https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

Geoscientific Model Development
Discussions

Evaluation of atmospheric rivers in reanalyses and climate models in a new metrics framework

Bo Dong[1], Paul Ullrich[1], Jiwoo Lee[1], Peter Gleckler[1], Kristin Chang[1], Travis A. O'Brien[2,3]

1. Lawrence Livermore National Laboratory, Livermore, CA, USA
2. Department of Earth and Atmospheric Sciences, Indiana University, Bloomington, IN, USA
3. Climate and Ecosystem Sciences Division, Lawrence Berkeley National Lab, Berkeley, CA, USA

Correspondence to: Bo Dong (dong12@llnl.gov)

Key points:

1. A metrics package designed for easy analysis of AR characteristics and statistics is presented
2. The tool is efficient for diagnosing systematic AR bias in climate models, and useful for evaluating new AR characteristics in model simulations
3. In climate models, landfalling AR precipitation shows dry biases globally, and AR tracks are farther poleward (equatorward) in the north and south Atlantic (south Pacific and Indian Ocean)

Abstract

We present a suite of new atmospheric river (AR) metrics that are designed for quick analysis of AR characteristics and statistics in gridded climate datasets such as model output and reanalysis. This package is expected to be particularly useful for climate model evaluation. The metrics include mean bias and spatial pattern correlation, which are efficient for diagnosing systematic AR biases in climate models. For example, the package identifies that in CMIP5 and CMIP6 models, AR tracks in the south Atlantic are positioned farther poleward compared to the ERA5 reanalysis, while in the south Pacific, tracks are generally biased towards the equator. For the landfalling AR peak season, we find that most climate models simulate a completely opposite seasonal cycle over western Africa. This tool is also useful for identifying and characterizing structural differences among different AR detectors (ARDTs). For example, ARs detected with the Mundhenk algorithm exhibit systematically larger size, width and length compared to the TempestExtremes (TE) method. The AR metrics developed

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

Geoscientific Model Development Discussions

from this work can be routinely applied for model benchmarking and during the development cycle to trace performance evolution across model versions or generations and set objective targets for the improvement of models. They can also be used by operational centers to perform near real-time climate and extreme events impact assessment as part of their forecast cycle.

1. Introduction

Atmospheric rivers (ARs) are dynamically driven, synoptic-scale filamentary structures of water vapor jets that play important roles in the global water cycle and regional weather and hydrology (Ralph et al. 2013; Gimeno et al. 2014; Shields et al. 2019; Payne et al. 2020; O'Brien et al., 2022). These narrow, concentrated corridors of moisture in the atmosphere can carry an immense amount of water, often compared to the flow of multiple major rivers combined (Ralph and Dettinger, 2011), and account for a substantial portion, more than 90% of the poleward water vapor transport (Zhu and Newell, 1998; Newman et al. 2012; Ullrich et al. 2021). When approaching landmasses or interacting with mountainous regions, ARs usually bring extreme weather inland, such as heavy rainfall and strong wind, leading to severe flooding and landslides, causing devastating damages to natural landscapes, agricultural fields, infrastructure, human settlements, and disruption to businesses and services with significant economic losses (Ralph et al., 2006; Leung and Qian, 2009; Neiman et al., 2011; Neiman et al., 2013; Gershunov et al., 2017).

Previous studies have developed numerical algorithms for objective identification of ARs (e.g., Neiman et al., 2009; Dettinger, 2011; Ralph et al., 2013; Mundhenk et al. 2016; Ullrich and Zarzycki 2017; Ullrich et al., 2021). As noted by O'Brien et al. (2022), the different choices made by ARDT developers essentially amount to different definitions of ARs, all of which are qualitatively consistent with the definition in the AMS glossary (Ralph et al., 2018). ARDTs are generally threshold-based, mostly using the intensity of moisture transport with some geographical constraints that limit the AR spatial extent and some geometrical constraints that preserve their nature as "long and narrow" filaments of moisture. For example, the Mundhenk algorithm (Mundhenk et al. 2016) calculates integrated water vapor transport (IVT) anomalies relative to the historical period and uses a fixed relative threshold to identify ARs that are above a certain percentile of the historical simulation. The TempestExtremes (TE; Ullrich et al. 2021) method, as another example, uses relative threshold on the Laplacian of the IVT field rather than the IVT field itself. Although AR detectors (ARDTs) are usually designed with particular research questions in mind, they have widely facilitated broader studies

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

Geoscientific Model Development Discussions

of AR characteristics and impacts (Shields et al., 2018; Rutz et al., 2019; O'Brien et al., 2022).

The number of climate models under active development and used in the research community has increased substantially in recent decades, with many supporting multiple configurations and parameterization choices. Meanwhile, newer versions of ARDTs have been developed , along with newer observational data products. As such, routine evaluation of ARs during model development lifecycles requires a quantitative climate data assessment evaluation workflow that is independent of ARDT and that allows comparing AR characteristics from different ARDTs. We believe progress in improving our understanding of ARs and their impacts could be accelerated with a dedicated tool for calculating AR statistics in climate models and gridded data products.

Metrics have been widely used to quantify climate model performance in recent decades (Taylor 2001; Gleckler et al. 2008; Wilks 2011; Zarzycki et al. 2021). Similarly, a set of common metrics are also increasingly employed in AR studies over the past few years, such as mean bias (Guan and Waliser 2017; Chapman et al. 2019), weighted ensemble mean bias (Massoud et al. 2019), RMS error and relative RMS error (Guan and Waliser 2017), spatial pattern correlation (Chapman et al. 2019; Huang et al. 2021), ratio of spatial standard deviation (O'Brien et al. 2022), and skill scores for assessing AR predictions (Wick et al. 2013, Nardi et al. 2018) and model performance (Zhang et al. 2024). While these quantitative measures are case-specific and depend on the aim of these studies, there is value in synthesizing commonly used metrics in one comprehensive analysis tool.

