et al. (2020)
CESM2-WACCM-FV2	NCAR (USA)	r1i1p1f1	Danabasoglu et al. (2020)
CIESM	Tsinghua University (China)	r1i1p1f1	Lin et al. (2020)
CMCC-CM2-SR5	Euro-Mediterranean Centre on Climate Change (Italy)	r1i1p1f1	Cherchi et al. (2019)
CNRM-CM6-1	CNRM, Météo-France / CNRS (France)	r1i1p1f2	Voldoire et al. (2019)
CNRM-CM6-1-HR	CNRM, Météo-France / CNRS (France)	r1i1p1f2	Voldoire et al. (2019)
CNRM-ESM2-1	CNRM, Météo-France / CNRS (France)	r1i1p1f2	Séférian et al. (2019)
E3SM-1-0	Department of Energy (USA)	r1i1p1f1	Golaz et al., (2019)
EC-Earth3	EC-Earth Consortium (Europe)	r1i1p1f1	Döscher et al. (2021)
EC-Earth3-Veg	EC-Earth Consortium (Europe)	r1i1p1f1	Döscher et al. (2022)
EC-Earth3-Veg-LR	EC-Earth Consortium (Europe)	r1i1p1f1	Döscher et al. (2022)
FGOALS-f3-L	Chinese Academy of Sciences (China)	r1i1p1f1	He et al. (2020)
FGOALS-g3	Chinese Academy of Sciences (China)	r1i1p1f1	Li et al. (2020)
GFDL-CM4	NOAA Geophysical Fluid Dynamics Laboratory (USA)	r1i1p1f1	Held et al. (2019)
GFDL-ESM4	NOAA Geophysical Fluid Dynamics Laboratory (USA)	r1i1p1f1	Dunne et al. (2020)
GISS-E2-1-G	NASA Goddard Institute for Space Studies (USA)	r1i1p1f1	Miller et al. (2021)
GISS-E2-1-H	NASA Goddard Institute for Space Studies (USA)	r1i1p1f1	Miller et al. (2021)
HadGEM3-GC31-LL	UK Met Office Hadley Centre (UK)	r1i1p1f3	Andrews et al. (2020)
HadGEM3-GC31-MM	UK Met Office Hadley Centre (UK)	r1i1p1f3	Andrews et al. (2020)
INM-CM4-8	Institute for Numerical Mathematics (Russia)	r1i1p1f1	Volodin et al. (2017)
INM-CM5-0	Institute for Numerical Mathematics (Russia)	r1i1p1f1	Volodin et al. (2017)
IPSL-CM6A-LR	Institut Pierre-Simon Laplace (France)	r1i1p1f1	Boucher et al. (2020)
KACE-1-0-G	National Institute of Meteorological Sciences (South Korea)	r1i1p1f1	Byun et al. (2019)
MIROC-ES2L	Japan Agency for Marine-Earth Science and Technology (Japan)	r1i1p1f1	Hajima et al. (2019)
MIROC6	Japan Agency for Marine-Earth Science and Technology (Japan)	r1i1p1f1	Tatebe et al. (2019)
MPI-ESM-1-2-HAM	Max Planck Institute for Meteorology (Germany)	r1i1p1f1	Neubauer et al. (2019)
MPI-ESM1-2-HR	Max Planck Institute for Meteorology (Germany)	r1i1p1f1	Gutjahr et al. (2019)

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

MPI-ESM1-2-LR	Max Planck Institute for Meteorology (Germany)	r1i1p1f1	Giorgetta et al. (2013)
MRI-ESM2-0	Meteorological Research Institute (Japan)	r1i1p1f1	Yukimoto et al. (2019)
NESM3	Nanjing University of Information Science and Technology (China)	r1i1p1f1	Cao et al. (2018)
NorCPM1	Norwegian Climate Centre (Norway)	r1i1p1f1	Bethke et al. (2019)
NorESM2-LM	Norwegian Climate Centre (Norway)	r1i1p1f1	Seland et al. (2020)
NorESM2-MM	Norwegian Climate Centre (Norway)	r1i1p1f1	Seland et al. (2020)
SAM0-UNICON	Seoul National University (South Korea)	r1i1p1f1	Park et al. (2019)
UKESM1-0-LL	UK Met Office Hadley Centre (UK)	r1i1p1f2	Mulcahy et al. (2023)

Table 1. The list of 45 CMIP6 models evaluated in this study. Models available for both historical (1981–2014) and future SSP3-7.0 (2015–2100) periods are shown in bold.

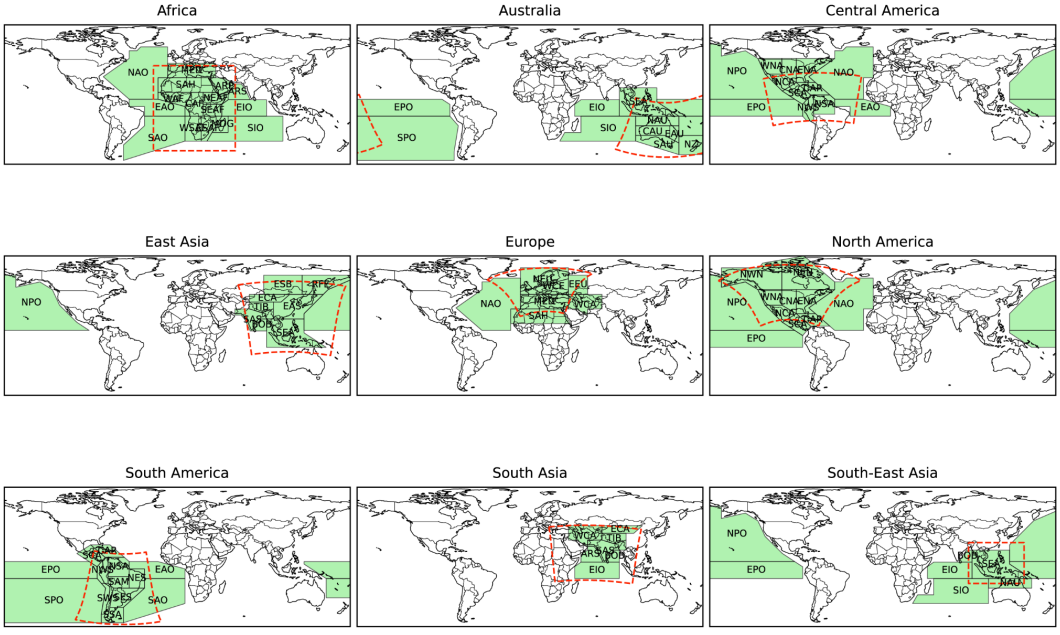

Figure 1. ***CORDEX domains and IPCC AR6 regions used in GCMs evaluation****. Red dashed lines indicate the boundaries of each CORDEX domain, while green polygons denote the AR6 continental and oceanic regions used to evaluate GCMs within each domain.*

2.1 Evaluation Metrics

The evaluation metrics used to assess GCMs across CORDEX domains in the historical period fall into two categories. The first set consists of common metrics that are applied uniformly across all CORDEX domains. These metrics represent fundamental climate characteristics of precipitation, temperature, winds, humidity, and sea surface temperature, and include measures of amplitude, seasonal means, annual variability, timing of peaks, and indices such as seasonality and relative entropy of precipitation.

The second set consists of region-specific metrics that capture unique dynamical and climatological processes important for individual CORDEX domains. Examples of these include the timing of monsoon onset and withdrawal, the position and strength of jet streams, vertical wind

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

shear in monsoon regions, the mean latitude of the Intertropical Convergence Zone (ITCZ), and the representation of circulation features such as the South Pacific High, the Aleutian Low, and the Azores High. Because these processes are only relevant for certain regions, a single skill score is computed for each metric and then distributed equally across all AR6 regions in the corresponding CORDEX domain to ensure fair weighting.

Across all nine CORDEX domains, a total of 98 evaluation metrics is used, combining both the common and region-specific categories (see Table 2). The number of metrics varies by domain depending on regional processes represented. For example, the African (AFR) domain includes 57 metrics, the largest number among all domains, reflecting its diverse climate regimes and the inclusion of diagnostics for West African, East African and Southeast African monsoons.

African Easterly Jet	HUS850 DJF	SAM MTG	TAS Peak Month
Atlantic ITCZ[1]	HUS850 JJA	SAM U_SHEAR	TAS Standard Deviation
AUS[2] Monsoon Demise	Indian Ocean Dipole	SAM V_SHEAR	TAS SON
AUS Monsoon Onset	Mascarene High	SAS[12] Monsoon Demise	TOS[16] DJF
AUS MTG[3]	Mozambique Channel Trough	SAS Monsoon Lows Index	TOS JJA
AUS U[4]_SHEAR	NAM[9] Monsoon Demise	SAS Monsoon Onset	TOS SON
AUS V[5]_SHEAR	NAM Monsoon Onset	SAS MTG	U200 DJF
Azores High	NAM MTG	SAS U_SHEAR	U200 JJA
Central America PR[6] Bimodality	NAM U_SHEAR	SAS V_SHEAR	U850 DJF
Caribbean Low-level Jet	NAM V_SHEAR	SEA[13] Monsoon Demise	U850 JJA
EAF[7] Low-level Jet	NAO[10]	SEA Monsoon Onset	V200 DJF
EAF Monsoon Demise	North Atlantic Polar Jet	SEA MTG	V200 JJA
EAF Monsoon Onset	PR Amplitude	SEA U_SHEAR	V850 DJF
EAF MTG	PR DJF	SEA V_SHEAR	V850 JJA
EAF U_SHEAR	PR Hovmöller	SEAF[14] Monsoon Demise	WAF[17] Monsoon Demise
EAF V_SHEAR	PR JJA	SEAF Monsoon Onset	WAF Monsoon Onset
EAS Monsoon Demise	PR MAM	SEAF MTG	WAF MTG
EAS[8] Monsoon Onset	PR Peak Month	SEAF U_SHEAR	WAF U_SHEAR
EAS MTG	PR Relative Entropy	SEAF V_SHEAR	WAF V_SHEAR
EAS Subtropical Westerly Jet	PR Standard D	South Atlantic High	Wang Index
EAS U_SHEAR	PR Seasonality	Subtropical Westerly Jet	Western Pacific Subtropical High
EAS V_SHEAR	PR SON	TAS[15] Amplitude	ZG[18]500 DJF
EAS Winter Monsoon Index	Southern Annular Mode	TAS DJF	ZG500 JJA
ENSO Niño-3.4	SAM[11] Monsoon Demise	TAS JJA	
Guo Index	SAM Monsoon Onset	TAS MAM	

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

ITCZ[1] = Intertropical Convergence Zone, AUS[2] = Australasian, MTG[3] = Meridional Temperature Gradient, U[4]=Zonal Winds, V[5] = Meridional Winds, PR[6] = Precipitation, EAF[7] = East African, EAS[8] = East Asian, NAM[9] = North American, NAO[10] = North Atlantic Oscillation, SAM[11] = South American, SAS[12] = South Asian, SEA[13]= Southeast Asian, SEAF[14] = Southeast African, TAS[15] = Surface Air Temperature, TOS[16] = Sea Surface Temperature, WAF[17] = West African, ZG[18] = Geopotential Height

Table 2. The list of 98 metrics used in GCMs evaluation

Although the evaluation suite is extensive, potential concerns about over-complexity are addressed in two ways. First, pairwise dependence among metrics is quantified within each AR6 region to identify redundancy and assign weights according to each metric's uniqueness (explained later in *Metrics Weighting Subsection*). Second, the sensitivity of model rankings to the number of included metrics is examined using a Monte Carlo convergence experiment. For each domain, subsets of metrics are progressively expanded to the full suite, repeatedly sampling orders (weighted by uniqueness) and tracking how model rankings deviate from the reference ranking obtained with all metrics. This analysis, described in detail in the *Results* section, illustrates how rapidly or slowly rankings stabilize as additional metrics are included, clarifying the practical benefit of using a broad metric set after redundancy has been accounted for.

2.2 Skill Quantification

Model skill for each metric is evaluated using up to three complementary statistical measures that capture different aspects of model performance. These include spatially centered pattern correlation (PC), which quantifies spatial agreement between simulated and observed fields; normalized bias, which measures systematic offsets in magnitude relative to observed variability; and normalized spatially centered root mean square error (RMSE), which estimates total spatial deviation relative to observed spatial variability. For spatial RMSE, normalization is based on the spatial standard deviation of the reference dataset, whereas for bias, normalization is based on the temporal standard deviation unless otherwise specified.

In most cases, all three measures are applied to provide a comprehensive evaluation of model skill. However, for some metrics, particularly those based on climatological values or single time series, only a subset of measures is used due to limitations in the dimensionality or structure of the underlying data. Metric-specific details, including which measures are applied, are provided in the Appendix.

Each statistical measure is transformed to a 0–100 scale. For this purpose, we introduce penalty-based exponential functions:

$$\text{Pattern correlation: } PC_{(0-100)} = max\,(0, 100 \times PC) \quad (1)$$

$$\text{Normalized bias: } Bias_{(0-100)} = 100 \times exp\,(-|Bias| \,/\, N) \quad (2)$$

$$\text{Normalized RMSE: } RMSE_{(0-100)} = 100 \times exp\,(-RMSE \,/\, N) \quad (3)$$

Pattern correlation is scaled using a linear function because it is already a bounded, dimensionless measure of spatial similarity (−1 to 1). Its values directly reflect proportional differences in pattern agreement, so a linear transformation preserves this interpretation while simply rescaling it to 0–100. Negative correlations are set to zero because they indicate an opposite spatial pattern and are treated as no skill rather than negative skill. An exponential penalty is unnecessary for PC and would artificially compress differences among high-correlation values, particularly near 1 where most well-performing models cluster. In contrast, bias and RMSE are unbounded error measures

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

in physical units. They are normalized by the reference standard deviation to express errors relative to natural variability, and an exponential function is applied to penalize large deviations more strongly while maintaining a smooth decline in skill.

Here, N determines the rate of decay and is set to 2 for most metrics, corresponding to twice the reference standard deviation used in the normalization. Because the errors are already expressed in normalized variability units, dividing by 2 effectively represents a threshold of two standard deviations in physical terms. This choice moderates the rate at which the penalty increases with error, avoiding overly harsh penalization of moderate deviations while still strongly down-weighting large departures. For example, a bias equal to twice the standard deviation results in a score of approximately 37. Under a Gaussian assumption, about 95 percent of values lie within ± two standard deviations of the mean, so a bias or RMSE exceeding this level represents a deviation larger than what is typically observed in the reference climate. This is noted only to illustrate the scale of the threshold. Even for variables that are not strictly Gaussian, such as precipitation, the two standard deviations level provides a practical, variability-based benchmark for identifying substantial deviations. For a few metrics with a low temporal coefficient of variation, bias normalization is based on the range rather than the standard deviation, and the division by N is omitted (see Appendix for details).

When multiple statistical measures are available for a given metric, their component scores (Eq. 1-3) are averaged to obtain a composite skill score (S) ranging from 0 to 100.

2.3 Metrics Weighting

To minimize redundancy among metrics and ensure that correlated measures do not disproportionately influence model evaluation, each metric is assigned a *uniqueness weight* based on its independence from other metrics. The weighting approach follows the method developed by Ashfaq et al. (2022).

Pairwise Pearson correlations are first computed among all metrics within each AR6 domain, and these correlations are converted into a correlation-based distance measure. Smaller distances indicate stronger relationships between metrics. A *similarity score* (SS) is then calculated for each metric pair as:

$$SS_{ij} = exp\left[-\frac{\left(1 - \left|C_{ij}\right|\right)}{D_x}\right] \quad (4)$$

where C_{ij} is the Pearson linear correlation between metrics i and j, $\left(1 - \left|C_{ij}\right|\right)$ is the correlation-based distance, and D_x is similarity radius. For this study, D_x is set as 0.2. This threshold focuses on high inter-metric relationships while allowing for the moderate correlations expected in interconnected climate variables. For each metric i, the *effective redundancy* (ER_i) is defined as the sum of its similarity scores with all other metrics:

$$ER_i = 1 + \sum_{j \neq i} SS_{ij} \quad (5)$$

The *uniqueness weight* (W_i) is then the inverse of this redundancy:

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

$$W_i = \frac{1}{ER_i} \quad (6)$$

Metrics that are largely independent from others receive higher weights, while those showing strong interdependence are proportionally down weighted. Weighted skill scores are subsequently averaged across all metrics to obtain regional and domain-level performance scores.

Figure 2 illustrates the weighting framework, showing how metric correlations, similarity scores, and resulting weights are related. The top panel (Fig. 2a) presents an example from one AR6 region in South America. The left triangle shows pairwise Pearson correlations among all metrics, while the right triangle shows the corresponding similarity scores. The line plot above displays the resulting weights of individual metrics within that AR6 region. For instance, high correlations among models' performances in capturing seasonal magnitudes of temperature, precipitation and humidity in the lower troposphere results in high similarity scores and reduces their weights to less than one third of their actual performance scores. We have provided the individual illustrations for all AR6 regions across CORDEX domains in the *Supplementary Material* (Figs. S1). The *Supplementary Material* is archived in a public repository (Ashfaq et al., 2026; https://doi.org/10.5281/zenodo.19556090). Figure 2b summarizes the global distribution of pairwise Pearson correlations, similarity scores and weights across all AR6 regions in nine CORDEX domains. Pairwise correlations range from −0.73 to 0.98, indicating that metric relationships span from moderately negative to very strongly positive. In this framework, redundancy is determined by the strength of the linear relationship between two metrics. Since the similarity score is calculated from the correlation distance $\left(1 - \left|C_{ij}\right|\right)$, both strong positive and strong negative correlations reduce this distance and therefore produce high similarity values. The similarity score is scaled by the prescribed radius $D_x = 0.2$. Because D_x is small, only correlations with large magnitudes produce sufficiently small distances to generate substantial similarity values.

Moderate and weak correlations fall outside this radius, causing their similarity scores to decay rapidly toward zero and limiting their contribution to redundancy (*Eq. 5)*. The resulting similarity scores range from 0 to 0.99, but most metric pairs yield values close to zero, showing that only a limited subset of strongly related pairs contribute meaningfully to redundancy. The corresponding raw pairwise weights span from 0.50 to 1.00. Weights near 1 dominate, reflecting that most metric pairs provide largely unique information. The lower bound of 0.5 occurs when two metrics are perfectly correlated in magnitude, meaning they contain the same information. Even in this case, each metric retains half of its contribution, so no single pairwise relationship can completely suppress a metric's influence. At the aggregate level, a metric's final uniqueness weight is obtained by summing its similarity scores across all its pairs to estimate effective redundancy (*Eq. 5*))and then taking the inverse of that total (i.e., *Eq 6*), so substantial down-weighting arises only when strong relationships persist across multiple pairings, indicating systematic redundancy within the metric set.

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

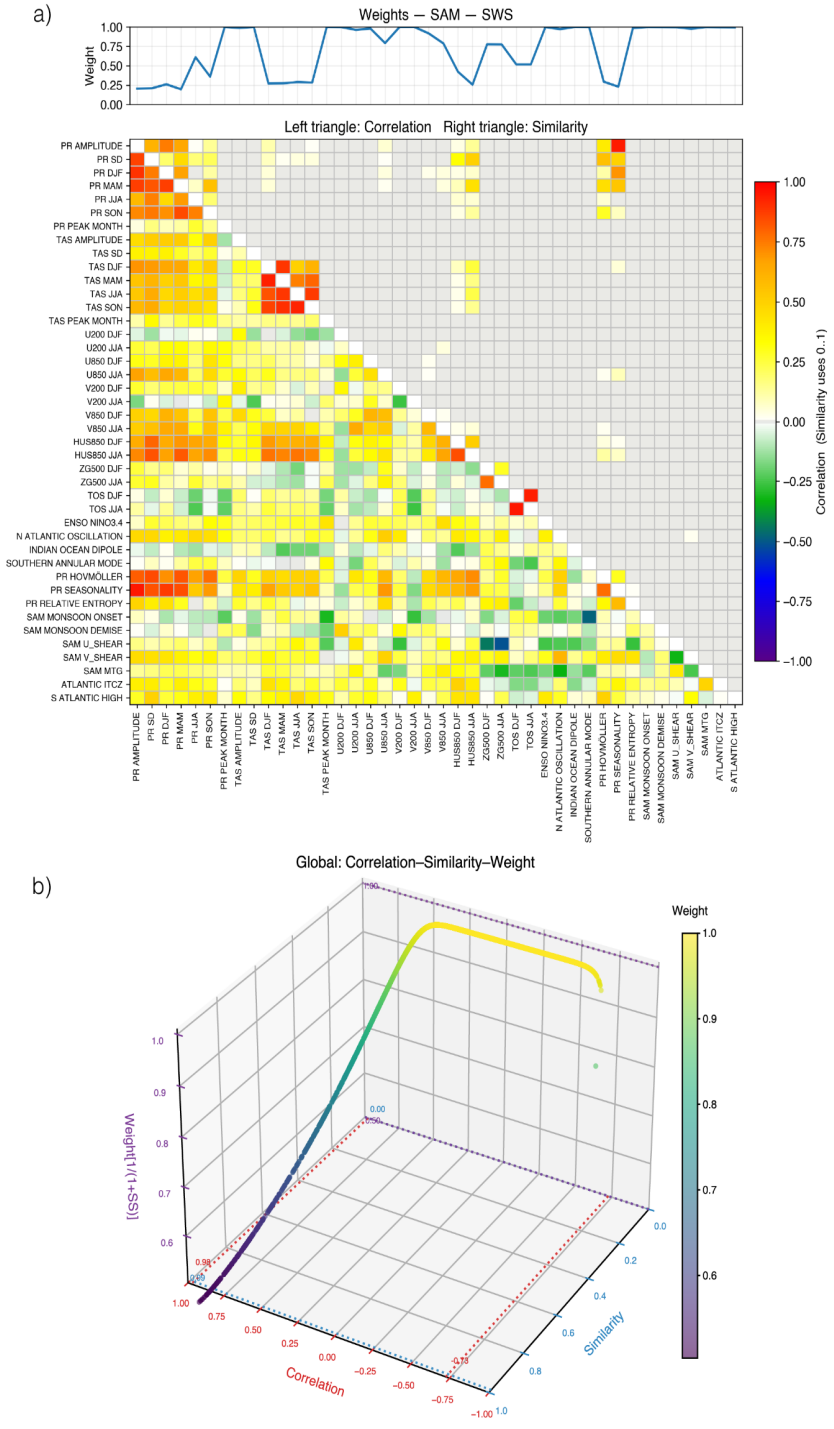

Figure 2 ***Illustration of the metrics weighting methodology****. (a) Pairwise correlations between metrics (bottom triangle) and their corresponding similarity scores (top triangle) for the SWS AR6 region within SAM domain, with the line plot above showing the overall weight of each metric. (b) Three-dimensional scatter plot showing the relationship between correlation, similarity, and pairwise weight for all metric pairs across all AR6 regions in nine CORDEX domains. The dotted colored lines with labels on each axis represent the global minimum and maximum of each measure. Colorbar indicates the weight magnitude.*

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

2.4 Models Independence

Model independence is evaluated to quantify the structural diversity of the GCM ensemble. Independence reflects how distinct models are in their performance across metrics and regions: high independence indicates distinct skill patterns that likely arise from differences in process representation or model structure, whereas low independence suggests redundancy due to shared components, parameterizations, or model lineage. Balancing model skill and independence helps avoid overrepresentation of structurally similar models as input for regional downscaling.

Several approaches have been proposed to quantify model independence, including methods based on model genealogy, structural similarity, and performance-based distance metrics (e.g., Knutti et al., 2013; Sanderson et al., 2015a,b; Abramowitz et al., 2019; Brunner et al., 2019; Brands 2022b). Here, we adopt the performance-based framework of Ashfaq et al. (2022), in which independence is evaluated using similarity in model skill patterns.

Independence is quantified using pairwise cosine similarity computed from skill scores across all metrics and AR6 regions within the nine CORDEX domains. For each model, a multidimensional vector is created by combining its scores. The cosine similarity between two models, A and B, is defined as:

$$CosSim(A,B) = \frac{\sum_{k=1}^{n} S_{A,k} S_{B,k}}{\sqrt{\sum_{k=1}^{n} S_{A,k}^2} \sqrt{\sum_{k=1}^{n} S_{B,k}^2}} \quad (7)$$

where $S_{A,k}$ and $S_{B,k}$ are the metric skill values and n is the total number of metrics. A cosine similarity of 1 indicates identical skill patterns across all regions and metrics, 0 indicates no linear alignment. Because all skill scores are defined on a 0–100 scale and are therefore non-negative, cosine similarity values in this framework are bounded between 0 and 1. Independence is represented as cosine distance, $Dcos = 1 - CosSim$.

Since cosine similarity measures the alignment of full skill vectors, very high values indicate that two models exhibit nearly proportional performance across all metrics and AR6 regions. Such strong similarity is unlikely to arise from incidental agreement and more plausibly reflects closely related structural characteristics or shared model development pathways, although the metric itself is based solely on evaluated performance. Pairwise distances are computed for all model pairs to form an $M \times M$ matrix, where higher *Dcos* values denote greater dissimilarity and thus higher independence. A high *CosSim* threshold is used to identify pairs of models exhibiting very strong similarity and potential redundancy. The rationale for the specific threshold choice is discussed in the Results section.

2.5 Global Aggregation and Ranking of Models

To identify models that perform consistently well across all CORDEX domains, a form of global aggregation of metric scores is required. However, the choice of aggregation strategy could potentially influence the resulting rankings. To assess this sensitivity, four approaches were tested. These approaches represent a subset of possible aggregation strategies and are designed to retain all models in the ranking process without imposing additional exclusion or penalty criteria based on performance in a single domain.

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

Method 1 averages all metric skill scores across all CORDEX domains and AR6 regions to obtain a single global score for each model. This approach is simple and transparent but implicitly overweights domains or regions with a larger number of available metrics. Method 2 first aggregates metric skill scores at the AR6-regional level by averaging all metrics for each model within each AR6 region and then ranks models within each region based on these mean scores. The resulting AR6-regional ranks are subsequently averaged to obtain a global rank. This approach ensures equal weighting of AR6 regions but collapses metric-level information into a single regional score prior to ranking. Method 3, the preferred approach, aggregates metric skill scores within each CORDEX domain to form domain-level mean scores, ranks models within each domain, and then averages these domain ranks globally. By preserving CORDEX domains as an explicit intermediate aggregation level, this approach provides a balanced and interpretable framework that is consistent with CORDEX's spatial structure. Method 4 distinguishes between metrics that are common across all CORDEX domains and metrics that are defined only for specific domains. Common metrics and domain-specific metrics are combined at the domain level prior to ranking, with domain-specific metrics contributing individually to the domain-level score. Domain-level ranks are then averaged globally, increasing the influence of metrics that are unique to individual CORDEX domains relative to the other approaches.

3. Results

3.1 Unweighted Skill Scores

Africa is selected in Fig. 3 as an illustrative example to describe the domain-level evaluation in detail, while results for the remaining eight CORDEX regions are provided in the *Supplementary Material* (Figs. S2) and summarized collectively in the radial dashboard (Fig. 3b). Models show moderate overall performance skill over Africa, with median values near 70 and interquartile ranges (IQRs) around 15–20, indicating reasonable model agreement across metrics and AR6 regions. Circulation- and thermodynamics-related metrics, such as upper-tropospheric winds, humidity, and large-scale pressure systems, are generally well captured, reflecting the models' ability to reproduce the broad dynamical structure of the African climate. In contrast, precipitation- and monsoon-timing metrics (for example, seasonal rainfall peaks, monsoon onset and demise) exhibit weaker performance and larger inter-model spread, underscoring persistent uncertainty in simulating regional hydrological processes and seasonal transitions. Overall, the models demonstrate robust skill for large-scale circulation but limited consistency for metrics tied to monsoon variability and precipitation seasonality.

When extended across all CORDEX domains (Fig. 4), the radial dashboard reveals that these patterns are broadly consistent worldwide with metrics related to regional hydrological cycles, such as precipitation peak month, monsoon onset and demise, and precipitation amplitude, are among the worst performers. The dashboard provides a compact comparison of model performance across regions by summarizing the distribution of metric-level skill within each domain and highlighting the subset of metrics with the lowest mean skill. Differences among domains are expressed both through the position and width of the IQR. The IQR is represented by the radial thickness of the colored annulus for each domain. The inner and outer edges of this band correspond to the 25th and 75th percentiles of mean model skill across all metrics in that domain, respectively, while the median is shown as a solid line within the band. A narrow annulus indicates that most metrics have similar skill values, reflecting relatively uniform model performance across processes, whereas a wider annulus signifies greater variability among metrics. Because the radial distance from the center corresponds to increasing skill, outward-shifted IQR bands denote higher overall performance, while inward-shifted bands indicate lower typical skill.

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

Temperate and subtropical CORDEX domains, including EUR, EAS, AUS, display higher median model skill (>73) and narrower IQRs (13–17), indicating relatively stronger and more uniform performance across metrics. Conversely, tropical regions such as WAS, SEA, CAM, and SAM exhibit comparable median skill but substantially wider spreads (about 17–23), reflecting greater metric-to-metric variability and higher model uncertainty. NAM stands out with the lowest and most uniform skill (median ≈ 61, IQR ≈ 11), suggesting generally weaker yet more homogeneous model behavior.

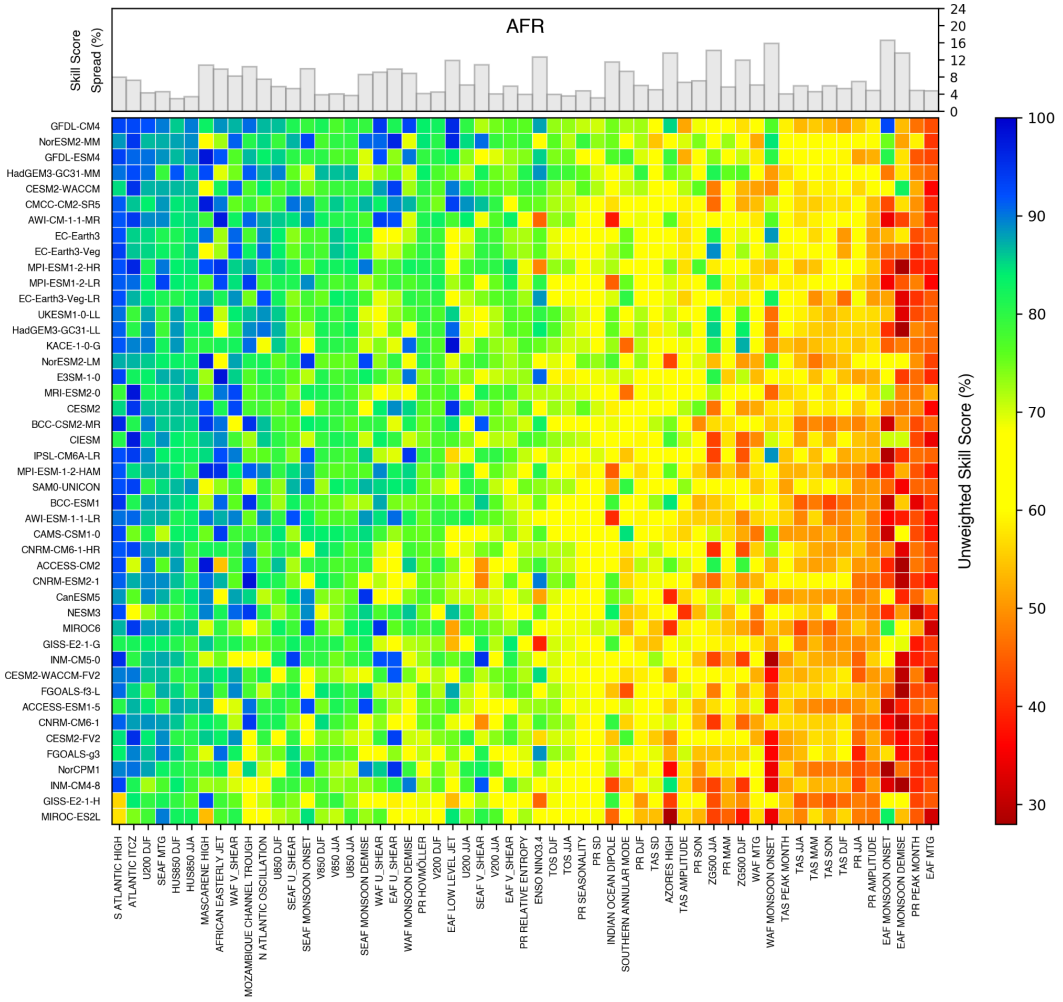

Figure 3 ***Climate models unweighted metrics skill score****. Individual metrics skill scores (heatmap) and its spread across models (bars) over Africa (AFR) domain. Models are ordered from top to bottom according to overall skill, and metrics are arranged from left to right based on average performance across models.*

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

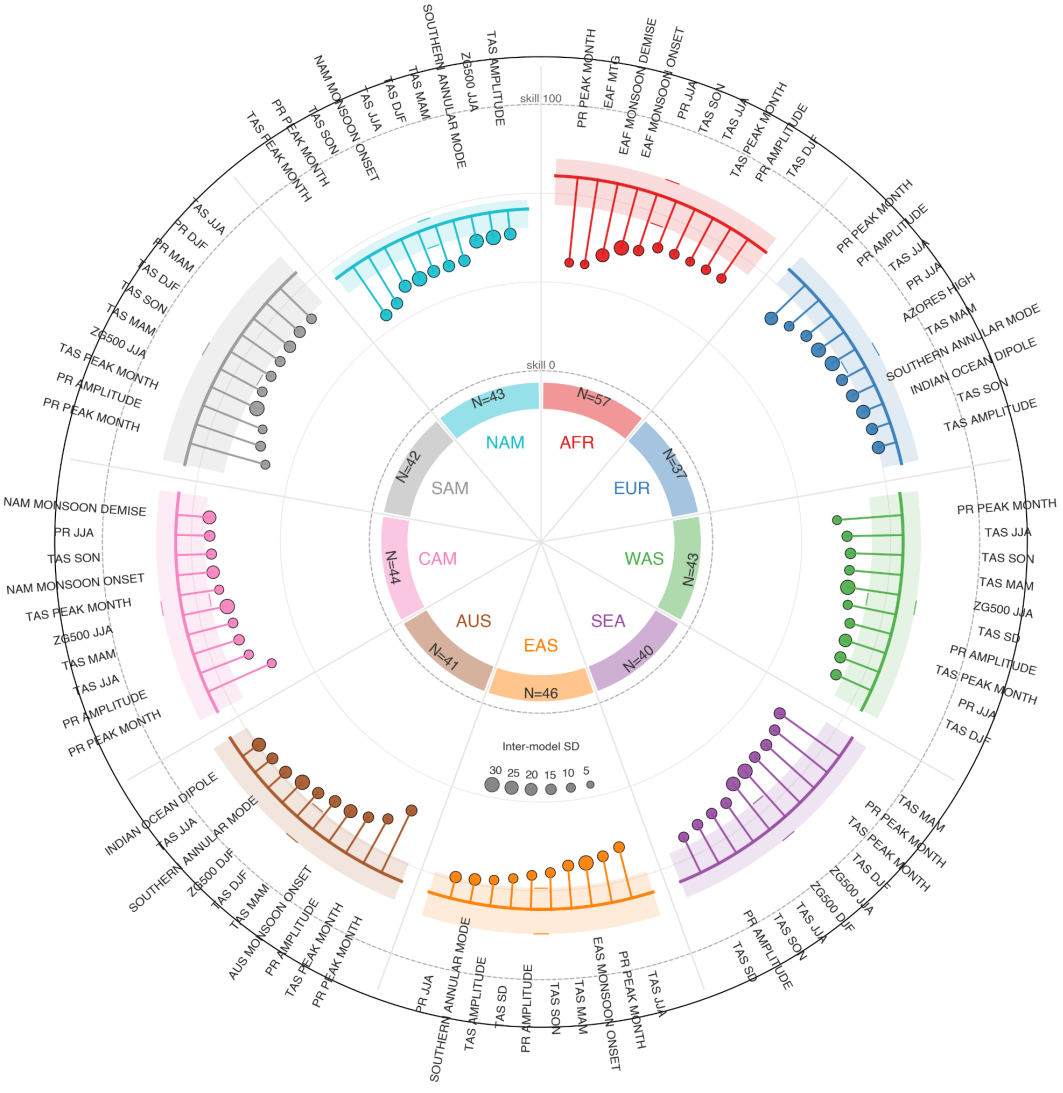

Figure 4 ***Climate models unweighted metrics skill score****. Summary statistics of climate model skill across CORDEX domains for unweighted metrics. The colored annulus shows the interquartile range of mean model skill across all metrics, with the median marked as a solid line. Outer radial spokes highlight the ten metrics with the lowest mean skill. The dot position denotes the mean model skill for that metric, the spoke length represents its deviation from the domain median (|mean−median|), and the dot size indicates inter-model standard deviation (spread in skill). Reference circles and N correspond to skill values of 0 and 100, and the number of evaluated metrics, respectively.*

3.2 Metrics Weights

To ensure that correlated metrics do not bias the multi-metric assessment of model skill, a correlation-based weighting scheme was applied (See *Methods*). Metric covariance was quantified by correlating the skill scores of the 45 GCMs for every pair of metrics within each CORDEX domain; metrics showing similar model-to-model performance patterns were considered interrelated and were therefore down-weighted. The resulting weights vary across domains (Fig. 5), from mean values of about 0.83 in SAM and 0.86 in CAM to approximately 0.90 in AFR and

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

SAS, indicating stronger inter-metric correlations in the American domains and greater metric independence in AFR, SAS, and EAS. Metrics representing large-scale circulation and dynamical features, including jet streams, vertical wind shear, monsoon onset and demise, and pressure systems such as the Mascarene (in the southern Indian Ocean) and Western Pacific highs, retain high weights (≈0.98–1.00) across regions, reflecting their distinct and non-redundant information content. In contrast, precipitation- and temperature-based metrics, particularly those describing the seasonal cycle (e.g., precipitation amplitude, seasonality index, and relative entropy), are consistently down-weighted (≈0.4–0.8) due to their strong inter-correlations among the 45 GCMs. This pattern is most pronounced in SAM, CAM, and EUR, whereas tropical domains such as AFR, SAS, and SEA show higher effective weighting across metrics, indicating greater diagnostic diversity among the evaluated variables.