In this paper, we propose a set of metrics that is designed for easy quantification of AR characteristics and statistics in all types of gridded climate data, with the expectation that such a metric suite would be useful for climate model evaluation. Following the introduction, section 2 describes the general design of the AR metrics. Section 3 presents several example model evaluation applications of using the metrics evaluation package. Conclusions and discussion are in section 4.

2. Data and method

2.1 Input data

The input data to the metrics package includes AR "tags" and optional climate variables of interest that are concurrent with AR activities, such as precipitation, winds, and temperature (Fig. 1). The AR tags can be products of any regional or global AR detector

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

Geoscientific Model Development
Discussions

(ARDT), including those based on relative (e.g., TempestExtremes or TE; Ullrich and Zarzycki 2017; Ullrich et al. 2021), fixed-relative (e.g., Mundhenk_v3; Mundhenk et al. 2016), and absolute (e.g., Lora_v2; Lora et al. 2017) thresholds to the moisture field.

For applications in section 3, we run and compare the TE ARDT on the 6-hourly integrated water vapor transport (IVT) data from three reanalysis products - ERA5 (Hersbach et al. 2020), MERRA-2 (Gelaro et al. 2017) and JRA-55C (Japan Meteorological Agency, Japan 2015) to obtain AR tags for reanalyses. Given its longer data record and finer model resolution, we subsequently use ERA5 as the default reference in this study. To demonstrate how results are sensitive to the choice of ARDTs, we then use the Mundhenk_v3 tags from ERA5 data.

To evaluate ARs in climate models, we use the archived AR tags from the Atmospheric River Tracking Method Intercomparison Project (ARTMIP) Tier 2 experiment, which is based on the coupled CMIP model simulations for the historical and 21st century projection periods. (Shield et al. 2019, Rutz et al 2019, O'Brien et al, 2022). The tag data include six of the CMIP5 models (CCSM4, CSIRO-Mk3-6, CanESM2, IPSL-CM5A-LR, IPSL- CM5B-L, and NorESM1-M) and 3 of the CMIP6 models (BCC-CSM2-MR, IPSL-CM6A-LR, MRI-ESM2-0). For model evaluation purposes in our application examples, only TE tags from the archive are selected.

We further use simulations from the Energy Exascale Earth System Model (E3SM; Golaz et al. 2019, Caldwell et al. 2019) high resolution (HR, 0.25°, ~28 km grid) and low resolution (LR, 1°, ~111 km grid) experiments to examine the sensitivity of ARs to model resolution. Except for their different horizontal grid spacing, both E3SM-HR and E3SM-LR use an identical set of physical parameters, and the simulations follow a similar protocol of the Coupled Model Intercomparison Project Phase 6 (CMIP6; Eyring et al. 2016).

For the evaluation of AR characteristics, statistics gauging the consistency of latitude, longitude, width, length, and size are required as the input for metrics. In our case, we use the 'BlobStats' tool (Ullrich et al. 2021) to calculate the statistics, where latitude and longitude are weighted by the moisture field, width and length are based on principle component analysis (PCA; Inda-Díaz et al. 2021), and size is based on a count of the number of contiguous grid cells in the feature. This tool can be called and run within the AR metrics workflow, with a separate installation. Users can also optionally use their preferred statistical package for AR geometry calculation and then feed the data back to the metrics workflow.

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

Geoscientific Model Development
Discussions
Open Access

2.2 Geographical Regions

In this tool package, the AR metrics are calculated based on the data in user-defined geographic domains. In Fig. 1, the upper right panel shows examples of regions that were selected for landfalling AR diagnostics (red boxes in the panel, lat-lon boundaries are listed in the supplementary table S1). These regions, mostly located in the west coast of continents, are known to have frequently observed AR landfalls (Guan and Waliser 2015, Algarra et al. 2020). We purposely use rectangular region boundaries for easy use of the metrics tool, such that rather than needing a regional mask file, users can quickly sub-select the data by declaring latitude and longitude bounds of any specific region. For AR statistics, we group global ARs into 5 major ocean basins – the North Pacific, South Pacific, North Atlantic, South Atlantic, and South Indian Ocean (blue boxes in Fig. 1 upper right panel; lat-lon coordinates in table S1 in the supplement).

2.3 Metrics

2.3.1 Mean bias

We use mean bias to measure how close a climate data product is with respect to the reference data, calculated as

$$\bar{b} = \bar{x} - \bar{y}$$

where $\bar{x}$ is the arithmetic mean of the test variable x with sample size *n*, given by

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

and similarly, the $\bar{y}$ is the arithmetic mean of the reference variable

The statistical significance of the mean bias is measured using the Z-test, with the test statistics (z-score) formulated as

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\bar{\mu}_1 - \bar{\mu}_2)}{\sqrt{\frac{s_1^2}{n_1 - 1} + \frac{s_2^2}{n_2 - 1}}}$$

where $\bar{x_i}$ is sample arithmetic mean, μ_i is population mean, s_i is sample variance, and n_i is sample size. A positive z-score indicates that the value is above the mean. The higher the z-score, the further above the mean the value is, and vice versa. A result is

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

Geoscientific Model Development
Discussions

considered statistically significant at the 95% confidence level if the magnitude of the z-score is greater than 1.96.