3.3 Models Ranking

The weighted metrics are used to distinguish between better-performing and worse-performing models. Models are ranked both within each CORDEX domain and globally across all nine domains. Regional rankings are derived by averaging all weighted metrics across the AR6 regions within a given CORDEX domain. As noted earlier, unique metrics contribute equally to each AR6 region to ensure unbiased weighting. By default, the global ranking is obtained by averaging ranks across the nine domains; however, we also assess the sensitivity of global rankings to the aggregation method used. In addition to the default method (Method 3 in Fig. 6, which first computes regional rankings within each CORDEX domain and then averages these regional ranks to obtain the global ranking), three alternative aggregation approaches are tested (described in the *Methods* section), and their results are illustrated in Fig. 5. Rankings remain largely consistent across methods, indicating that the choice of aggregation scheme has minimal influence on the overall conclusions. However, Method 4, which places greater emphasis on metrics unique to individual CORDEX domains, produces noticeably larger deviations for a subset of models. These differences highlight the sensitivity of some models to domain-specific performance characteristics. The relative insensitivity of model rankings to the choice of aggregation method otherwise reflects the general consistency of their performance across domains.

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

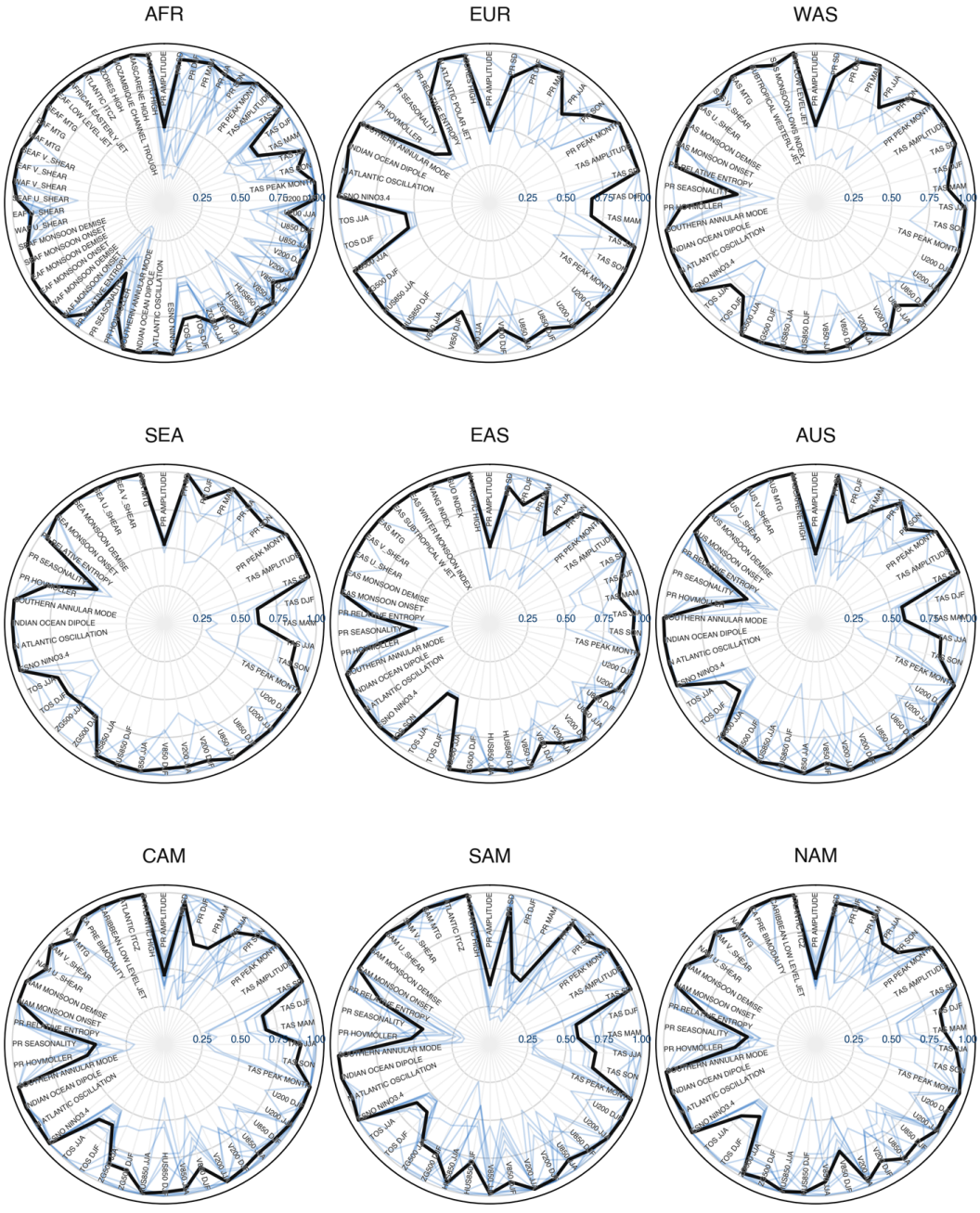

Figure 5 ***Metric weights for the nine CORDEX domains****. Each panel shows weights assigned to all metrics within a domain; grey concentric rings represent the weight values; thin colored lines represent individual AR6 regions and the thick black line denotes the median weight across regions.*

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

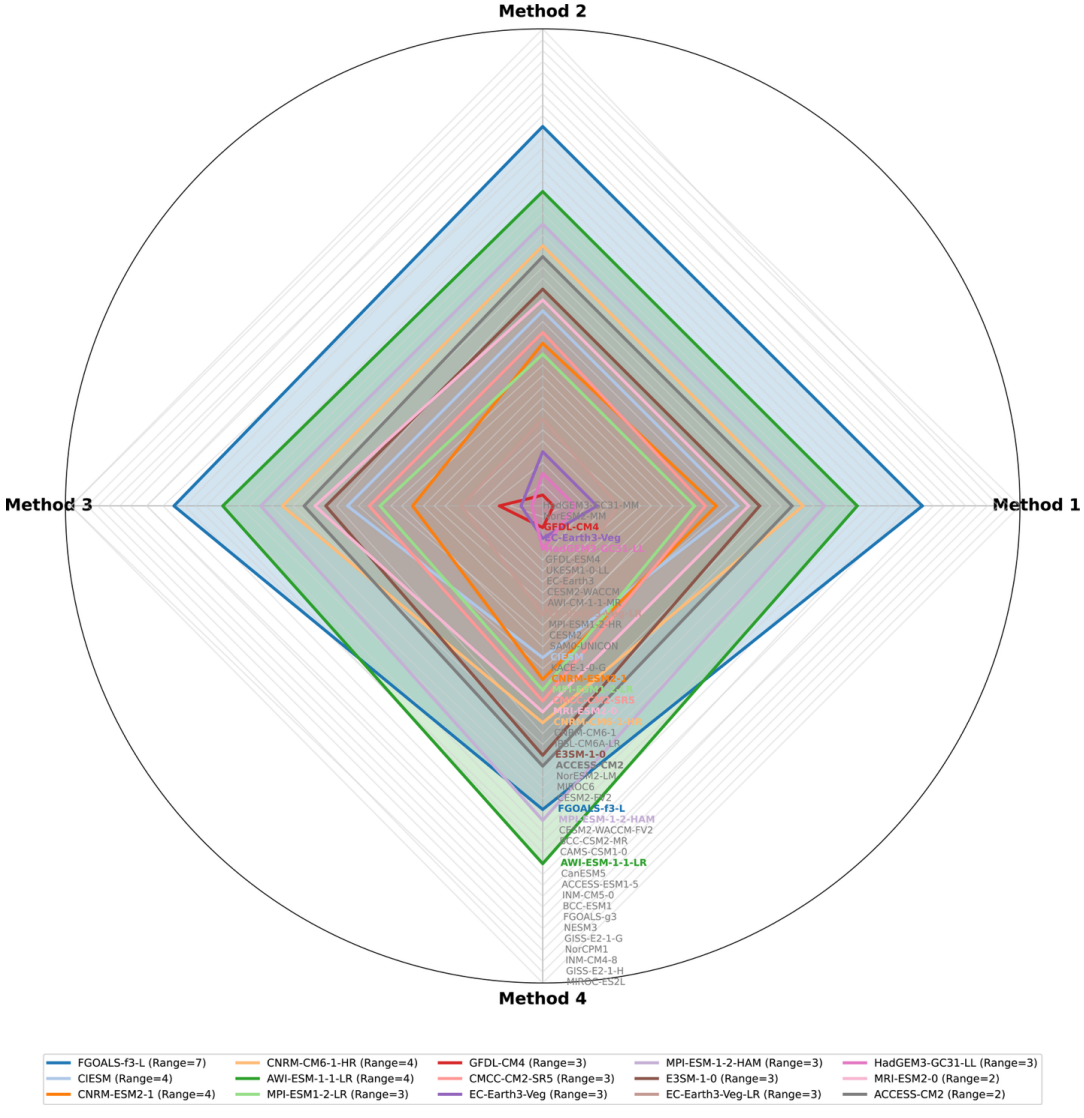

Figure 6. ***Comparison of global model rankings obtained using different aggregation methods****. The top fifteen models exhibiting the most noticeable changes in rank across methods are highlighted in colors other than gray and shown in the list at the bottom.*

We present the outcome of weighted model ranking using a heatmap of regional and global ranks to illustrate the degree of consistency in climate model performance across CORDEX domains (Fig. 7). Several key highlights emerge from this analysis, illustrated by the following examples. Models such as HadGEM3-GC31-MM, GFDL-ESM4 and EC-Earth3-Veg rank high across all domains, with relatively small inter-domain rank spread, indicating robust and geographically transferable performance across both tropical and extratropical regions. NorESM2-MM also shows generally strong performance but with noticeable degradation in EUR and AUS, reflecting moderate regional dependence. In contrast, several models exhibit pronounced inconsistency. CMCC-CM2-SR5 and CNRM-CM6-1 perform reasonably well in selected domains but rank poorly in others, particularly AFR, resulting in substantially lower global ranks despite strengths in some domains. The MPI family (MPI-ESM1-2-HR and MPI-ESM1-2-LR) displays relatively

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

stable mid-range performance, with stronger ranks in EUR and NAM and weaker performance in monsoon-dominated domains, while UKESM1-0-LL performs better in tropical regions than in midlatitudes, especially EUR and NAM, which limits its global ranking. At the lower end, models such as INM-CM4-8 and MIROC-ES2L rank consistently poorly across all domains, indicating systematic deficiencies rather than region-specific weaknesses. Overall, the global ranking is largely determined by cross-domain consistency: models that maintain solid performance across all domains rank higher than those that perform very well in only a few regions but poorly in others.

3.4 Need for Large Suite of Evaluation Metrics

To assess the robustness of model rankings to the number of metrics used, we perform a Monte Carlo convergence analysis (Fig. 7). The goal is to determine how large a fraction of the available metrics is needed for rankings to remain consistent with those obtained from the full suite. For each CORDEX domain, we generate up to 5,000 random realizations, each representing a unique random order in which metrics are added. In the initial steps, when only a few metrics are available, the total number of possible combinations is smaller than 5,000, so fewer realizations contribute to the statistics. Once the number of metrics exceeds twelve, the number of possible subsets becomes larger than 5,000, and from that point onward the full ensemble of 5,000 realizations is used for all subsequent steps. For reference, with 13 metrics, there are 8,191 possible subsets, and with 40 metrics, the number exceeds one trillion ($\approx 1.1\times10^{12}$), illustrating how rapidly the combination space expands as metrics are added. Metrics are sampled once per realization (without repetition), with selection probabilities weighted by their influence, defined as the product of the metric's assigned weight and the inter-model standard deviation of scores. This weighting favors metrics that both carry greater importance and better discriminate among models. A fixed random seed ensures reproducibility.

Each realization begins with a single metric, after which metrics are added sequentially according to the randomized order until the full set is included. After each addition, model scores are recomputed, domain ranks are recalculated, and a global rank is derived as the mean of the domain-specific ranks. The absolute deviation of this global rank from the reference global ranking (computed using all metrics) is then recorded. This procedure tracks how the global ranking converges as progressively more metrics are incorporated.

The results are summarized using probability density functions of global rank deviation as a function of the fraction of metrics included. Rankings derived from small metric subsets exhibit large variability across realizations, indicating poor robustness. Convergence improves steadily as additional metrics are included, with the median global rank deviation falling below ±2 ranks once approximately 75% of the metrics are used, and below ±1 rank only after more than 90% are included. These results demonstrate that a large fraction of the full metric suite is required to obtain a stable and reproducible global ranking, and that rankings based on limited metric subsets are highly sensitive to metric selection, highlighting the importance of comprehensive evaluation approaches in climate model assessment (Sa'adi et al., 2026).

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

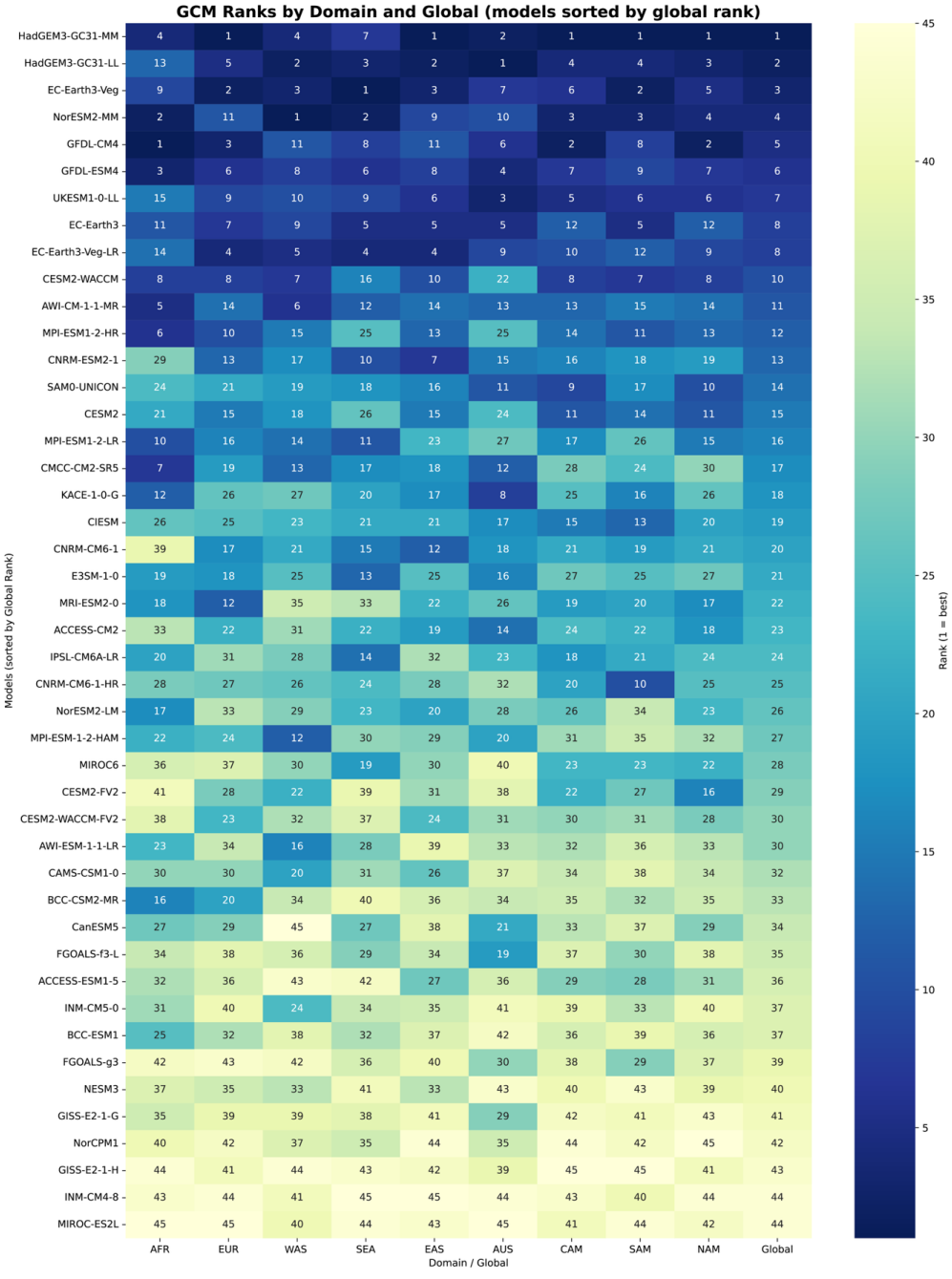

Figure 7 ***Global and regional ranks of models****. Heatmap of regional and global ranks for 45 CMIP6 models across nine CORDEX domains.*

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

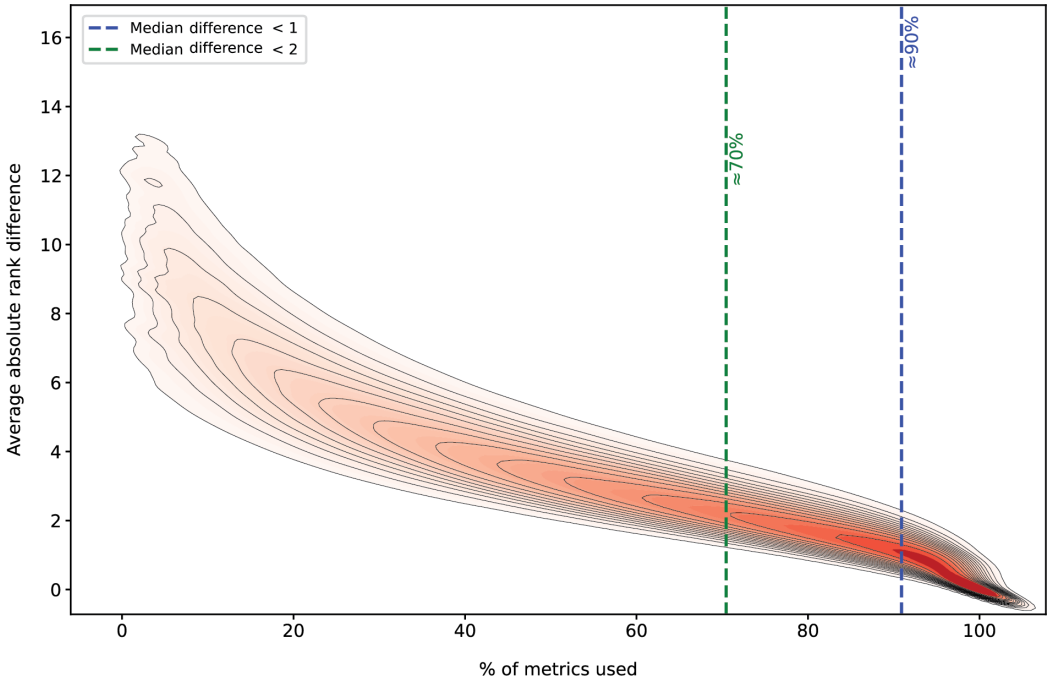

Figure 8. ***Sensitivity of ranking to the number of evaluation metrics****. Probability density of global rank deviation as a function of the fraction of metrics included, based on Monte Carlo convergence analysis.*

Pairwise cosine similarity across the GCM ensemble is generally high, with values ranging from approximately 0.68 to 0.99 and a mean near 0.85, but this overall similarity masks clear structure within the ensemble (Fig. 8). Several closely related model configurations exhibit extremely high similarity, for example HadGEM3-GC31-LL / HadGEM3-GC31-MM (≈0.96) and CESM2 / CESM2-WACCM (≈0.97), indicating near-identical behavior across the evaluated metrics. At the same time, other model pairs show substantially lower similarity, reflecting meaningful divergence in regional or process-level responses despite comparable large-scale behavior. This spread in similarity values indicates that, while many models share strong common structure, others retain distinct signatures that are not captured by ensemble-mean agreement alone. To identify pairs of models that are effectively redundant in a multi-metric sense, a cosine similarity threshold corresponding to approximately the 97th percentile of the pairwise similarity distribution (Fig. 9a; grey line), thereby retaining only the top ~3% of the most similar model pairs. For an ensemble of 45 models, this criterion isolates roughly 30 out of 990 unique model pairs, ensuring that only near-collinear behaviors in the high-dimensional metric space are classified as strongly similar. Model pairs exceeding this threshold are thus considered highly redundant, while those below it retain sufficient independence to contribute distinct information to the ensemble.