When comparing across different variables, a commonly used measure is the normalized bias, with the data normalized by the standard deviation of the reference field. In this study, we simply use z-score as the normalized bias, as it incorporates both bias and statistical significance in one succinct formula.

2.3.2 Spatial pattern similarity

The spatial pattern correlation is a measure used to quantify the similarity between two spatial fields without reflecting the magnitude of the difference. Here we compute the spatial pattern correlation using the Pearson correlation coefficient:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

where, x_i and y_i are the values of the two spatial patterns at location *i* (or grid point *i* in gridded data product), $\bar{x}$ and $\bar{y}$ are the means of the values of the two patterns, and *n* is the total number of locations. This equation essentially measures the degree to which the values of the two spatial patterns vary together. If they vary together perfectly, r will be 1. If they vary together inversely, r will be -1. If there's no linear relationship between the patterns, r will be 0.

The statistical significance of correlation is determined by the two-tailed p-value of the cumulative distribution function (CDF) of the t-statistic, as

$$p = 2 \times (1 - \mathrm{CDF}(t))$$

The the t-statistic *t* is given by

$$t = r \times \frac{\sqrt{n_e}}{\sqrt{1 - r^2}}$$

where r is the correlation coefficient, and n_e is the effective sample size. Although there are a number of methods to estimate the effective geographic sample size (e.g., Griffth 2013), given that ARs present notable seasonal and interannual latitudinal shift patterns, we propose a new method to estimate n_e as the number of Principal Component Analysis (PCA) modes required to explain more than 95% of the total

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

variance in the AR tag data. The cumulative variance explained by the principal components is expressed as

$$n_e = \min\left\{ n_e \mid \frac{\sum_{i=1}^{n_e} \lambda_i}{\sum_{i=1}^{p} \lambda_i} > 0.95 \right\}$$

where the λ_i are the eigenvalues of the spatial correlation matrix of the data, and p is the total number of principal components. Estimating n_e based on ERA5 reanalysis data, we find that the effective sample sizes for spatial pattern correlation are generally small, ranging from 14 - 27 for the 5 ocean basins (Table S2 in supplementary information).

2.3.3 Temporal detection similarity

The AR binary occurrence time series refers to a binary time series equal to one when an AR is present in a given region and zero otherwise. The overlap between two AR occurrence time series is measured by the Intersection over Union (IoU) metric. The metric is written as

$$IoU(A,B) = \frac{\sum |A \cap B|}{\sum |A \cup B|}$$

where, A and B are binary AR occurrence time series. The IoU is useful for gauging the degree of temporal similarity of ARs detected in different ARDTs.

2.3.4 Metrics and diagnostics implementation

The metrics and diagnostics are pre-defined in the metrics framework, and they are fully customizable. Table 1 lists all the AR metrics and diagnostics used in this study. The AR metrics are composed of AR properties (as shown in the top row) and evaluation metrics. Similarly, the AR diagnostics are composed of AR properties and statistical diagnostics. The number of regions that these metrics are applied to are indicated by the numbers in the table. The metrics code is python-based, and it handles gridded AR tag and climate data using xCDAT (Xarray Climate Data Analysis Tools, https://xcdat.readthedocs.io), which is an extension of the Xarray package (https://xarray.pydata.org).

3. Metrics applications

In this section, we present five example applications using the metrics tool for assessing ARs in climate models, including evaluation of AR frequency and characteristics,

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

Geoscientific Model Development
Discussions

comparison of ARs in high- and low-resolution simulations, sensitivity of ARs to choice of ARDT, precipitation bias associated with ARs and landfalling AR seasonality.

3.1 AR characteristics in CMIP5 and CMIP6 models

3.1.1 AR frequency

We first analyze the pattern of AR occurrence frequency over a 10-year period (1979-1988) for the five major ocean basins from section 2.2. From the spatial distribution of the AR frequency, we calculate the pattern correlation between selected climate models and ERA5. The spatial pattern correlation coefficient is shown in Fig. 2. Notably the correlations are statistically significant for all models and regions. This suggests that climatologically, climate models simulate AR density and spatial distribution that broadly resemble reanalysis on planetary scale. This is evidenced in the spatial AR occurrence density maps in Fig. 3 (a-b) and (d-e).

The high spatial correlation is mainly a result of the similar spatial gradient of the AR frequencies, rather than the similar magnitude of the frequency at each grid point in two datasets. For instance, if the AR frequency values in one map are doubled compared to those on the other map, the spatial patterns, or spatial structures of the two, can still be perfectly correlated. Since climatologically ARs are largely clustered along the storm track, with nearly no presence over a large portion of the basin domain, it is natural that the pattern correlations are significant in most cases. Similar high pattern correlations of AR frequencies are also noted in other studies (e.g., Huang et al. 2020; Guan et al. 2023). In other words, the spatial correlation coefficient is not that indicative for the magnitude resemblance of the AR spatial frequency. Therefore, these metric results can be better interpreted together with AR frequency maps.

While the spatial correlation coefficient synthesizes the level of pattern consistency, difference maps further reveal the spatial discrepancies. For example, Fig. 3c shows that South Pacific AR tracks shift farther towards the equator in the CSIRO model than in ERA5. While in the North Atlantic basin (Fig. 3f), AR tracks are displaced more poleward in the BCC model. The further north AR location is likely associated with the poleward jet stream bias in CMIP6 models (Bracegirdle et al. 2020; Harvey et al. 2020).