The hierarchical clustering dendrogram translates pairwise cosine similarity into a structured representation of ensemble organization, revealing clusters of closely related models as well as more weakly connected models (Fig. 9b). Based on the dendrogram, applying the top 3% threshold (grey line) isolates a small set of highly redundant configurations. The EC-Earth variants (EC-Earth3, EC-Earth3-Veg, EC-Earth3-Veg-LR) form a tight cluster, as do CNRM-CM6-1/CNRM-ESM2-1, CESM2/CESM2-WACCM, and GFDL-CM4/GFDL-ESM4, indicating near-collinear behavior across the full multi-metric space. Strong links are also evident among HadGEM3-GC31-

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

LL/MM and between HadGEM3-GC31-LL and UKESM1-0-LL, reflecting closely aligned responses despite nominal model differences, which are likely due to use of the same atmospheric general circulation model. In contrast, models such as GISS-E2-1-G, INM-CM4-8, and INM-CM5-0 do not form high-similarity links beyond their immediate counterparts, indicating more independent behavior relative to the tightly clustered groups. This dendrogram-based view clarifies which models are effectively interchangeable and which contribute distinct information, directly supporting redundancy-aware model sub-selection.

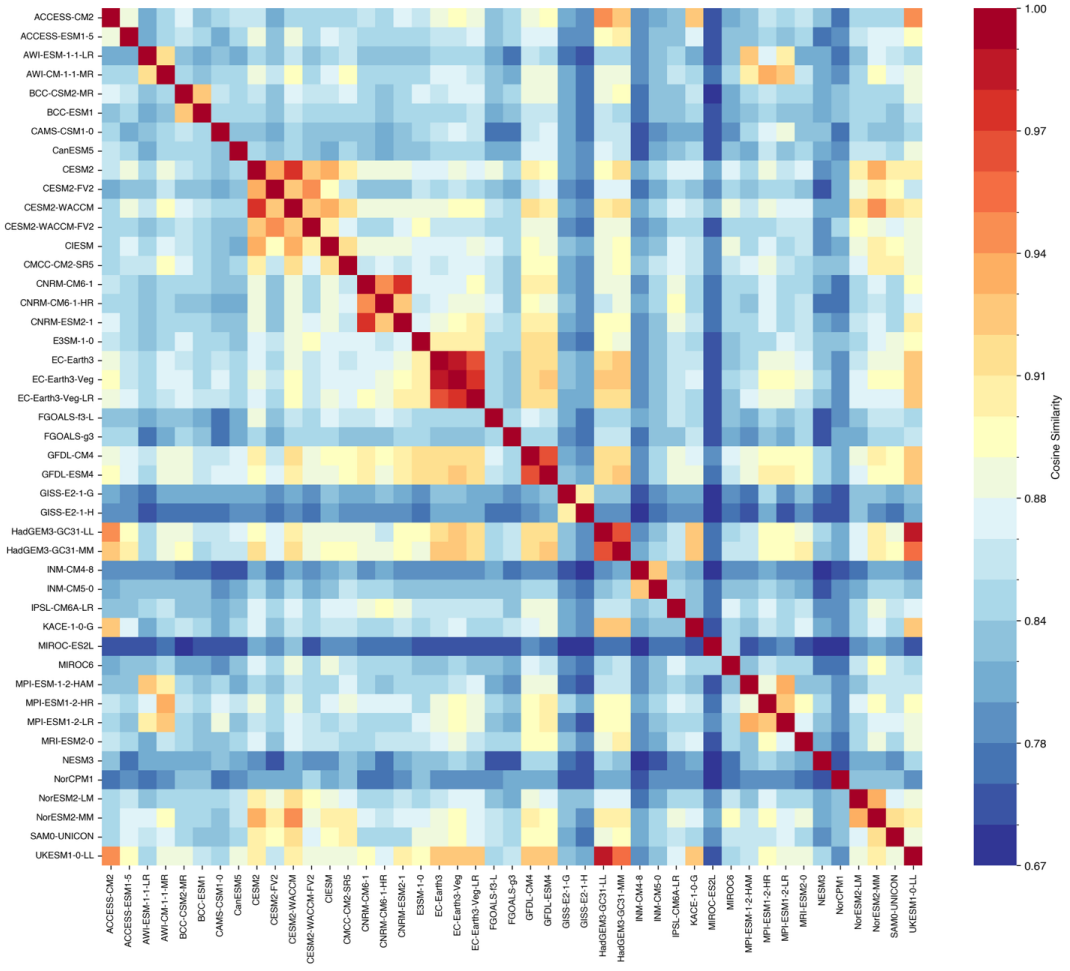

Figure 9 ***Model redundancy in the CMIP6 ensemble****. (a) Pairwise cosine similarity heatmap for all GCM pairs, with warmer colors indicating higher similarity.*

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

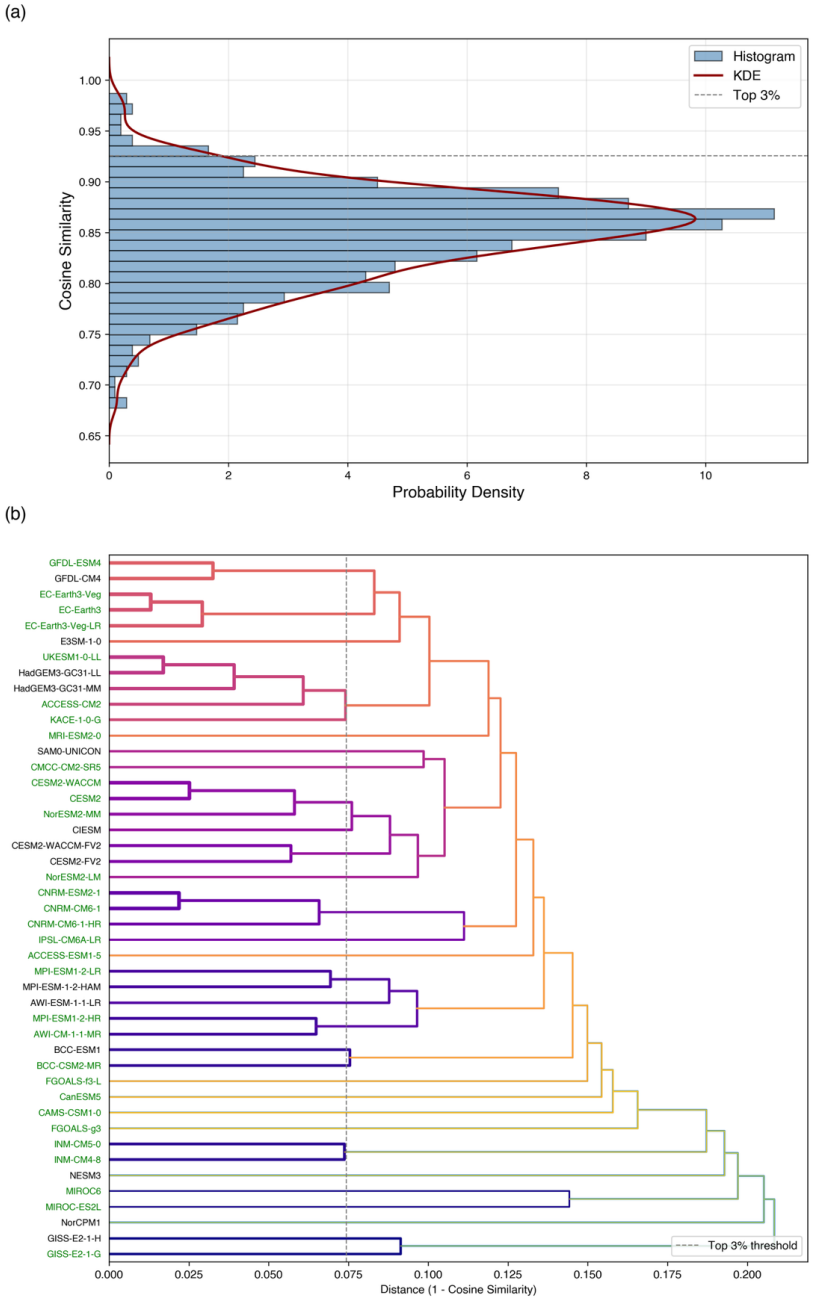

Figure 10. ***Model redundancy in the CMIP6 ensemble****. (a) Probability density distribution of cosine similarity values for the 990 unique model pairs. (b) Hierarchical clustering dendrogram constructed using average linkage and cosine distance; merge distance denotes the cosine distance at which clusters are joined. Models labeled in black do not have SSP3-7.0 output.*

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

3.5 Regional Climate Sensitivity

We use regional effective climate sensitivity to assess CMIP6 GCMs across CORDEX domains, analyzing projected annual temperature changes over 2015–2100 relative to the 1995–2014 reference period using 31 GCMs with available SSP3-7.0 future projections. This analysis is illustrated for the EUR domain by a heatmap showing mean annual warming for each year during 2071–2100 and the linear warming trend over 2015–2100, while broader tendencies across domains and models are summarized using a heatmap of projected temperature change for the final 30 years of the 21st century (Fig. 11). Corresponding heatmaps for the remaining CORDEX regions are provided in the *Supplementary Material* (Figs. S3*)*

Projected temperature changes over EUR during 2015–2100 exhibit substantial inter-model spread in both the magnitude of warming and its late-century amplification. Models such as UKESM1-0-LL and CanESM5 occupy the upper end of the ensemble, combining strong full-period trends with the largest late-century warming, indicating persistently high warming rates throughout the century. IPSL-CM6A-LR, EC-Earth3 and its variants also project strong warming, although with greater contrast between full-period means and late-century behavior.

Differences in temporal evolution further distinguish model behavior. Several models show relatively steady warming, with late-century trends comparable to or only modestly higher than full-period trends. For example, AWI-CM-1-1-MR, CESM2, MIROC6, and NorESM2-MM exhibit moderate late-century warming without pronounced amplification relative to their full-period behavior. In contrast, a subset of models shows clear late-century strengthening. UKESM1-0-LL, IPSL-CM6A-LR, EC-Earth3, and EC-Earth3-Veg exhibit substantially higher warming rates during the final 30 years than implied by their full-period trends, resulting in larger late-century mean warming. This contrast indicates that inter-model spread in end-of-century warming arises not only from differences in mean warming rates but also from differences in how warming accelerates or persists during the latter half of the century.

At the lower end of the ensemble, models such as MRI-ESM2-0, CAMS-CSM1-0, FGOALS-g3, and NorESM2-LM project weaker warming, with both full-period means and late-century warming remaining near or below 3 °C. Some of these models, including MRI-ESM2-0 and AWI-CM-1-1-MR, also exhibit relatively weak late-century trends, consistent with limited amplification toward the end of the century. Overall, the EUR results show that differences in late-century temperature projections reflect a combination of spread in full-period warming and systematic differences in the temporal evolution of warming across models.

Across domains (Fig. 11b) models largely preserve their relative position in terms of late-century warming compared to EUR. Models such as UKESM1-0-LL and CanESM5 consistently project the strongest late-century warming in all CORDEX regions, with last-30-year anomalies exceeding 5 °C in AFR, EUR, and EAS and remaining among the highest in SEA and AUS. IPSL-CM6A-LR and EC-Earth3, EC-Earth3-Veg, and EC-Earth3-Veg-LR also consistently project high late-century warming across regions, occupying the upper portion of the ensemble in most domains, with particularly strong warming over EUR and EAS.

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

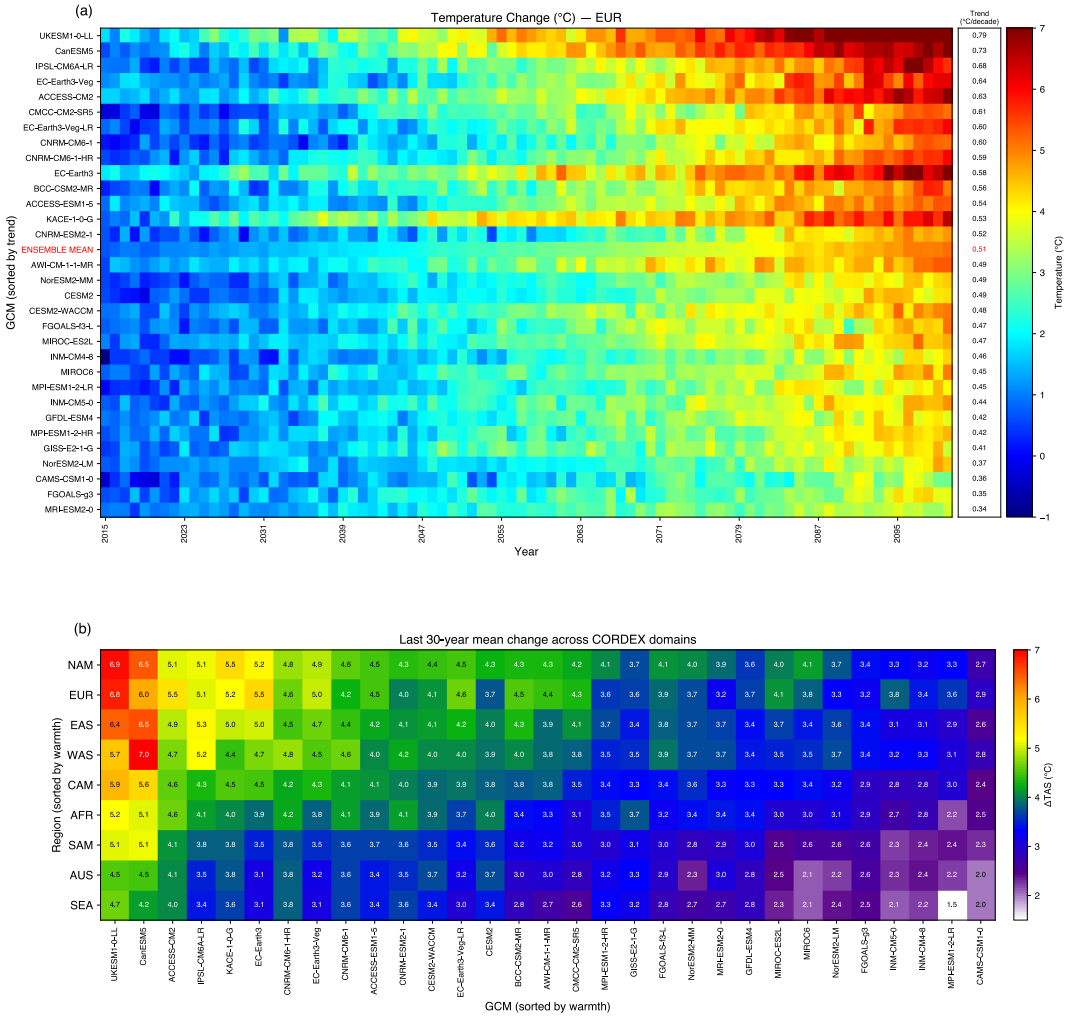

Figure 11 ***Projected temperature anomalies across CORDEX regions under the SSP3-7.0 scenario****. (a) Annual temperature anomalies for EUR (2015–2100) with decadal trend in the rightmost column. (b) Mean temperature anomaly for the last 30 years across all nine CORDEX domains.*

At the lower end of the ensemble, CAMS-CSM1-0, MRI-ESM2-0, NorESM2-LM, and MPI-ESM1-2-LR consistently remain among the coolest models across all domains, with late-century warming generally near or below 3 °C, mirroring their behavior over EUR. A smaller set of models exhibits clearer regional variation. For example, MIROC6 and FGOALS-f3-L project moderate late-century warming over EUR and EAS but substantially weaker warming over SEA and AUS, while CESM2 and CESM2-WACCM remain near the ensemble median across most regions. Overall, these results show that inter-model differences in late-century temperature projections are primarily controlled by intrinsic model sensitivity that is consistent across domains, with regional processes modulating the magnitude of warming without substantially altering the relative ordering of models.

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

3.6 Spread in Precipitation Projections

In Fig. 12, we analyze projected annual precipitation changes over 2015–2100 relative to the 1995–2014 reference period to assess projections spread among CMIP6 GCMs across CORDEX domains. This analysis is illustrated for the CAM domain by a heatmap showing mean annual changes during 2015–2100. Corresponding heatmaps for the remaining CORDEX regions are provided in the *Supplementary Material* (Figs. S4). Additionally, broader tendencies across CORDEX domains and models are summarized using the Wet–Dry Dominance Index (WDI), which quantifies the cumulative imbalance between positive (wet) and negative (dry) precipitation anomalies over the analysis period.

$$WDI = \sum_{P_t>0} P_t - \sum_{P_t<0} |P_t| \quad (8)$$

where P_t represents the precipitation anomaly in year t. Positive WDI values indicate that wet-year anomalies dominate over dry-year anomalies, whereas negative values indicate dominance of dry anomalies.

Projected precipitation changes over CAM during 2015–2100 exhibit substantial inter-model spread in both long-term trends and temporal evolution. Full-period trends range from weak to moderate drying across most models, but the temporal consistency of these trends varies markedly, indicating that differences among models arise not only from trend magnitude but also from how precipitation anomalies evolve through time.

Some models display coherent wet–dry dominance behavior despite overall drying. For example, EC-Earth3, EC-Earth3-Veg, and EC-Earth3-Veg-LR show similar drying trends together with consistently positive WDI values, indicating that wet-year anomalies dominate their long-term response. A similar pattern is evident in the CNRM models, which exhibit modest trends and near-neutral to positive WDI values with limited divergence between full-period and late-century behavior. In contrast, CanESM5, CESM2, CESM2-WACCM, MPI-ESM1-2-HR, MPI-ESM1-2-LR, and GFDL-ESM4 consistently show strongly negative WDI values, indicating dry-year dominance that is reflected in the prevalence of negative anomalies in the annual time series. Importantly, weak or modest trends do not imply uniform behavior. Models such as NorESM2-LM and FGOALS-f3-L exhibit relatively small long-term trends but near-neutral WDI values and alternating wet and dry anomalies through time, demonstrating that temporal variability and anomaly structure can differ substantially even when mean changes are small.

To place the CAM-specific results in a broader regional context, we further examine the WDI across all CORDEX domains using a summary visualization based solely on WDI (Fig. 10b). This multi-domain perspective reveals that several of the behaviors identified over CAM are not region-specific but recur across domains. Models such as INM-CM5-0 and INM-CM4-8, ACCESS-CM2, EC-Earth3 and its variants, and CNRM and its variants consistently exhibit strongly positive WDI values across most domains, indicating systematic dominance of wet-year anomalies irrespective of region. In contrast, CESM2, CESM2-WACCM, and GFDL-ESM4 show persistently negative WDI across domains, reflecting robust dry-year dominance that aligns with their behavior over CAM. Other models display pronounced regional dependence. For example, ACCESS-ESM1-5 and NorESM2-MM exhibit positive WDI in some domains (such as AFR and EAS) but negative values in others (including AUS and EUR), indicating that wet–dry dominance is sensitive to

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

regional processes even when long-term trends are comparable. Overall, comparisons of precipitation responses across domains (Figs. S6) and the associated WDI (Fig. 10b) values indicate that wet–dry dominance is not determined solely by the sign or magnitude of the mean precipitation change, but by how anomalies are distributed across wet and dry years.

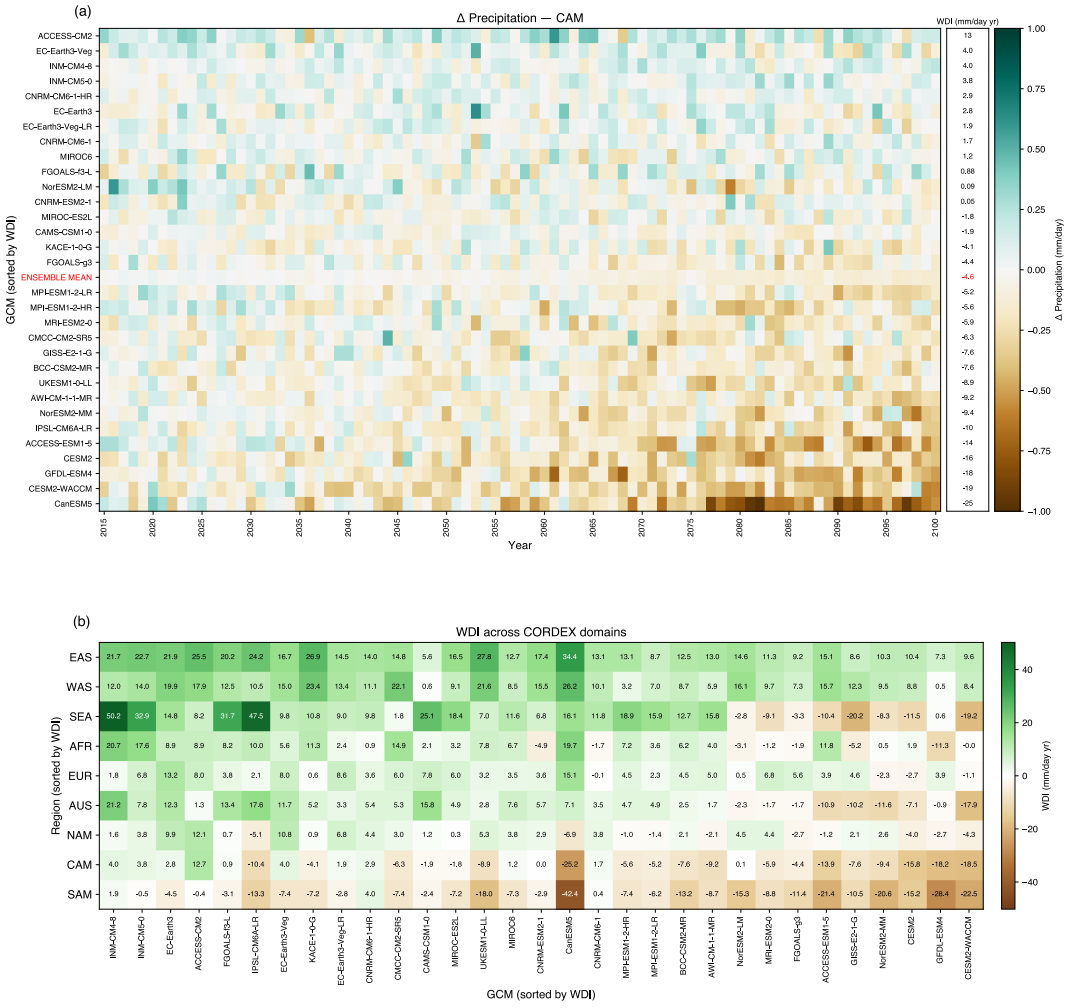

Figure 12 ***Precipitation anomalies and wet-dry index (WDI) across CORDEX regions****. (a) Annual PR anomalies for CAM (2015–2100) with model WDI values in the rightmost column. (b) WDI for all nine CORDEX domains across 31 models.*

3.7 ΔPR versus ΔTAS Relationship

In addition to the separate assessment of temperature and precipitation responses, we examine the joint relationship between ΔPR and ΔTAS to evaluate how regional precipitation responds to warming (Fig. 13). Precipitation responses reflect a combination of thermodynamic effects associated with increasing atmospheric moisture and dynamic effects linked to circulation changes (Shepherd 2014). Large inter-model differences in ΔPR therefore indicate a stronger role for

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

circulation variability, whereas consistent scaling across models suggests a response constrained primarily by warming.

Across the CORDEX domains, precipitation–temperature scaling varies substantially. Strong positive coupling occurs in EAS ($r = 0.89$) and WAS ($r = 0.83$), moderate coupling in EUR ($r = 0.37$), weak relationships in AFR ($r = 0.21$), AUS ($r = 0.11$), SEA ($r = 0.14$), and NAM ($r = -0.07$), and negative scaling in CAM ($r = -0.41$) and SAM ($r = -0.53$). These correlations are calculated from the 93 model–period pairs derived from the three 25-year transitions considered in this study.

The strong scaling in EAS and WAS indicates a precipitation response largely constrained by warming, with models projecting larger precipitation increases as temperatures rise. In contrast, AFR and NAM show little systematic relationship between precipitation and warming despite substantial temperature differences across models. This suggests that domain-mean precipitation responses are influenced by compensating subregional signals within the large CORDEX domains, heterogeneous wetting and drying patterns cancel in the domain mean.

CAM and SAM exhibit a tendency toward decreasing precipitation with increasing warming, indicating that circulation changes likely play a dominant role in shaping regional precipitation responses. Meanwhile, SEA and AUS show large inter-model spread with weak temperature scaling, implying strong influence from model-dependent circulation responses and regional feedbacks. In SEA in particular, some moderate-warming models project larger precipitation increases than models with stronger warming.

Overall, the joint ΔPR–ΔTAS analysis highlights strong regional contrasts in precipitation sensitivity to warming. While thermodynamic constraints dominate precipitation responses in some regions, others are primarily influenced by circulation variability or compensating subregional signals. This coupled perspective therefore provides a useful diagnostic for model selection by identifying regions where precipitation responses are robust versus those where uncertainty remains dominated by dynamic processes.

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

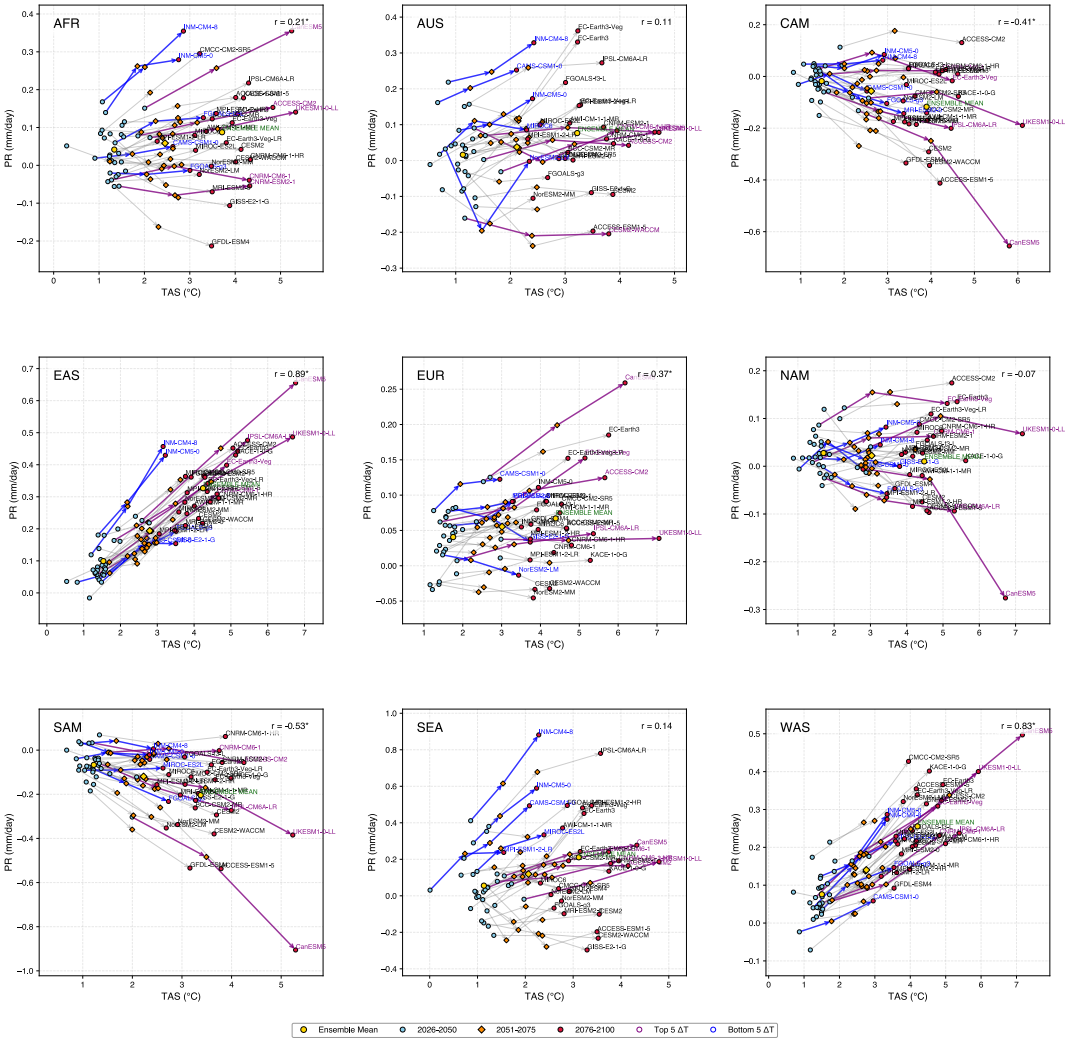

Figure 13 ***ΔPR versus ΔTAS Relationship****. Scatter plots of projected mean precipitation (PR, mm/day) versus mean near-surface air temperature (TAS, °C) across 31 CMIP6 GCMs for nine CORDEX domains. The last 75 years of the projection period (2026–2100) are divided into three equal 25-year windows: 2026–2050 (blue circles), 2051–2075 (orange diamonds), and 2076–2100 (red circles). Arrows trace each model's trajectory. Models in the top 5 and bottom 5 of the temperature change ranking are highlighted in bold purple and blue, respectively. Pearson correlation coefficients (r) between PR and TAS, computed from all 93 data pairs (31 models × 3 periods), are shown in the upper-right corner of each panel; asterisks denote statistical significance at $p < 0.05$.

4. Synthesis of Model Selection Framework

The Sankey diagram (Fig. 14) is based on global aggregation across the CORDEX domains and summarizes the four criteria used in the model sub-selection framework: historical performance, model independence, climate sensitivity, and projected precipitation change. The leftmost nodes represent the full CMIP6 ensemble of 45 models, each labeled by global performance rank. The top 14 ranked models in the Sankey diagram do not provide SSP3-7.0 simulations and are therefore

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

used only to establish the relative ranking of the remaining models within the broader CMIP6 ensemble. Subsequent analysis focuses on the 31 models with available SSP3-7.0 data. The regional Sankey diagrams are provided in the *Supplementary Material* (Figs. S5). An interactive HTML version is also provided to facilitate interpretation of the synthesis.

For these 31 models, the intermediate nodes represent tercile bins of projected near-surface air temperature change (ΔTAS) over the last 30 years of the 21st century, categorized as low, mid, and high warming. The rightmost nodes represent precipitation change (ΔPR) categories, classified as wet (>5 mm yr^{-1}), neutral (within ±5 mm yr^{-1}), or dry (<−5 mm yr^{-1}). Model independence is indicated by colored rectangles adjacent to the model labels. Models exceeding a cosine similarity threshold of top 3% are treated as strongly similar and share the same color, indicating cluster membership, while models below this threshold are shown with white rectangles and are considered relatively independent. This structure provides a compact view of model performance, projected warming, precipitation response, and independence within a single summary framework. The global signal indicates that the majority of the models project on average wetter conditions independently of the projected warming, but as shown in Fig. 11 there are a variety of regional responses.

Applying the cosine similarity threshold yields 19 clusters among the 31 models with future projections, comprising six multi-model clusters and 13 singleton models. Several coherent clusters are evident. The CESM/NorESM family, all using versions of the Community Atmosphere Model (Brands 2022b), forms a four-model cluster (NorESM2-MM, CESM2-WACCM, CESM2, NorESM2-LM), largely occupying the low-to-mid warming and dry precipitation bins. The EC-Earth3 family (EC-Earth3-Veg, EC-Earth3-Veg-LR, EC-Earth3) consistently projects high warming and wet precipitation changes. A similar high-warming, wet-precipitation signal is observed in the UKESM/KACE/ACCESS cluster (UKESM1-0-LL, KACE-1-0-G, ACCESS-CM2, using Met Office UM in the atmosphere), with UKESM1-0-LL exhibiting particularly strong warming and precipitation responses relative to most other models. The MPI/AWI cluster (AWI-CM-1-1-MR, MPI-ESM1-2-HR, MPI-ESM1-2-LR, using ECHAM6 in the atmosphere) is characterized by mid-to-low warming combined with wet precipitation responses, while the CNRM family (CNRM-ESM2-1, CNRM-CM6-1, CNRM-CM6-1-HR) spans mid-to-high warming with predominantly wet outcomes. The INM models (INM-CM5-0, INM-CM4-8) form a distinct low-warming cluster with exceptionally strong wetting signals. The GCMs precipitation responses are consistent with Fig. 10b.

When interpreted in the context of historical performance, the Sankey diagram indicates that the set of viable candidates for sub-selection is limited to a small number of similarity clusters. Clusters anchored by consistently high-performing models, including the CESM/NorESM and EC-Earth3 families, emerge as the most plausible sources of candidate models. The MPI/AWI cluster also remains competitive due to the presence of multiple well-performing members. In addition, high-performing singleton models can also enter the candidate pool. For example, GFDL-ESM4 appears as an independent singleton cluster with a high-performance rank, projecting moderate warming and drying, and is classified within the low-warming, dry-precipitation category.

Across these performance-competitive clusters, models separate consistently into three warming ranges defined by late-century ΔTAS: low (< ~3 °C), mid (~3–4 °C), and high (> ~4.5 °C), with corresponding shifts in precipitation behavior from predominantly dry/neutral to increasingly wet conditions. The viable candidate pool is therefore effectively confined to a small number of models

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

within each of these ranges, each represented by closely related configurations identified through the independence analysis. Selecting across these three ranges ensures retention of the dominant spread in both temperature and precipitation responses while avoiding redundancy within clusters.

In contrast, clusters such as the INM family, which are dominated by lower-ranked models, are unlikely to contribute candidates despite their relative independence and pronounced precipitation responses. Within this framework, independence and projected warming and precipitation characteristics serve to differentiate among performance-competitive models and clusters rather than compensate for weak historical performance.

Finally, sub-selection is subject to practical data constraints. Regional dynamical downscaling requires the availability of high-resolution, three-dimensional atmospheric variables at sub-daily temporal resolution; consequently, model choice is further constrained by the availability of the required data products, especially within the CORDEX-CMIP context.

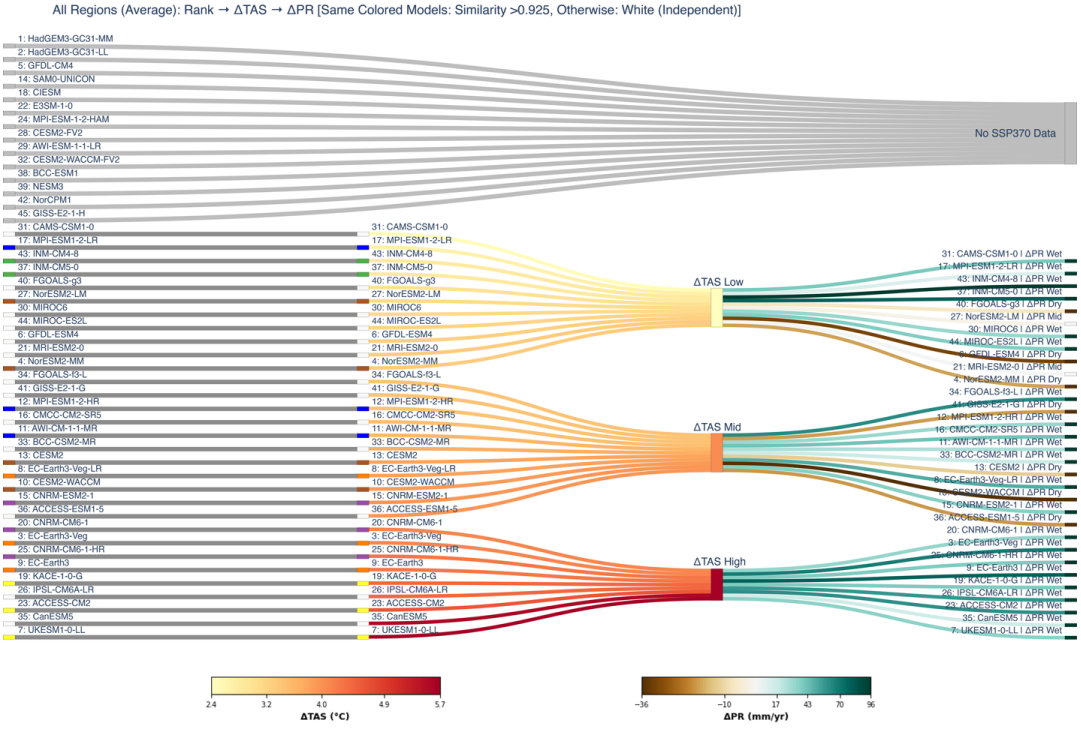

Figure 14 ***Synthesis of multi-criteria model evaluation framework****. Sankey diagram illustrating model rankings (numbers), independence (colored rectangles), and climate sensitivity and precipitation spread based on global aggregation across nine CORDEX domains.*

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

5 Appendix

This appendix presents the definitions of the indices and the criteria used to evaluate model performance for each of them.