3.1.2 AR geometric features in major ocean basins

The portrait plots in Fig. 4 show normalized biases (as z-score) of AR characteristics in climate models for the 5 major ocean basins. Several striking results emerge. For instance, in the North Pacific, the CMIP5 and CMIP6 AR geometry, in terms of width and length, are significantly smaller than the ERA5 reanalysis. One possible cause of

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

such bias is that the AR blobs detected with TE in the relatively lower resolution climate models are geometrically less curvy, and less pointy at the ends. Fig. S1 shows an example time slice of AR blobs in the ERA5 and BCC model. It is clear that the highlighted AR blob in the BCC model exhibits a “cut-off” feature at both ends, thus shorter in length than the ERA5 reanalysis. And although visually the blob is wider, the PCA based width is actually narrower due to its less curvy blob geometry. In contrast, for all other ocean basins, the AR sizes (area) are generally bigger in climate models. The figures also show notable latitudinal model AR biases, such that compared to the reanalysis, ARs tend to shift towards higher latitudes in the North and South Atlantic and biased towards the equator in the South Pacific and Indian Ocean.

Fig. 4 also helps identify outliers of a specific model or variable. For example, although most climate models tend to simulate larger ARs than observed (indicated by the positive values in the area columns), one notable exception is the CanESM2 model which has significantly smaller AR width, length, and area than other models and ERA5 reanalysis. Taking a closer look into the AR width and length in the North Pacific in Fig. 5, we see that CanESM2 simulates more smaller ARs and fewer bigger ARs than the reanalysis, resulting in negative mean biases. This type of histogram helps us better understand the AR distribution discrepancies.

Another example is from the CCSM4 model simulations. The higher bounds of the model histogram in nearly all fields indicate that the CCSM4 model simulates more ARs than the reanalysis, with bigger size indicated as taller area bars in Fig. 5c. The higher ARs counts in the model are mostly located in the high latitudes and the tropics south of 20°N (Fig. 5a), spreading across all longitude (Fig. 5b). Fig. 5d and 5e show that the additional ARs in CCSM4 are narrower and/or longer in shape.

3.2 ARs in high and low resolution E3SM simulations

We now apply the metrics and diagnostics identified in section 2.3.4, including the mean bias of AR latitude, longitude, area, width and length over 5 ocean basins, and AR induced precipitation over 16 landfall regions, to evaluate and compare AR characteristics in the E3SM HR and LR simulations. ARs in both HR and LR exhibit similar structural differences compared to the ERA5 (Fig. 6a, b). They are bigger in terms of area, width, and length, and biased towards higher latitudes in the North Pacific and South Atlantic. Zonally, ARs in E3SM are more westward distributed in the North Pacific, and more eastward distributed in the North Atlantic and South Pacific. One difference we see between the two experiments is that in the North Atlantic basin, AR tracks in the HR are shifted more northward than in the LR simulation.

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

Geoscientific Model Development
Discussions

Figure 6c shows AR differences between E3SM HR and LR models. The most noticeable differences are that the HR simulates wider and longer ARs than the LR model over all ocean basins. The AR size, in the area column, however, shows mixed results which are not consistent with systematic biases in width and length. This is probably because of different AR geometric properties in the HR and LR simulations. For example, in Supplementary Figure S2, the highlighted AR blob in the North Atlantic is longer but smaller in the LR compared to the one in the HR simulation. Latitudinally, AR distributions show hemispheric contrast, as compared to the LR, ARs in HR are located more southward in the Pacific sector but more northward in the Atlantic sector.

Figure 7 shows AR characteristic distribution in the North Pacific for E3SM HR, LR and ERA5. Apparently, E3SM produces more AR events than the reanalysis in nearly all fields and across all scales. We also evaluated the precipitation associated with landfalling ARs in California in both HR and LR simulations, as in Fig. 8. It is notable that both models simulate systematically higher precipitation than ERA5 for all rainfall intensity categories. It is also clear that the precipitation bias in HR simulation is larger than LR simulation, except in the light rainfall (< ~6mm/day) category. Similarly, better topographic representation in high resolution version of the model does not improve precipitation simulation is also reported in Harrop et al. (2023), especially when the bias in the low resolution model is substantially high.

3.3. Sensitivity of AR characteristics to ARDT

In this application of the metrics package, we examine how ARs in ERA5 are sensitive to the choice of ARDT. In addition to TE-based AR tags, we use AR tags detected using the Mundhenk_v3 algorithm for comparison. Despite significant differences in their associated algorithms, results from ARTMIP showed their performance was similar and close to the mean among all ARDTs (Shields et al., 2018). Table 2 shows agreement of landfalling ARs detected using these two ARDTs, as % values of IoU (AR concurrence normalized by total occurrence of the ARs in both methods). The level of consistency ranges from 56% to 83%, which suggests that TE and Mundhenk detect ARs concurrently most of the time, but with asynchronous discrepancies, possibly at the timing of the landfall and the end of the AR life cycle.

For AR characteristics over the oceans, the Mundhenk method detects larger ARs in area, width, and length compared to TE (Fig. 9). ARs are also present at more northward latitudes with Mundhenk than TE. Zonally, AR distributions exhibit more hemispherical contrast, with Mundhenk showing more westward located ARs in the Pacific sector but more eastward located ARs in the Atlantic sector.