5.1 Amplitude

The amplitude is defined as the difference between the maximum and minimum of the monthly climatological mean values of precipitation or temperature over the annual cycle. Let X_m represent the climatological mean for month m, where $m = 1, 2, \ldots, 12$. Then the amplitude A_p is given by:

$$A_p = \max_{1 \le m \le 12} (X_m) - \min_{1 \le m \le 12} (X_m) \tag{A1}$$

Here, X_m is computed as the multi-year average of monthly mean of precipitation or temperature for each calendar month over the analysis period:

$$X_m = \frac{1}{N} \sum_{y=1}^{N} X_{m,y} \tag{A2}$$

where $X_{m,y}$ is the mean in month m of year y, and N is the number of years in the climatological period.

Models are evaluated using three statistical measures. Spatially centered pattern correlation (PC) quantifies the similarity between simulated and observed spatial distributions of amplitude. Bias, normalized by the temporal standard deviation of ERA5 during the reference period, measures systematic offsets in magnitude. Spatially centered RMSE, normalized by the spatial standard deviation of ERA5, estimates the total spatial error. Each statistical measure is bounded between 0 and 100 using penalty-based functions:

$$PC_{[0,100]} = \max\left(0,\ 100 \times PC\right), \tag{A3}$$

$$Bias_{[0,100]} = \max\left(0,\ 100 \times e^{-\frac{Bias}{N}}\right), \tag{A4}$$

$$RMSE_{[0,100]} = \max\left(0,\ 100 \times e^{-\frac{RMSE}{N}}\right), \tag{A5}$$

where N is a scaling factor, set to twice the reference standard deviation. For example, if $|Bias| = 2\sigma$ and $N = 2\sigma$, then

$$Bias_{[0,100]} = 100 \times e^{-1} \approx 36.8, \tag{A6}$$

illustrating that a bias equal to two standard deviations reduces the score to about 37. Finally, the three statistical measures are averaged to provide an overall skill score on a 0–100 scale.

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

5.2 Annual Standard Deviation

The annual standard deviation (SD) quantifies the year-to-year variability in total annual magnitude over a specified climatological period. Let X_y represent the total precipitation or temperature in year y, where $y = 1, 2, \ldots, N$ and N is the number of years in the reference period. The annual SD, denoted as σ_X, is defined as:

$$\sigma_X = \sqrt{\frac{1}{N}\sum_{y=1}^{N}\left(X_y - \bar{X}\right)^2} \tag{A7}$$

where $\bar{X}$ is the mean annual mean over the period:

$$\bar{X} = \frac{1}{N}\sum_{y=1}^{N} X_y \tag{A8}$$

Model evaluation follows same methodology as that for amplitudes.

5.3 Seasonal Averages

The seasonal average for a season is calculated by averaging monthly values within the season. As an example, the seasonal average for DJF is calculated by averaging monthly values for December, January, and February, taking care to account for the year boundary. Let $X_{m,y}$ represent the climate variable (e.g., precipitation or winds) for month m and year y, where $m \in \{1, 2, \ldots, 12\}$. For each year y, the DJF seasonal mean X_y^{DJF} is given by:

$$X_y^{\mathrm{DJF}} = \frac{1}{3}\left(X_{12,y-1} \quad X_{1,y} \quad X_{2,y}\right) \tag{A9}$$

Model evaluation follows same methodology as that for amplitudes.

5.4 Peak Timing

The peak timing of precipitation or temperature is defined as the calendar month during which the climatological monthly mean of the variable reaches its maximum. Let X_m denote the climatological mean of a variable (e.g., precipitation or temperature) for month m, where $m = 1, 2, \ldots, 12$. The peak timing m_{peak} is then given by:

$$m_{\mathrm{peak}} = \arg\max_{1 \le m \le 12} X_m \tag{A10}$$

Here, $m_{\mathrm{peak}} \in \{1, 2, \ldots, 12\}$ corresponds to the month (1 = January, ..., 12 = December) in which the variable attains its maximum climatological value.

Model evaluation for this metric quantifies only the similarity between simulated and observed spatial distributions using pattern correlation.

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

5.5 Seasonality and Entropy

Feng et al. (2013) introduced the use of relative entropy (RE) and the dimensionless seasonality index (DSI) as robust measures of hydroclimate seasonality. RE quantifies the concentration of annual precipitation into specific months relative to a uniform distribution, while DSI combines RE with normalized annual precipitation to account for both the seasonality and the overall magnitude of rainfall. This framework allows identification of regions with strong monsoonal characteristics, where rainfall is both highly seasonal and abundant. Let r_m be the mean precipitation in month m, and let total annual precipitation be $R = \sum_{m=1}^{12} r_m$. Define the fractional monthly distribution as $p_m = \frac{r_m}{R}$, and the uniform reference distribution as $q_m = \frac{1}{12}$. The relative entropy (RE) is given by:

$$\mathrm{RE} = \sum_{m=1}^{12} p_m \log_2 \left(\frac{p_m}{q_m} \right) \tag{A11}$$

The dimensionless seasonality index (DSI) combines RE with normalized annual rainfall:

$$\mathrm{DSI} = \left(\frac{R}{R_{\max}} \right) \cdot \mathrm{RE} \tag{A12}$$

Here, $R_{\max}$ is the maximum annual precipitation among all locations considered, used to normalize DSI values across space.

The model evaluation method is the same as that used for amplitude. However, because both seasonality and entropy are based on climatological values, no temporal standard deviation is available to standardize the bias. Instead, bias is expressed as the fractional departure from the reference value and penalized using $N = 2$, which corresponds to twice the reference value.

5.6 Monsoon Onset and Withdrawal

The timing of monsoon onset and withdrawal is a critical aspect of regional hydroclimate because it determines the effective length of the rainy season, influences agricultural planning, and affects water resource availability. Shifts in these timings can lead to substantial socio-economic impacts, making their accurate characterization and projection essential for both climate research and adaptation strategies. Ashfaq et al. (2020) determine monsoon onset and withdrawal from the climatological cumulative precipitation fraction curve. The onset is defined as the pentad corresponding to the *minimum* of the curve, representing the beginning of sustained rainfall accumulation. The withdrawal is defined as the pentad corresponding to the *maximum* of the curve, representing the end of the rainy season. Let $P_{k,y}$ denote the precipitation in pentad k of year y, where $k = 1, 2, \ldots, 73$. Define the climatological mean pentad precipitation:

$$\bar{P}_k = \frac{1}{N} \sum_{y=1}^{N} P_{k,y} \tag{A13}$$

where N is the number of years in the historical period. Define the cumulative precipitation fraction up to pentad k:

$$F_k = \frac{\sum_{i=1}^{k} \bar{P}_i}{\sum_{i=1}^{K} \bar{P}_i} \tag{A14}$$

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

Then onset pentad k_{onset} is the smallest k such that:

$$F_{k_{\text{onset}}} \geq f_{\text{onset}} \tag{A15}$$

Similarly, withdrawal pentad k_{withdraw} is the largest k such that:

$$F_{k_{\text{withdraw}}} \leq f_{\text{withdraw}} \tag{A16}$$

For monsoon onset and withdrawal, evaluation is based only on bias. Because these indices are circular in nature, onset and withdrawal dates are first converted to circular values using

$$\theta_{\text{onset}} = \frac{2\pi}{73} k_{\text{onset}}, \qquad \theta_{\text{withdraw}} = \frac{2\pi}{73} k_{\text{withdraw}} \tag{A17}$$

The bias is then computed as the absolute difference between the ERA5 and model circular values. Bounding is applied with $N = 3$, which corresponds to three pentads (15 days). Thus, when the model onset or withdrawal differs from ERA5 by three pentads, the score decays to about 37 (since $e^{-1} \approx 0.37$), with larger differences further penalized toward zero.

5.7 Vertical Shear of Winds

The vertical shear of winds between 200 hPa and 850 hPa is computed using monthly climatological means of the zonal or meridional wind component. For instance, let $\overline{u}_{200,m}$ and $\overline{u}_{850,m}$ denote the climatological monthly mean zonal wind at 200 hPa and 850 hPa, respectively, for month m, where $m = 1, 2, \ldots, 12$. The vertical wind shear for month m, denoted S_m, is given by

$$S_m = \overline{u}_{200,m} - \overline{u}_{850,m}. \tag{A18}$$

A negative value of S_m indicates stronger upper-level easterlies relative to the lower-level winds, which is important for diagnosing monsoon dynamics and tropical wave activity. This analysis is conducted only for the monsoon regions as defined in Ashfaq et al. (2021). It yields 12 climatological monthly values that are used to evaluate spatially centered pattern correlation (PC) and RMSE. For RMSE, normalization is based on the standard deviation of the 12 monthly values. For bias, only the monsoon months for each region are considered; the bias in each month is normalized by the observed standard deviation for that month and then averaged over the season.

5.8 Zonally Averaged Precipitation (Hovmöller)

Zonally averaged precipitation (Hovmöller diagrams; Hovmöller, 1949) provides a latitude–time representation of rainfall obtained by averaging precipitation along longitudes. This diagnostic is commonly used to evaluate the meridional structure, seasonal migration, and variability of precipitation belts, including the position and strength of the Intertropical Convergence Zone (ITCZ) and other large-scale precipitation features. Let $\overline{P}(\lambda, \phi, m)$ be the climatological mean

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

precipitation for longitude λ, latitude ϕ, and month m. The zonal average at latitude ϕ for month m, denoted $\overline{P}(\phi, m)$, is defined as:

$$\overline{P}(\phi, m) = \frac{1}{N} \sum_{i=1}^{N} \overline{P}(\lambda_i, \phi, m) \tag{A19}$$

where N is the number of longitudinal grid points in the averaging interval. Model evaluation follows same methodology as that for amplitudes.

5.9 Caribbean Low-Level Jet Index

The Caribbean Low-Level Jet (CLLJ) is a key circulation feature of the Intra-Americas region. Centered near 925 hPa, it transports large amounts of moisture from the tropical Atlantic into the Caribbean and Central America, strongly influencing seasonal rainfall variability, the development of tropical storms, and moisture availability for the North and South American monsoon systems. Because of its role in regional hydroclimate and tropical cyclone activity, the CLLJ serves as an important metric for evaluating the ability of models to reproduce low-level circulation and moisture transport. Following Wang (2007), The CLLJ is represented here as the average 925-hPa zonal wind over the region 12°–18°N latitude and 280°–290°E longitude. Let $u(\phi, \lambda, 925, t)$ denote the zonal wind at latitude ϕ, longitude λ, and time t. The index is given by

$$CLLJ(t) = \frac{1}{N} \sum_{\phi=12^\circ}^{18^\circ} \sum_{\lambda=280^\circ}^{290^\circ} u(\phi, \lambda, 925, t), \tag{A20}$$

where N is the number of grid points within the averaging region. The evaluation of the Caribbean Low-Level Jet (CLLJ) follows the same procedure as for vertical wind shear, except that the bias is computed using all months of the annual cycle rather than only a specific season.

5.10 Atlantic ITCZ Position

The Atlantic Intertropical Convergence Zone (ITCZ) is the dominant rainfall band over the tropical Atlantic, and its seasonal migration strongly influences regional climate and rainfall variability in adjacent continental areas. To characterize its mean position, the ITCZ is analyzed over the domain ϕ_S–ϕ_N in latitude and λ_W–λ_E in longitude, where $\phi_S = -10^\circ$, $\phi_N = 25^\circ$, $\lambda_W = 260^\circ$, and $\lambda_E = 305^\circ$. The ITCZ position is defined here as the precipitation-weighted centroid, extending the commonly used precipitation-weighted latitude definition (e.g., Biasutti et al., 2006) to both latitude and longitude. The monthly climatological ITCZ longitude (λ_{ITCZ}) and latitude (ϕ_{ITCZ}) are then computed as precipitation-weighted averages of the precipitation field.

$$\left(\lambda_{\mathrm{ITCZ}}(t), \phi_{\mathrm{ITCZ}}(t)\right) = \frac{\sum_{\phi=\phi_S}^{\phi_N} \sum_{\lambda=\lambda_W}^{\lambda_E} (\lambda, \phi)\, F(\phi, \lambda, t)}{\sum_{\phi=\phi_S}^{\phi_N} \sum_{\lambda=\lambda_W}^{\lambda_E} F(\phi, \lambda, t)}. \tag{A21}$$

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

For the Atlantic ITCZ, evaluation is based on bias and the seasonal cycle of its latitude and longitude locations. The bias reflects the strength of the ITCZ, while the monthly intensity-weighted latitude and longitude positions ($m = 1, \ldots, 12$) are compared between model and reference using pattern correlation. This provides a measure of the seasonal march of the ITCZ in both directions. The two pattern correlations (latitude and longitude) together with the bias score are averaged to obtain the final skill score.

5.11 Zonally Averaged Upper Tropospheric Temperature

This diagnostic follows the formulation of Ashfaq et al. (2021) and is limited to the nine global monsoon regions defined in that study. The zonal mean distribution of upper-tropospheric temperature is a key diagnostic for evaluating large-scale circulation features in climate models, as it captures the meridional gradients and vertical structure of warming that influence the Hadley circulation, subtropical jets, and monsoon dynamics.

Let $\overline{T}(p, \phi, \lambda, m)$ denote the monthly climatological mean temperature at pressure level p, latitude ϕ, longitude λ, and month m ($m = 1, 2, \ldots, 12$). The zonally averaged monthly climatology is defined as

$$\bar{T}(p, \phi, m) = \frac{1}{N} \sum_{i=1}^{N} \overline{T}(p, \phi, \lambda_i, m), \tag{A22}$$

where N is the number of longitudinal grid points. To emphasize the seasonal march of meridional temperature gradients, the annual mean zonal temperature is removed:

$$\bar{T}'(p, \phi, m) = \bar{T}(p, \phi, m) - \frac{1}{12} \sum_{m=1}^{12} \bar{T}(p, \phi, m). \tag{A23}$$

The anomaly field $\bar{T}'(p, \phi, m)$ is typically displayed as a latitude–pressure cross-section for each calendar month, providing a clear representation of the seasonal cycle in upper-tropospheric temperature gradients and associated circulation features. For zonally averaged upper-tropospheric temperature, the evaluation follows the same procedure as for amplitude, except that the data are organized along latitude and month dimensions, with months representing climatological means. Pattern correlation and RMSE are computed across the latitude–month field. Bias is calculated for each month after normalizing by the observed monthly standard deviation, and then averaged across all months to obtain the overall bias score.

5.12 East Asian Summer Monsoon Index

The strength of the East Asian summer monsoon is defined using the zonal sea-level pressure (SLP) gradient between land at 110°E and ocean at 160°E across latitudes 10°N–50°N during JJA (Guo, 1983).

$$\text{EASMI} = \sum_{\phi=10^\circ N}^{50^\circ N} [\text{SLP}(160^\circ E, \phi) - \text{SLP}(110^\circ E, \phi)] \tag{A24}$$

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

A stronger land–sea pressure contrast, reflected in more negative values of the index, corresponds to a stronger summer monsoon.

Given that this index produces a single seasonal value (JJA mean) for each year, model evaluation is restricted to normalized bias. Range is used in place of interannual variability due to its larger magnitude in the reference data. Accordingly, $N = 1$ is applied in the bounded bias formulation, with scores constrained between 0 and 100.

5.13 Western Pacific Subtropical High Index

The Western Pacific Subtropical High (WPSH) is a dominant circulation system in the boreal summer that strongly influences the East Asian climate (Lu, 2002; Zhou et al., 2009). Its intensity and westward extension modulate monsoon rainfall distribution, typhoon tracks, and heatwave frequency across China, Korea, and Japan. Accurate representation of the WPSH is therefore a critical benchmark for assessing the skill of climate models in simulating East Asian summer climate. The evaluation is restricted to June–July–August (JJA), when the WPSH exerts its maximum influence.

The WPSH index is defined as the area-averaged 850-hPa geopotential height over the region 15°–25°N, 115°–150°E:

$$\mathrm{WPSHI} = \frac{1}{N} \sum_{\phi=15^\circ}^{25^\circ} \sum_{\lambda=115^\circ}^{150^\circ} Z_{850}(\phi, \lambda), \tag{A25}$$

where $Z_{850}(\phi, \lambda)$ is the climatological mean 850-hPa geopotential height, and N is the number of grid points within the specified domain.

5.14 African Easterly Jet Index

The African Easterly Jet (AEJ) is a prominent mid-tropospheric circulation feature that develops during boreal summer as a result of the strong meridional temperature gradient between the hot Sahara and the cooler Gulf of Guinea (Cook, 1999). It plays a central role in modulating West African rainfall, the formation of African easterly waves, and the development of tropical cyclones over the Atlantic. Evaluating the AEJ is therefore crucial for assessing model skill in representing the dynamics of the West African monsoon system.

The AEJ intensity index is defined as the average 600-hPa zonal wind over the domain 9°–21°N latitude and 10°W–20°E longitude. Let $u(\phi_i, \lambda_j, 600, t)$ denote the zonal wind at latitude ϕ_i, longitude λ_j, and time t. The index is computed as

$$AEJ_{600}(t) = \frac{1}{N} \sum_{\phi=9^\circ}^{21^\circ} \sum_{\lambda=10^\circ W}^{20^\circ E} u(\phi, \lambda, 600, t), \tag{A26}$$

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

where N is the total number of grid points in the averaging domain.

In addition to its intensity, the latitudinal location of the AEJ is an important diagnostic. Following the approach used for the Atlantic ITCZ, the jet latitude is defined as the zonal-wind-weighted mean latitude within the same domain:

$$\phi_{AEJ}(t) = \frac{\sum_{\phi=9^\circ}^{21^\circ} \sum_{\lambda=10^\circ W}^{20^\circ E} \phi\, u(\phi, \lambda, 600, t)}{\sum_{\phi=9^\circ}^{21^\circ} \sum_{\lambda=10^\circ W}^{20^\circ E} u(\phi, \lambda, 600, t)}. \tag{A27}$$

This formulation captures both the strength and meridional position of the AEJ, allowing evaluation of its seasonal cycle and interannual variability.

5.15 South Asian Monsoon Lows Index

Monsoon lows and depressions are critical synoptic systems of the South Asian summer monsoon, supplying a large fraction of seasonal rainfall and modulating intraseasonal variability. Their development depends on a favorable background environment characterized by both sufficient mid-tropospheric moisture and strong large-scale circulation. To represent this background state, Following Rastogi et al. (2018), we defined Monsoon Lows Index (MLI) as the product of lower–mid tropospheric moisture and dynamical forcing terms. This diagnostic provides a compact measure of the environmental conditions conducive to MD genesis and propagation.

$$MLI(t) = Q_{850\text{–}500}(t) \times (|U_{850}(t)| \quad |V_{200}(t)|)\,, \tag{A28}$$

where $Q_{850\text{–}500}$ is the specific humidity averaged over 10°–30°N, 80°–100°E and vertically between 850–500 hPa, U_{850} is the 850-hPa zonal wind averaged over 5°S–18°N, 50°–80°E, and V_{200} is the 200-hPa meridional wind averaged over 5°S–30°N, 80°–100°E.

5.16 Mozambique Channel Trough

The Mozambique Channel Trough (MCT) is a recurrent low-level circulation feature during austral summer that influences rainfall variability across southeastern Africa and modulates moisture transport into the South Indian Ocean Convergence Zone (Barimalala et al., 2020). Its strength and year-to-year variability are therefore key diagnostics for assessing the representation of regional circulation– precipitation coupling in climate models.

The MCT strength index is defined as the DJF mean 850-hPa geopotential height averaged over 25°S–15°S and 35°E–45°E:

$$MCT(y) = \langle Z_{850}(\phi, \lambda, y) \rangle_{\phi \in [-25^\circ, -15^\circ],\, \lambda \in [35^\circ, 45^\circ]}, \tag{A29}$$

where $Z_{850}(\phi, \lambda, y)$ is the seasonal mean 850-hPa geopotential height for year y, and $\langle \cdot \rangle$ denotes

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

a spatial average. The amplitude of the trough is defined as the interannual standard deviation of the DJF means:

$$A_{MCT} = \mathrm{SD}\big(MCT(y)\big), \quad y = 1, 2, \ldots, T. \tag{A30}$$

5.17 East African Low-Level Jet Stream

The East African Low-Level Jet Stream (EALLJ) is a prominent feature of the boreal summer circulation, characterized by strong southerly flow at 850 hPa along the East African coast (Findlater, 1969; Yang et al., 2015). It plays a crucial role in transporting moisture from the western Indian Ocean into East Africa and the Horn of Africa, thereby influencing rainfall variability and the strength of the East African monsoon. Its intensity and position are key diagnostics for evaluating model skill in capturing regional hydroclimate processes.

The EALLJ strength is defined as the seasonal (JJAS) mean of the maximum southerly wind at 850 hPa within the domain 5°–15°N and 40°–60°E. Let $v_{850}(\phi, \lambda, t)$ denote the 850-hPa meridional wind at latitude ϕ, longitude λ, and time t. The jet strength index is given by

$$EALLJ(t) = \frac{1}{T} \sum_{m \in JJAS} \max_{\phi \in [5°,15°],\, \lambda \in [40°,60°]} v_{850}(\phi, \lambda, m), \tag{A31}$$

where T is the number of JJAS months.

The latitudinal location of the EALLJ is defined as the wind-weighted mean latitude of the jet core within the same region:

$$\phi_{EALLJ}(t) = \frac{\sum_{\phi=5°}^{15°} \sum_{\lambda=40°}^{60°} \phi\, v_{850}(\phi, \lambda, t)}{\sum_{\phi=5°}^{15°} \sum_{\lambda=40°}^{60°} v_{850}(\phi, \lambda, t)}. \tag{A32}$$

These diagnostics together capture both the seasonal strength and meridional position of the EALLJ.

5.18 Azores High

The Azores High is a semi-permanent subtropical anticyclone of the North Atlantic that strongly influences climate variability over Europe, North Africa, and the Mediterranean (Rodwell and Hoskins, 2001; Hurrell et al., 2003). Its intensity and seasonal migration modulate storm tracks, precipitation regimes, and surface climate across these regions. Accurate representation of the Azores High is therefore essential for evaluating the skill of climate models in reproducing North Atlantic– Mediterranean circulation patterns.

The Azores High index is defined using climatological monthly mean sea level pressure (PSL) averaged over 25°–35°N and 40°W–20°W (longitudes converted to the −180°–180° format).

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

Let $PSL(\phi, \lambda, m)$ denote the monthly climatological mean sea level pressure at latitude ϕ, longitude λ, and month m. The index is

$$AH(m) = \langle PSL(\phi, \lambda, m) \rangle_{\phi \in [25°, 35°],\, \lambda \in [-40°, -20°]}, \tag{A33}$$

where $\langle \cdot \rangle$ denotes the spatial average over the defined domain. Intra-annual characteristics of the Azores High are assessed from the monthly cycle of $AH(m)$, $m = 1, \ldots, 12$.

5.19 Mascarene High Index

The Mascarene High is a dominant subtropical anticyclone of the Southern Hemisphere, located over the southwest Indian Ocean. During boreal summer it plays a central role in modulating the cross-equatorial pressure gradient and driving low-level monsoon flow into South Asia (Morioka et al., 2012). Its intensity and variability are therefore key indicators of monsoon strength and moisture transport.

The Mascarene High index is defined as the JJA mean sea level pressure (SLP) averaged over 40°S–25°S and 45°E–100°E. Let $SLP(\phi, \lambda, y)$ denote the seasonal mean SLP at latitude ϕ, longitude λ, and year y. The index for year y is given by

$$MH(y) = \langle SLP(\phi, \lambda, y) \rangle_{\phi \in [-40°, -25°],\, \lambda \in [45°, 100°]}. \tag{A34}$$

Evaluation focuses on two aspects: (i) the mean bias in $MH(y)$ relative to observations, which represents the climatological strength of the Mascarene High, and (ii) its interannual variability, computed as the standard deviation of the seasonal means:

$$\sigma_{MH} = \mathrm{SD}\big(MH(y)\big), \quad y = 1, 2, \ldots, T. \tag{A35}$$

5.20 Caribbean Bimodal Rainfall

The Caribbean Basin exhibits a distinct bimodal rainfall, with peaks typically occurring in early and late summer separated by a midsummer drought (Magaña et al., 1999; Wang et al., 2010). This seasonal cycle is controlled by the migration of the Intertropical Convergence Zone (ITCZ), the strength of the Caribbean Low-Level Jet, and local ocean–atmosphere interactions. Accurate simulation of the bimodal pattern is essential for evaluating model performance in representing rainfall variability and water resource seasonality across the region.

For the domain 10°–20°N and 65°–83°W, the monthly mean precipitation is first averaged spatially to obtain the climatological seasonal cycle:

$$P(m) = \langle P(\phi, \lambda, m) \rangle_{\phi \in [10°, 20°],\, \lambda \in [-83°, -65°]}, \quad m = 1, 2, \ldots, 12. \tag{A36}$$

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

Evaluation focuses on (i) mean bias in $P(m)$ relative to observations, and (ii) the representation of intraseasonal characteristics, including the timing and amplitude of the two seasonal rainfall maxima and the intervening midsummer drought.

5.21 South Atlantic High Pressure

The South Atlantic High (SAH) is a semi-permanent subtropical anticyclone that plays a key role in modulating the South American monsoon system, cross-equatorial moisture transport, and the position of the South Atlantic Convergence Zone (Carvalho et al., 2004). Its seasonal cycle in intensity and location strongly influences rainfall variability over South America and adjacent ocean basins, making it an important diagnostic for climate model evaluation.

The SAH index is defined using monthly climatological mean sea level pressure (PSL) averaged over 45°S–15°S and 20°W–10°E. Let $PSL(\phi, \lambda, m)$ denote the climatological monthly PSL at latitude ϕ, longitude λ, and month m. The mean strength of the high is

$$SAH(m) = \langle PSL(\phi, \lambda, m) \rangle_{\phi \in [-45°, -15°], \lambda \in [-20°, 10°]}, \tag{A37}$$

where $\langle \cdot \rangle$ denotes a spatial average.

To evaluate its seasonal migration, the latitude and longitude of the SAH center are defined as pressure-weighted means within the same domain:

$$\phi_{SAH}(m) = \frac{\sum_\phi \sum_\lambda \phi \, PSL(\phi, \lambda, m)}{\sum_\phi \sum_\lambda PSL(\phi, \lambda, m)}, \quad \lambda_{SAH}(m) = \frac{\sum_\phi \sum_\lambda \lambda \, PSL(\phi, \lambda, m)}{\sum_\phi \sum_\lambda PSL(\phi, \lambda, m)} \tag{A38}$$

These diagnostics provide measures of the climatological strength, latitudinal location, and longitudinal location of the SAH throughout the annual cycle, with evaluation focusing on mean biases relative to observations.

5.22 Natural Modes of Variability

We analyze model skill for four major modes of natural climate variability: the El Niño–Southern Oscillation (ENSO), the Indian Ocean Dipole (IOD), the North Atlantic Oscillation (NAO), and the Southern Annular Mode (SAM). For all indices, monthly anomalies are computed by removing the climatological monthly mean over a common reference period.

ENSO variability is represented using the Niño-3.4 index, defined as sea surface temperature (SST) anomalies averaged over the Niño-3.4 region (5°S–5°N, 170°W–120°W;) (Trenberth, 1997):

$$\mathrm{ENSO}(y, m) = X(y, m) - \overline{X}_m, \tag{A39}$$

where $X(y, m)$ is the regional monthly mean SST and $\overline{X}_m$ is the climatological mean for calendar month m.

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

The Indian Ocean Dipole is represented using the Dipole Mode Index (DMI), defined as the difference between SST anomalies in the western equatorial Indian Ocean (10°S–10°N, 50°E–70°E) and the southeastern equatorial Indian Ocean (10°S–0°, 90°E–100°E) (Saji et al., 1999):

$$\mathrm{IOD}(y,m) = \left[X_{\mathrm{W}}(y,m) - \overline{X}_{\mathrm{W},m}\right] - \left[X_{\mathrm{E}}(y,m) - \overline{X}_{\mathrm{E},m}\right]. \tag{A40}$$

The Southern Annular Mode is represented using a zonal-mean sea level pressure based index, defined as the difference between pressure anomalies at 40°S and 65°S (Marshall, 2003):

$$\mathrm{SAM}(y,m) = \left[X_{40}(y,m) - \overline{X}_{40,m}\right] - \left[X_{65}(y,m) - \overline{X}_{65,m}\right]. \tag{A41}$$

The North Atlantic Oscillation is represented using an EOF-based index derived from sea level pressure anomalies over the North Atlantic sector (Hurrell, 1995). The anomaly field is area weighted following standard practice prior to EOF analysis. The NAO index is defined as the principal component associated with the leading EOF mode:

$$\mathrm{NAO}(t) = \mathrm{PC}_1(t). \tag{A42}$$

Model evaluation for these indexes follows same methodology as that for amplitudes.

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

Code availability

The analysis code used in this study is currently being prepared for archiving in a public repository and will be made openly available upon publication of the manuscript.

Data availability

All data used in this study are publicly available. CMIP6 model output is available from the Earth System Grid Federation (ESGF, https://esgf-node.llnl.gov), and ERA5 reanalysis data are available from the Copernicus Climate Data Store (https://cds.climate.copernicus.eu). Supplementary figures (Figs. S1–S5) are archived in a public repository https://doi.org/10.5281/zenodo.19556090 (Ashfaq et al., 2026).

Author contribution

M. Ashfaq designed the study, performed the analysis, and wrote the manuscript. All co-authors contributed through discussion and provided feedback on the methodology, analysis, visualization, and manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

The authors acknowledge the World Climate Research Programme (WCRP) Coordinated Regional Climate Downscaling Experiment (CORDEX) Coordinated Output for Regional Evaluations (CORE) Task Force for their efforts in coordinating regional climate modeling activities that informed this study. This work is supported by the U.S. Air Force Numerical Weather Modeling Program and the National Climate-Computing Research Center, located within the National Center for Computational Sciences at Oak Ridge National Laboratory (ORNL), and supported under Strategic Partnership Project 2316-T849-08 between the U.S. Department of Energy (DOE) and NOAA. This research used resources of the Oak Ridge Leadership Computing Facility, a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725.

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

References

Andrews, M. B., Ridley, J. K., Wood, R. A., Andrews, T., Blockley, E. W., Booth, B., Burke, E., Dittus, A. J., Florek, P., Gray, L. J., Haddad, S., Hardiman, S. C., Hermanson, L., Hodson, D., Hogan, E., Jones, G. S., Knight, J. R., Kuhlbrodt, T., Misios, S., Mizielinski, M. S., Ringer, M. A., Robson, J., and Sutton, R. T.: Historical Simulations With HadGEM3-GC3.1 for CMIP6, J. Adv. Model. Earth Syst., 12, e2019MS001995, https://doi.org/10.1029/2019MS001995, 2020.

Ashfaq, M., Coppola, E., Lennard, C., Teichmann, C., Rastogi, D., Massoud, E., Buonomo, E., Im, E.-S., Zittis, G., Evans, J., Fernández, J., Evans, K., da Silva, M. L., Adinolfi, M., Bukovsky, M., da Rocha, R. P., hasson, . shabeh . ul ., Solman, S., Sobolowski, S., … Gao, X. (2026). Supplementary Figures for: A Multi-Criteria Framework for CORDEX-CORE2 GCM Selection. Zenodo. https://doi.org/10.5281/zenodo.19556090

Ashfaq, M., Rastogi, D., Kitson, J., Abid, M. A., and Kao, S. C.: Evaluation of CMIP6 GCMs over the CONUS for downscaling studies, J. Geophys. Res. Atmos., 127, e2022JD036659, https://doi.org/10.1029/2022JD036659, 2022.

Arias, P. A., Correa, I. C., Fita, L., et al.: How well CMIP6 models simulate key boundary conditions affecting South American climate? Insights for regional modeling efforts, Clim. Dyn., 63, 231, https://doi.org/10.1007/s00382-025-07704-w, 2025.

Barimalala, R., Blamey, R. C., Desbiolles, F., and Reason, C. J. C.: Variability in the Mozambique Channel Trough and impacts on southeast African rainfall, Journal of Climate, 33, 811–829, https://doi.org/10.1175/JCLI-D-19-0267.1, 2020.

Bethke, I., Wang, Y., Counillon, F., Keenlyside, N., Kimmritz, M., Fransner, F., Samuelsen, A., Langehaug, H., Svendsen, L., Chiu, P.-G., Passos, L., Bentsen, M., Guo, C., Gupta, A., Tjiputra, J., Kirkevåg, A., Olivié, D., Seland, Ø., Solsvik Vågane, J., Fan, Y., and Eldevik, T.: NorCPM1 and its contribution to CMIP6 DCPP, Geosci. Model Dev., 14, 7073–7116, https://doi.org/10.5194/gmd-14-7073-2021, 2021.

Biasutti, M., Sobel, A. H., and Kushnir, Y.: AGCM precipitation biases in the tropical Atlantic, Journal of Climate, 19, 935–958, https://doi.org/10.1175/JCLI3673.1, 2006.

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., Davini, P., de Lavergne, C., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J.-L., Dupont, E., Éthé, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M.-A., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J.-Y., Guenet, B., Guez, L. E., Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luyssaert, S., Madec, G., Madeleine, J.-B., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rochetin, N., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thiéblemont, R., Traore, A. K., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N., and Vuichard,

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

N.: Presentation and Evaluation of the IPSL-CM6A-LR Climate Model, J. Adv. Model. Earth Syst., 12, e2019MS002010, https://doi.org/10.1029/2019MS002010, 2020.

Brands, S.: A circulation-based performance atlas of the CMIP5 and CMIP6 models for regional climate studies in the Northern Hemisphere mid-to-high latitudes, Geosci. Model Dev., 15, 1375–1411, https://doi.org/10.5194/gmd-15-1375-2022, 2022a.

Brands, S.: Common error patterns in the regional atmospheric circulation simulated by the CMIP multi-model ensemble, Geophys. Res. Lett., 49, e2022GL101446, https://doi.org/10.1029/2022GL101446, 2022b.

Brient, F.: Reducing uncertainties in climate projections with emergent constraints: Concepts, examples and prospects, Adv. Atmos. Sci., 37, 1–15, https://doi.org/10.1007/s00376-019-9140-8, 2020.

Byun, Y.-H., Lim, Y.-J., Sung, H. M., Kim, J., Sun, M., and Kim, B.-H.: NIMS-KMA KACE1.0-G model output prepared for CMIP6 CMIP historical, Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.8378, 2019.

Carvalho, L. M. V., Jones, C., and Liebmann, B.: The South Atlantic convergence zone: Intensity, form, persistence, and relationships with intraseasonal to interannual activity and extreme rainfall, Journal of Climate, 17, 88–108, 2004.

Cao, J., Wang, B., Yang, Y.-M., Ma, L., Li, J., Sun, B., Bao, Y., He, J., Zhou, X., and Wu, L.: The NUIST Earth System Model (NESM) version 3: description and preliminary evaluation, Geosci. Model Dev., 11, 2975–2993, https://doi.org/10.5194/gmd-11-2975-2018, 2018.

Cherchi, A., Fogli, P. G., Lovato, T., Peano, D., Iovino, D., Gualdi, S., Masina, S., Scoccimarro, E., Materia, S., Bellucci, A., and Navarra, A.: Global Mean Climate and Main Patterns of Variability in the CMCC-CM2 Coupled Model, J. Adv. Model. Earth Syst., 11, 185–209, https://doi.org/10.1029/2018MS001369, 2019.

Cook, K. H.: Generation of the African easterly jet and its role in determining West African precipitation, Journal of Climate, 12, 1165–1184, 1999.

Coppola, E., Raffaele, F., Giorgi, F., Giuliani, G., Gao, X., Ciarlo, J. M., et al.: Climate hazard indices projections based on CORDEX-CORE, CMIP5 and CMIP6 ensemble, Clim. Dyn., 57, 1293–1383, https://doi.org/10.1007/s00382-021-05640-z, 2021.

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, W. H., Mills, M. J., Neale, R., Oleson, K. W., Otto-Bliesner, B., Phillips, A. S., Sacks, W., Tilmes, S., van Kampenhout, L., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J. E., Kinnison, D., Kushner, P. J., Larson, V. E., Long, M. C., Mickelson, S., Moore, J. K., Nienhouse, E., Polvani, L., Rasch, P. J., and Strand, W. G.: The Community Earth System Model

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

Version 2 (CESM2), J. Adv. Model. Earth Syst., 12, e2019MS001916, https://doi.org/10.1029/2019MS001916, 2020.

Desmet, Q. and Ngo-Duc, T.: A novel method for ranking CMIP6 global climate models over the Southeast Asian region, Int. J. Climatol., 42, 97–117, https://doi.org/10.1002/joc.7234, 2022.

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.

Dix, M., Bi, D., Dobrohotoff, P., Fiedler, R., Harman, I., Law, R., Mackallah, C., Marsland, S., O'Farrell, S., Rashid, H., Srbinovsky, J., Sullivan, A., Trenham, C., Vohralik, P., Watterson, I., Williams, G., Woodhouse, M., Bodman, R., Dias, F. B., Domingues, C., Hannah, N., Heerdegen, A., Savita, A., Wales, S., Allen, C., Druken, K., Evans, B., Richards, C., Ridzwan, S. M., Roberts, D., Smillie, J., Snow, K., Ward, M., and Yang, R.: CSIRO-ARCCSS ACCESS-CM2 model output prepared for CMIP6 CMIP historical, Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.4271, 2019.

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arsouze, T., Bergman, T., Bernardello, R., Boussetta, S., Caron, L.-P., Carver, G., Castrillo, M., Catalano, F., Cvijanovic, I., Davini, P., Dekker, E., Doblas-Reyes, F. J., Docquier, D., Echevarria, P., Fladrich, U., Fuentes-Franco, R., Gröger, M., v. Hardenberg, J., Hieronymus, J., Karami, M. P., Keskinen, J.-P., Koenigk, T., Makkonen, R., Massonnet, F., Ménégoz, M., Miller, P. A., Moreno-Chamarro, E., Nieradzik, L., van Noije, T., Nolan, P., O'Donnell, D., Ollinaho, P., van den Oord, G., Ortega, P., Prims, O. T., Ramos, A., Reerink, T., Rousset, C., Ruprich-Robert, Y., Le Sager, P., Schmith, T., Schrödner, R., Serva, F., Sicardi, V., Sloth Madsen, M., Smith, B., Tian, T., Tourigny, E., Uotila, P., Vancoppenolle, M., Wang, S., Wårlind, D., Willén, U., Wyser, K., Yang, S., Yepes-Arbós, X., and Zhang, Q.: The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6, Geosci. Model Dev., 15, 2973–3020, https://doi.org/10.5194/gmd-15-2973-2022, 2022.