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

Geoscientific Model Development Discussions

3.4 Landfalling AR precipitation in CMIP5/6 models

Precipitation is an important indicator of the intensity of a landfalling AR. Here we evaluate landfalling AR precipitation in the CMIP5 and CMIP6 models, with the ERA5 reanalysis and MSWEP (Beck et al. 2017) gridded product as reference. Fig. 10 shows that compared to the observations, landfalling precipitation differences in the models are generally much larger than in reanalysis. The models show dry biases in most regions, particularly large in California, Pacific Northwest, Iceland and Greenland.

As it is unclear if these biases are mainly due to general precipitation biases, or AR activity bias, we further examine model precipitation bias diagnostics regardless of AR activity (Fig. 11a) and AR frequency bias metrics (Fig. 11b) separately. For total precipitation in the models, structural biases as in Fig. 10 are absent, but AR landfalls are less frequent in the Pacific Northwest, Iceland, and Greenland. This suggests that the systematic dry AR precipitation biases over these regions are primarily due to the insufficient number of landfalling ARs in the models. For California, similar results do not hold for all the models, for example, total precipitation in CCSM4 is higher than the reanalysis and AR landfalls are more frequent, but the AR-related rainfall has a significant dry bias. This suggests that landfalling ARs in CCSM4 are less intense, suggesting a potential direction for model improvement.

3.5 Landfalling AR peak day

3.5.1 Comparison among reanalyses

Seasonality of AR landfalls is one of the important metrics for understanding AR variability and impacts. Here we analyze landfalling AR seasonality over various regions of the globe among three reanalysis products. We perform a Fourier transform on the 10-year long-term daily mean AR histogram to find its peak date based on the phase of the first Fourier mode. Results indicate that the AR peak days agree well among reanalyses for most regions, with small differences of only a few days. Large discrepancies are noted for Australia and western Africa: In Australia, AR landfall peaks nearly a month behind in JRA-55C than MERRA-2, while in west Africa, AR landfall in MERRA-2 peaks 46 days behind ERA5.

Details of these differences are depicted in the histogram plots. For West Africa, AR landfalls have two peaks in ERA5 and MERRA-2, one being in September, followed by another peak in November. In ERA5, the peak in November is the main peak, while in MERRA-2, the September peak is comparable to the November peak, resulting in an earlier peak day from the Fourier phase spectrum. JRA-55C, in contrast, has only one

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

peak in November, and the AR landfall event counts are fewer than the other two products over the entire year, indicative of smaller year to year variability.

Seasonal distribution of AR landfalls in Australia in the three reanalyses exhibit similar differences to those in western Africa. In ERA5 and MERRA-2, there are two peaks in February and June, but only one peak presents in JRA-55C in June. This explains the relative late peak day in JRA-55C. While the main peak in ERA5 is in June, in MERRA-2, the main peak is in February, which is consistent with the metrics result that MERRA-2 has the earliest peak day. Similarly, the JRA-55C has a smaller number of landfalling ARs, although the interannual variability is comparable to the other two reanalyses.

3.5.2 Evaluation of climate models

Figure 13 shows CMIP5 and CMIP6 models' performance in simulating AR peak season compared to ERA5 reanalysis. To explore how model biases compare to the discrepancies among reanalyses, we also include AR peak day bias for MERRA-2 and JRA-55C reanalysis in the left two columns of the metrics plot. Perhaps unsurprisingly, the model spread is much larger than the spread among reanalysis products, which are tightly constrained by data assimilation.

In regions like South America, Baja, UK and Western Europe, the models show systematic late peak biases, and in South Africa, AR peaks earlier than the reanalyses. The exact cause of these structural biases in the models is likely indicative of persistent and ubiquitous timing issues in the shift of the storm track that is common among models. It is worth noting that the model biases in the West Africa region are significantly larger than other regions, with peak day difference up to 6 months as compared to the reanalysis. Looking at the AR counts histograms over the course of the year in this region in the CCSM3 and MRI-ESM2-0 models (Fig. 14), it is clear that AR landfall seasonality in both models is completely out of phase with ERA5. This is especially true for the MRI-ESM2-0 model, where AR landfall peaks in June, which is in opposition to the climatology in ERA5. The large discrepancy is probably because of the large spread in the atmospheric circulations in climate models over the West Africa region, as large spread among CMIP5/6 models in capturing atmospheric dynamic responses (Monerie et al. 2020), the lack of jet-rainfall coupling (Whittleson et al. 2017), and bias in simulating mesoscale convective systems (Jenkins et al. 2002) in climate models are noted. Although high resolution regional modeling may be capable of improving rainfall in this region (Sylla et al. 2009), the dynamics-rainfall coupling does not appear to be improved in high resolution global models such as the E3SM (Caldwell et al. 2019; Golaz et al. 2019). Therefore, challenges remain in modeling the AR water cycle in west Africa.

Geoscientific Model Development Discussions

4. Summary and discussion

In this study we have introduced a workflow for the objective evaluation of ARs in climate models and reanalysis, and have illustrated the potential for its use with five example case-studies to illustrate the scope of potential applications. The metrics-based analyses are designed for systematic diagnosis of AR biases in climate models. For example, applying the package to CMIP5 and CMIP6 models, we have shown that AR tracks in the south Atlantic are positioned farther poleward compared to the ERA5 reanalysis, while in the south Pacific, tracks are biased towards the equator. Over western Africa, we found that most climate models do a poor job at capturing the AR peak season. In addition to model evaluation, we have shown how our tool can be used to identify structural differences resulting from the choice of AR detector (ARDT). For instance, we demonstrated that ARs detected with the Mundhenk method are systematically larger in size, width and length compared to TE.

The workflow and metrics presented in this study can be used for a variety applications, e.g., to contrast the differences between AR features in historical and future scenarios as simulated by climate models. Objectively quantifying projected changes in landfall frequency, duration, and intervals between landfall events are of particular interest. Further confidence in this and other model evaluation applications can be gained by assessing what impact the choice of the ARDT can have on any conclusions concerning model quality. Our tool makes this and other sensitivity tests more tractable.

Our tool also pools a diverse suite of established and newly introduced AR metrics into one framework, facilitating objective evaluation of ARs with a diverse suite of input data, as well as intercomparison of ARs as simulated by multiple climate models. These metrics can be routinely applied for model benchmarking and during development cycles to monitor changes in AR characteristics across model versions or generations and set objective targets for the improvement of models. One expected application is the routine benchmarking of AR in simulations with increasingly higher resolution models. More frequent metrics evaluation of simulated ARs such as this could further our understanding of model bias and error characteristics, and potentially assist developers in making choices associated with new model versions. Furthermore, it effectively provides a quantitative measure for operational centres to perform near real-time climate and extreme events impact assessment along with their forecast cycles, which can facilitate their decision-making process.

Our metrics tool is developed with Xarray (Hoyer et al., 2017), XCDAT (Vo et al., 2024), and the PCMDI Metrics Package (PMP; Lee et al. 2024), which are compatible with one another, readily available and easy to install. At the time of the submission of this manuscript, our tool is being configured to be a part of the PMP. Looking forward, we welcome community contributions to successive development of the package. Inspired by Zarzycki et al. (2021), there is also a potential that these metrics can be applied for research beyond ARs, such as mesoscale meteorological features, regional hydrological extremes such as floods and droughts, and large-scale climate modes.

Acknowledgment

This work is performed under the auspices of the U.S. Department of Energy (DOE) by Lawrence Livermore National Laboratory (LLNL) under Contract No. DE-AC52-07NA27344 and is mainly supported by the Regional and Global Model Analysis (RGMA) program of the U.S. DOE Office of Science (OS) Biological and Environmental Research (BER) program. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Climate and Environmental Sciences Division, Regional & Global Model Analysis Program. Resources of the National Energy Research Scientific Computing Center (NERSC) were used. The research was partially supported by the Office of Science of the U.S. Department of Energy under Contract Number DE-AC02-05CH11231 and under award Number DE-SC0023519. This research was also supported in part by the Environmental Resilience Institute, funded by Indiana University's Prepared for Environmental Change Grand Challenge initiative and in part by Lilly Endowment, Inc., through its support for the Indiana University Pervasive Technology Institute. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modeling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies that support CMIP6 and ESGF. The authors acknowledge Antony Hoang and Ana Ordonez for computing and technical support, and Christine Shields, Yang Zhou and Allison Collow for their help on data and discussion.

Code and data availability

The metrics framework code is available on github. Users are recommended to install full PMP package (http://pcmdi.github.io/pcmdi_metrics/install.html) in order to have the environment and python packages to run the metrics code.

Author contribution

BD implemented the codes and developed the diagnostic results. All authors contributed to the writing of the manuscript.

Competing interests

At least one of the (co-)authors is a member of the editorial board of Geoscientific Model Development.

References

Algarra, I., Nieto, R., Ramos, A. M., Eiras-Barca, J., Trigo, R. M., & Gimeno, L. (2020). Significant increase of global anomalous moisture uptake feeding land-falling atmospheric rivers. Nature communications, 11 (1), 5082.

Beck, H. E., Van Dijk, A. I., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., & De Roo, A. (2017). MSWEP: 3-hourly 0.25 global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences*, *21*(1), 589-615.

Caldwell, P. M., Mametjanov, A., Tang, Q., Van Roekel, L. P., Golaz, J. C., Lin, W., ... & Zhou, T. (2019). The DOE E3SM coupled model version 1: Description and results at high resolution. Journal of Advances in Modeling Earth Systems, 11(12), 4095-4146.

Chapman, W. E., Subramanian, A. C., Delle Monache, L., Xie, S. P., & Ralph, F. M. (2019). Improving atmospheric river forecasts with machine learning. Geophysical Research Letters, 46(17-18), 10627-10635.

DeFlorio, M. J., Waliser, D. E., Guan, B., Lavers, D. A., Ralph, F. M., & Vitart, F. (2018). Global assessment of atmospheric river prediction skill. Journal of Hydrometeorology, 19(2), 409-426.

Dettinger, M. D., Ralph, F. M., Das, T., Neiman, P. J., & Cayan, D. R. (2011). Atmospheric rivers, floods and the water resources of california. Water , 3 (2), 445–478.

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

Gelaro, R., McCarty, W., Su´arez, M. J., Todling, R., Molod, A., Takacs, L., . . .others (2017). The modern-era retrospective analysis for research and applications, version 2 (merra-2). Journal of climate, 30 (14), 5419–5454.

Gershunov, A., Shulgina, T., Clemesha, R. E., Guirguis, K., Pierce, D. W., Dettinger, M. D., . . . others (2019). Precipitation regime change in western north america: the role of atmospheric rivers. Scientific reports, 9 (1), 9944.

Gimeno, L., Nieto, R., V´azquez, M., & Lavers, D. A. (2014). Atmospheric rivers: Amini-review. Frontiers in Earth Science, 2 , 2.