Feng, X., Porporato, A., and Rodriguez-Iturbe, I.: Changes in rainfall seasonality in the tropics, Nature Climate Change, 3, 811–815, https://doi.org/10.1038/nclimate1907, 2013.

Guo, Q.-Y.: The summer monsoon intensity index in East Asia and its variation, Acta Geographica Sinica, 38, 207–217, 1983.

He, B., Yu, Y., Bao, Q., Lin, P., Liu, H., Li, J., Wang, L., Liu, Y., Wu, G., Chen, K., Guo, Y., Zhao, S., Zhang, X., Song, M., and Xie, J.: CAS FGOALS-f3-L model dataset descriptions for CMIP6 DECK experiments, Atmos. Ocean. Sci. Lett., 13, 582–588, https://doi.org/10.1080/16742834.2020.1778419, 2020.

Held, I. M., Guo, H., Adcroft, A., Dunne, J. P., Horowitz, L. W., Krasting, J., Shevliakova, E., Winton, M., Zhao, M., Bushuk, M., Wittenberg, A. T., Wyman, B., Xiang, B., Zhang, R., Anderson, W., Balaji, V., Donner, L., Dunne, K., Durachta, J., Gauthier, P. P. G., Ginoux, P., Golaz, J.-C., Griffies, S. M., Hallberg, R., Harris, L., Harrison, M., Hurlin, W., John, J., Lin, P., Lin, S.-J., Malyshev, S., Menzel, R., Milly, P. C. D., Ming, Y., Naik, V., Paynter, D., Paulot, F.,

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

Rammaswamy, V., Reichl, B., Robinson, T., Rosati, A., Seman, C., Silvers, L. G., Underwood, S., and Zadeh, N.: Structure and Performance of GFDL's CM4.0 Climate Model, J. Adv. Model. Earth Syst., 11, 3691–3727, https://doi.org/10.1029/2019MS001829, 2019.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., and Simmons, A.: The ERA5 global reanalysis, Q. J. R. Meteorol. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Holtanová, E., Belda, M., Crespo, N. M., et al.: On the relation of CMIP6 GCM errors at RCM driving boundary condition zones and inner region for the Central Europe region, Clim. Dyn., 62, 6433–6446, https://doi.org/10.1007/s00382-024-07216-z, 2024.

Hovmöller, E.: The trough-and-ridge diagram, Tellus, 1, 62–66, https://doi.org/10.1111/j.2153-3490.1949.tb01260.x, 1949.

Hurrell, J. W.: Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation, Science, 269, 676–679, https://doi.org/10.1126/science.269.5224.676, 1995.

Knutti, R., Masson, D., and Gettelman, A.: Climate model genealogy: Generation CMIP5 and how we got there, Geophys. Res. Lett., 40, 1194–1199, https://doi.org/10.1002/grl.50256, 2013.

Lin, Y., Huang, X., Liang, Y., Qin, Y., Xu, S., Huang, W., Xu, F., Liu, L., Wang, Y., Peng, Y., Wang, L., Xue, W., Fu, H., Zhang, G. J., Wang, B., Li, R., Zhang, C., Lu, H., Yang, K., Luo, Y., Bai, Y., Song, Z., Wang, M., Zhao, W., Zhang, F., Xu, J., Zhao, X., Lu, C., Chen, Y., Luo, Y., Hu, Y., Tang, Q., Chen, D., Yang, G., and Gong, P.: Community Integrated Earth System Model (CIESM): Description and Evaluation, J. Adv. Model. Earth Syst., 12, e2019MS002036, https://doi.org/10.1029/2019MS002036, 2020.

Longmate, J. M., Risser, M. D., and Feldman, D. R.: Prioritizing the selection of CMIP6 model ensemble members for downscaling projections of CONUS temperature and precipitation, Clim. Dyn., 61, 5171–5197, https://doi.org/10.1007/s00382-023-06846-z, 2023.

Lu, R.: Indices of the summertime western North Pacific subtropical high, Advances in Atmospheric Sciences, 19, 1004–1028, 2002.

Magaña, V., Amador, J. A., and Medina, S.: The midsummer drought over Mexico and Central America, Journal of Climate, 12, 1577–1588, https://doi.org/10.1175/1520-0442(1999)012<1577:TMDOMA>2.0.CO;2, 1999.

Marshall, G. J.: Trends in the Southern Annular Mode from observations and reanalyses, Journal of Climate, 16, 4134–4143, https://doi.org/10.1175/1520-0442(2003)016<4134:TITSAM>2.0.CO;2, 2003.

McSweeney, C. F., Jones, R. G., and Booth, B. B. B.: Selecting ensemble members to provide regional climate change information, J. Climate, 25, 7100–7121, https://doi.org/10.1175/JCLI-D-11-00526.1, 2012.

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

McSweeney, C. F., Jones, R. G., Lee, R. W., and Rowell, D. P.: Selecting CMIP5 GCMs for downscaling over multiple regions, Clim. Dyn., 44, 3237–3260, https://doi.org/10.1007/s00382-014-2418-8, 2015.

Merrifield, A. L., Brunner,L., Lorenz, R., Humphrey, V., and Knutti, R.: Climate model selection by independence, performance, and spread (ClimSIPS v1.0.1) for regional applications, Geosci. Model Dev., 16, 4715–4747, https://doi.org/10.5194/gmd-16-4715-2023, 2023.

Miller, R. L., Schmidt, G. A., Nazarenko, L. S., Bauer, S. E., Kelley, M., Ruedy, R., Russell, G. L., Ackerman, A. S., Aleinov, I., Bauer, M., Bleck, R., Canuto, V., Cesana, G., Cheng, Y., Clune, T. L., Cook, B. I., Cruz, C. A., Del Genio, A. D., Elsaesser, G. S., Faluvegi, G., Kiang, N. Y., Kim, D., Lacis, A. A., Leboissetier, A., LeGrande, A. N., Lo, K. K., Marshall, J., Matthews, E. E., McDermid, S., Mezuman, K., Murray, L. T., Oinas, V., Orbe, C., Pérez García-Pando, C., Perlwitz, J. P., Puma, M. J., Rind, D., Romanou, A., Shindell, D. T., Sun, S., Tausnev, N., Tsigaridis, K., Tselioudis, G.,

Morioka, Y., Tozuka, T., and Yamagata, T.: Local SST impacts on the summertime Mascarene High variability, Journal of Climate, 25, 6787–6804, https://doi.org/10.1175/JCLI-D-11-00614.1, 2012.

Mulcahy, J. P., Jones, C. G., Rumbold, S. T., Kuhlbrodt, T., Dittus, A. J., Blockley, E. W., Yool, A., Walton, J., Hardacre, C., Andrews, T., Bodas-Salcedo, A., Stringer, M., de Mora, L., Harris, P., Hill, R., Kelley, D., Robertson, E., and Tang, Y.: UKESM1.1: development and evaluation of an updated configuration of the UK Earth System Model, Geosci. Model Dev., 16, 1569–1600, https://doi.org/10.5194/gmd-16-1569-2023, 2023.

Neubauer, D., Ferrachat, S., Siegenthaler-Le Drian, C., Stoll, J., Folini, D. S., Tegen, I., Wieners, K.-H., Mauritsen, T., Stemmler, I., Barthel, S., Bey, I., Daskalakis, N., Heinold, B., Kokkola, H., Partridge, D., Rast, S., Schmidt, H., Schutgens, N., Stanelle, T., Stier, P., Watson-Parris, D., and Lohmann, U.: HAMMOZ-Consortium MPI-ESM1.2-HAM model output prepared for CMIP6 AerChemMIP, Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.1621, 2019.

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 for CORDEX dynamical downscaling over Southeast Asia using a standardized benchmarking framework, Geosci. Model Dev., 17, 7285–7315, https://doi.org/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, https://doi.org/10.5194/gmd-9-3461-2016, 2016.

Park, S., Shin, J., Kim, S., Oh, E., and Kim, Y.: Global climate simulated by the Seoul National University atmosphere model version 0 with a unified convection scheme (SAM0-UNICON), J. Climate, 32, 2917–2949, https://doi.org/10.1175/JCLI-D-18-0796.1, 2019.

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

Parding, K. M., Dobler, A., McSweeney, C. F., Landgren, O. A., Benestad, R., Erlandsen, H. B., Mezghani, A., Gregow, H., Räty, O., Viktor, E., and El Zohbi, J.: GCMeval – an interactive tool for evaluation and selection of climate model ensembles, Clim. Serv., 18, 100167, https://doi.org/10.1016/j.cliser.2020.100167, 2020.

Pendergrass, A. G., Byrne, M. P., Watt-Meyer, O., Maher, P., and Webb, M. J.: Impact of ITCZ width on global climate: ITCZ-MIP, Geosci. Model Dev., 17, 6365–6378, https://doi.org/10.5194/gmd-17-6365-2024, 2024.

Prein, A. F., Gobiet, A., Truhetz, H., Keuler, K., Goergen, K., Teichmann, C., et al.: Precipitation in the EURO-CORDEX 0.11° and 0.44° simulations: High resolution, high benefits?, Clim. Dyn., 46, 383–412, https://doi.org/10.1007/s00382-015-2589-y, 2016.

Raju, K. S. and Kumar, D. N.: Review of approaches for selection and ensembling of GCMs, J. Water Clim. Change, 11, 577–599, https://doi.org/10.2166/wcc.2019.197, 2020.

Rampal, N., Hobeichi, S., Gibson, P. B., Baño-Medina, J., Abramowitz, G., Beucler, T., González-Abad, J., Chapman, W., Harder, P., and Gutiérrez, J. M.: Enhancing regional climate downscaling through advances in machine learning, Artif. Intell. Earth Syst., 3, 230066, https://doi.org/10.1175/AIES-D-23-0066.1, 2024.

Rong, X.: CAMS CAMS_CSM1.0 model output prepared for CMIP6 CMIP historical, Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.9754, 2019.

Sanderson, B. M., Wehner, M., and Knutti, R.: Skill and independence weighting for multi-model assessments, Geosci. Model Dev., 10, 2379–2395, https://doi.org/10.5194/gmd-10-2379-2017, 2017.

Saji, N. H., Goswami, B. N., Vinayachandran, P. N., and Yamagata, T.: A dipole mode in the tropical Indian Ocean, Nature, 401, 360–363, https://doi.org/10.1038/43854, 1999.

Seland, Ø., Bentsen, M., Olivié, D., Toniazzo, T., Gjermundsen, A., Graff, L. S., Debernard, J. B., Gupta, A. K., He, Y.-C., Kirkevåg, A., Schwinger, J., Tjiputra, J., Aas, K. S., Bethke, I., Fan, Y., Griesfeller, J., Grini, A., Guo, C., Ilicak, M., Karset, I. H. H., Landgren, O., Liakka, J., Moseid, K. O., Nummelin, A., Spensberger, C., Tang, H., Zhang, Z., Heinze, C., Iversen, T., and Schulz, M.: Overview of the Norwegian Earth System Model (NorESM2) and key climate response of CMIP6 DECK, historical, and scenario simulations, Geosci. Model Dev., 13, 6165–6200, https://doi.org/10.5194/gmd-13-6165-2020, 2020.

Semmler, T., Danilov, S., Gierz, P., Goessling, H. F., Hegewald, J., Hinrichs, C., Koldunov, N., Khosravi, N., Mu, L., Rackow, T., Sein, D. V., Sidorenko, D., Wang, Q., and Jung, T.: Simulations for CMIP6 With the AWI Climate Model AWI-CM-1-1, J. Adv. Model. Earth Syst., 12, e2019MS002009, https://doi.org/10.1029/2019MS002009, 2020.

Shepherd, T. G.: Atmospheric circulation as a source of uncertainty in climate change projections, Nat. Geosci., 7, 703–708, https://doi.org/10.1038/ngeo2253, 2014.

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

Shiogama, H., Watanabe, M., Kim, H., et al.: Emergent constraints on future precipitation changes, Nature, 602, 612–616, https://doi.org/10.1038/s41586-021-04310-8, 2022.

Sidorenko, D., Rackow, T., Jung, T., Semmler, T., Barbi, D., Danilov, S., Dethloff, K., Dorn, W., Fieg, K., Goessling, H. F., Handorf, D., Harig, S., Hiller, W., Juricke, S., Losch, M., Schröter, J., Sein, D. V., and Wang, Q.: Towards multi-resolution global climate modeling with ECHAM6–FESOM. Part I: model formulation and mean climate, Clim. Dyn., 44, 757–780, https://doi.org/10.1007/s00382-014-2290-6, 2015.

Sobolowski, S., Somot, S., Fernandez, J., et al.: GCM selection and ensemble design: Best practices and recommendations from the EURO-CORDEX community, Bull. Am. Meteorol. Soc., 106, E1834–E1850, https://doi.org/10.1175/BAMS-D-23-0189.1, 2025.

Swart, N. C., Cole, J. N., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Solheim, L., von Salzen, K., Yang, D., Winter, B., and Sigmond, M.: CCCma CanESM5 model output prepared for CMIP6 CMIP historical, Earth System Grid Federation [data set], https://doi.org/10.22033/ESGF/CMIP6.3610, 2019.

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., Sudo, K., Sekiguchi, M., Abe, M., Saito, F., Chikira, M., Watanabe, S., Mori, M., Hirota, N., Kawatani, Y., Mochizuki, T., Yoshimura, K., Takata, K., O'ishi, R., Yamazaki, D., Suzuki, T., Kurogi, M., Kataoka, T., Watanabe, M., and Kimoto, M.: Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6, Geosci. Model Dev., 12, 2727–2765, https://doi.org/10.5194/gmd-12-2727-2019, 2019.

Toolan, C. A., Amooli, J. A., Wilcox, L. J., Samset, B. H., Turner, A. G., and Westervelt, D. M.: Strong intermodel differences and biases in CMIP6 simulations of PM2.5, aerosol optical depth, and precipitation over Africa, Atmos. Chem. Phys., 25, 10523–10557, https://doi.org/10.5194/acp-25-10523-2025, 2025.

Trenberth, K. E.: The definition of El Niño, Bulletin of the American Meteorological Society, 78, 2771–2777, https://doi.org/10.1175/1520-0477(1997)078<2771:TDOENO>2.0.C

Voldoire, A., Saint-Martin, D., Sénési, S., Decharme, B., Alias, A., Chevallier, M., Colin, J., Guérémy, J.-F., Michou, M., Moine, M.-P., Nabat, P., Roehrig, R., Salas y Mélia, D., Séférian, R., Valcke, S., Beau, I., Belamari, S., Berthet, S., Cassou, C., Cattiaux, J., Deshayes, J., Douville, H., Ethé, C., Franchistéguy, L., Geoffroy, O., Lévy, C., Madec, G., Meurdesoif, Y., Msadek, R., Ribes, A., Sanchez-Gomez, E., Terray, L., and Waldman, R.: Evaluation of CMIP6 DECK Experiments With CNRM-CM6-1, J. Adv. Model. Earth Syst., 11, 2177–2213, https://doi.org/10.1029/2019MS001683, 2019.

Volodin, E. M., Mortikov, E. V., Kostrykin, S. V., Galin, V. Y., Lykossov, V. N., Gritsun, A. S., Diansky, N. A., Gusev, A. V., and Iakovlev, N. G.: Simulation of the present-day climate with the climate model INMCM5, Clim. Dyn., 49, 3715–3734, https://doi.org/10.1007/s00382-017-3539-7, 2017.

https://doi.org/10.5194/egusphere-2026-2649
Preprint. Discussion started: 21 May 2026

Wang, C.: Variability of the Caribbean Low-Level Jet and its relations to climate, Climate Dynamics, 29, 411–422, https://doi.org/10.1007/s00382-007-0243-z, 2007.

Wang, C., Lee, S.-K., and Enfield, D. B.: Origins of the Caribbean rainfall bimodal behavior, Journal of Geophysical Research: Atmospheres, 115, D11101, https://doi.org/10.1029/2009JD012990, 2010.

Weng, E., Wu, J., and Yao, M.-S.: CMIP6 Historical Simulations (1850–2014) With GISS-E2.1, J. Adv. Model. Earth Syst., 13, e2019MS002034, https://doi.org/10.1029/2019MS002034, 2021.

Wu, J., Gao, X. J., Shi, Y., and Giorgi, F.: Projection of future changes in tropical cyclone activity over the western North Pacific based on multi-RegCM4 simulations, Adv. Atmos. Sci., 39, 284–303, https://doi.org/10.1007/s00376-021-0286-9, 2022.

Wu, T., Lu, Y., Fang, Y., Xin, X., Li, L., Li, W., Jie, W., Zhang, J., Liu, Y., Zhang, L., Zhang, F., Zhang, Y., Wu, F., Li, J., Chu, M., Wang, Z., Shi, X., Liu, X., Wei, M., Huang, A., Zhang, Y., and Liu, X.: The Beijing Climate Center Climate System Model (BCC-CSM): the main progress from CMIP5 to CMIP6, Geosci. Model Dev., 12, 1573–1600, https://doi.org/10.5194/gmd-12-1573-2019, 2019.

Wu, T., Zhang, F., Zhang, J., Jie, W., Zhang, Y., Wu, F., Li, L., Yan, J., Liu, X., Lu, X., Tan, H., Zhang, L., Wang, J., and Hu, A.: Beijing Climate Center Earth System Model version 1 (BCC-ESM1): model description and evaluation of aerosol simulations, Geosci. Model Dev., 13, 977–1005, https://doi.org/10.5194/gmd-13-977-2020, 2020.

Yakubu, F., Böhner, J., Schickhoff, U., et al.: Global bias-corrected CORDEX datasets at half-degree resolution, Sci. Data, 12, 1781, https://doi.org/10.1038/s41597-025-06200-4, 2025.

Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., Yabu, S., Yoshimura, H., Shindo, E., Mizuta, R., Obata, A., Adachi, Y., and Ishii, M.: The Meteorological Research Institute Earth System Model version 2.0, MRI-ESM2.0: Description and basic evaluation of the physical component, J. Meteorol. Soc. Jpn., 97, 931–965, https://doi.org/10.2151/jmsj.2019-051, 2019.

Zhou, T., Zhang, J., He, J., and Li, B.: Why the western Pacific subtropical high has extended westward since the late 1970s, Journal of Climate, 22, 2199–2215, 2009.

Ziehn, T., Chamberlain, M. A., Law, R. M., Lenton, A., Bodman, R. W., Dix, M., Stevens, L., Wang, Y.-P., and Srbinovsky, J.: The Australian Earth System Model: ACCESS-ESM1.5, J. South. Hemisph. Earth Syst. Sci., 70, 193–214, https://doi.org/10.1071/ES19035, 2020.