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

Gleckler, P. J., Taylor, K. E., & Doutriaux, C. (2008). Performance metrics for climate models. Journal of Geophysical Research: Atmospheres, 113(D6).

Golaz, J. C., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., ... & Zhu, Q. (2019). The DOE E3SM coupled model version 1: Overview and evaluation at standard resolution. Journal of Advances in Modeling Earth Systems, 11(7), 2089-2129.

Griffith, D. A. (2013). Establishing qualitative geographic sample size in the presence of spatial autocorrelation. Annals of the Association of American Geographers, 103(5), 1107-1122.

Guan, B., Molotch, N. P., Waliser, D. E., Fetzer, E. J., & Neiman, P. J. (2010). Extreme snowfall events linked to atmospheric rivers and surface air temperature via satellite measurements. Geophysical Research Letters, 37 (20).

Guan, B., & Waliser, D. E. (2017). Atmospheric rivers in 20 year weather and climate simulations: A multimodel, global evaluation. Journal of Geophysical Research: Atmospheres, 122(11), 5556-5581.

Guan, B., Waliser, D. E., & Ralph, F. M. (2023). Global application of the atmospheric river scale. Journal of Geophysical Research: Atmospheres, 128(3), e2022JD037180.

Harrop, B., Leung, L. and Ullrich P. (2023). Improving Simulations of Atmospheric Rivers and Heat Waves in the Coupled E3SM. FY2023 First Quarter Performance Metric. DOE/SC-CM-23-001

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Hor´anyi, A., Mu˜noz-Sabater, J.,. . . others (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146 (730), 1999–2049.

Hoyer, S. & Hamman, J., (2017). xarray: N-D labeled Arrays and Datasets in Python. Journal of Open Research Software. 5(1), p.10. DOI: https://doi.org/10.5334/jors.148

Huang, X., Swain, D. L., & Hall, A. D. (2020). Future precipitation increase from very high resolution ensemble downscaling of extreme atmospheric river storms in California. Science advances, 6(29), eaba1323.

Huang, J., Zhang, C., & Prospero, J. M. (2009). African aerosol and large-scale precipitation variability over West Africa. *Environmental Research Letters*, *4*(1), 015006.

Hui, W. J., Cook, B. I., Ravi, S., Fuentes, J. D., & D'Odorico, P. (2008). Dust-rainfall feedbacks in the West African Sahel. *Water Resources Research*, *44*(5).

Inda-Díaz, H. A., O'Brien, T. A., Zhou, Y., & Collins, W. D. (2021). Constraining and characterizing the size of atmospheric rivers: A perspective independent from the detection algorithm. *Journal of Geophysical Research: Atmospheres*, *126*(16), e2020JD033746.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., . . . others (2015). The JRA-55 reanalysis: General specifications and basic characteristics. Journal of the Meteorological Society of Japan. Ser. II , 93 (1), 5–48.

Lee, J., P. J. Gleckler, M.-S. Ahn, A. Ordonez, P. Ullrich, K. R. Sperber, K. E. Taylor, Y. Y. Planton, E. Guilyardi, P. Durack, C. Bonfils, M. D. Zelinka, L.-W. Chao, B. Dong, C. Doutriaux, C. Zhang, T. Vo, J. Boutte, M. F. Wehner, A. G. Pendergrass, D. Kim, Z. Xue, A. T. Wittenberg, and J. Krasting, (2024): Systematic and Objective Evaluation of Earth System Models: PCMDI Metrics Package (PMP) version 3. Geoscientific Model Development, 17, 3919–3948, doi: 10.5194/gmd-17-3919-2024.

Leung, L. R., & Qian, Y. (2009). Atmospheric rivers induced heavy precipitation and flooding in the western us simulated by the WRF regional climate model. Geophysical research letters, 36 (3).

Lora, J. M., Mitchell, J. L., Risi, C., & Tripati, A. E. (2017). North pacific atmospheric rivers and their influence on western north America at the last glacial maximum. Geophysical Research Letters, 44 (2), 1051–1059.

Massoud, E. C., Espinoza, V., Guan, B., & Waliser, D. E. (2019). Global Climate Model Ensemble Approaches for Future Projections of Atmospheric Rivers. Earth's Future, 7: 1136–11511151.

Mundhenk, B. D., Barnes, E. A., & Maloney, E. D. (2016). All-season climatology and variability of atmospheric river frequencies over the north pacific. Journal of Climate, 29 (13), 4885–4903.

Nardi, K. M., Barnes, E. A., & Ralph, F. M. (2018). Assessment of numerical weather prediction model reforecasts of the occurrence, intensity, and location of atmospheric rivers along the West Coast of North America. Monthly Weather Review, 146(10), 3343-3362.

Neiman, P. J., Ralph, F. M., Moore, B. J., Hughes, M., Mahoney, K. M., Cordeira, J. M., & Dettinger, M. D. (2013). The landfall and inland penetration of a flood-producing atmospheric river in arizona. part i: Observed synoptic-scale, orographic, and hydrometeorological characteristics. Journal of Hydrometeorology, 14 (2), 460–484.

Neiman, P. J., Schick, L. J., Ralph, F. M., Hughes, M., & Wick, G. A. (2011). Flooding in western washington: The connection to atmospheric rivers. Journal of Hydrometeorology, 12 (6), 1337–1358.

Neiman, P. J., White, A. B., Ralph, F. M., Gottas, D. J., & Gutman, S. I. (2009). A water vapour flux tool for precipitation forecasting. In Proceedings of the institution of civil engineers-water management (Vol. 162, pp. 83–94).

Newman, M., Kiladis, G. N., Weickmann, K. M., Ralph, F. M., & Sardeshmukh, P. D. (2012). Relative contributions of synoptic and low-frequency eddies to time-mean

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

atmospheric moisture transport, including the role of atmospheric rivers. *Journal of climate*, *25*(21), 7341-7361.

O'Brien, T., Wehner, M., Payne, A., Shields, C., Rutz, J., Leung, L.-R., . . . others (2022). Increases in Future AR Count and Size: Overview of the ARTMIP Tier 2 CMIP5/6 Experiment. Journal of Geophysical Research: Atmospheres, 127 (6).

Payne, A. E., Demory, M. E., Leung, L. R., Ramos, A. M., Shields, C. A., Rutz, J. J., ... & Ralph, F. M. (2020). Responses and impacts of atmospheric rivers to climate change. *Nature Reviews Earth & Environment*, *1*(3), 143-157.

Ramachandran, J., & Aschheim, M. A. (2005). Sample size and error in the determination of mode shapes by principal components analysis. Engineering structures, 27(14), 1951-1967.

Ralph, F., Coleman, T., Neiman, P., Zamora, R., & Dettinger, M. (2013). Observed impacts of duration and seasonality of atmospheric-river landfalls on soil moisture and runoff in coastal northern california. Journal of Hydrometeorology, 14 (2), 443–459.

Ralph, F. M., Neiman, P. J., Kiladis, G. N., Weickmann, K., & Reynolds, D. W. (2011). A multiscale observational case study of a pacific atmospheric river exhibiting tropical—-extratropical connections and a mesoscale frontal wave. Monthly Weather Review , 139 (4), 1169–1189.

Ralph, F. M., Neiman, P. J., Wick, G. A., Gutman, S. I., Dettinger, M. D., Cayan, D. R., & White, A. B. (2006). Flooding on california's russian river: Role of atmospheric rivers. Geophysical Research Letters, 33 (13).

Ralph, F.M., Dettinger, M.D., Cairns, M.M., Galarneau, T.J. and Eylander, J. (2018). Defining "atmospheric river": How the Glossary of Meteorology helped resolve a debate. Bulletin of the American Meteorological Society, 99(4), pp.837-839.

Rutz, J. J., Steenburgh, W. J., & Ralph, F. M. (2014). Climatological characteristics of atmospheric rivers and their inland penetration over the western United states. Monthly Weather Review , 142 (2), 905–921.

Rutz, J.J., Shields, C.A., Lora, J.M., Payne, A.E., Guan, B., Ullrich, P., O'brien, T., Leung, L.R., Ralph, F.M., Wehner, M. and Brands, S. (2019). The atmospheric river tracking method intercomparison project (ARTMIP): Quantifying uncertainties in atmospheric river climatology. *Journal of Geophysical Research: Atmospheres*, *124*(24), pp.13777-13802.

Shields, C. A., Rosenbloom, N., Bates, S., Hannay, C., Hu, A., Payne, A. E., . . Truesdale, J. (2019). Meridional heat transport during atmospheric rivers in high-resolution cesm climate projections. Geophysical Research Letters, 46 (24), 14702–14712.

Shields, C. A., Rutz, J. J., Leung, L. R., Ralph, F. M., Wehner, M., O'Brien, T., & Pierce, R. (2019). Defining uncertainties through comparison of atmospheric river tracking methods. Bulletin of the American Meteorological Society, 100 (2), ES93–ES96.

Solmon, F., Mallet, M., Elguindi, N., Giorgi, F., Zakey, A., & Konaré, A. (2008). Dust aerosol impact on regional precipitation over western Africa, mechanisms and sensitivity to absorption properties. *Geophysical Research Letters*, *35*(24).

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. Journal of geophysical research: atmospheres, 106(D7), 7183-7192.

Ullrich, P. A., & Zarzycki, C. M. (2017). Tempestextremes: A framework for scale-insensitive pointwise feature tracking on unstructured grids. Geoscientific Model Development, 10 (3), 1069–1090.

Ullrich, P. A., Zarzycki, C. M., McClenny, E. E., Pinheiro, M. C., Stansfield, A. M., & Reed, K. A. (2021). Tempestextremes v2. 1: A community framework for feature detection, tracking and analysis in large datasets. Geoscientific model development discussions, 2021 , 1–37.

Vo et al., (2024). xCDAT: A Python Package for Simple and Robust Analysis of Climate Data. Journal of Open Source Software, 9(98), 6426, https://doi.org/10.21105/joss.06426

Wilks, D. S. (2011). Statistical methods in the atmospheric sciences (Vol. 100). Academic press.

Zarzycki, C. M., Ullrich, P. A., & Reed, K. A. (2021). Metrics for evaluating tropical cyclones in climate data. Journal of Applied Meteorology and Climatology, 60 (5), 643–660.

Zhang, L., Zhao, Y., Cheng, T. F., & Lu, M. (2024). Future changes in global atmospheric rivers projected by CMIP6 models. Journal of Geophysical Research: Atmospheres, 129, e2023JD039359

Zhao, A., Ryder, C. L., & Wilcox, L. J. (2022). How well do the CMIP6 models simulate dust aerosols?. *Atmospheric Chemistry and Physics*, *22*(3), 2095-2119.

Zhu, Y., & Newell, R. E. (1998). A proposed algorithm for moisture fluxes from atmospheric rivers. Monthly weather review , 126 (3), 725–735

Supplementary information

In a separate document

Figures and tables

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

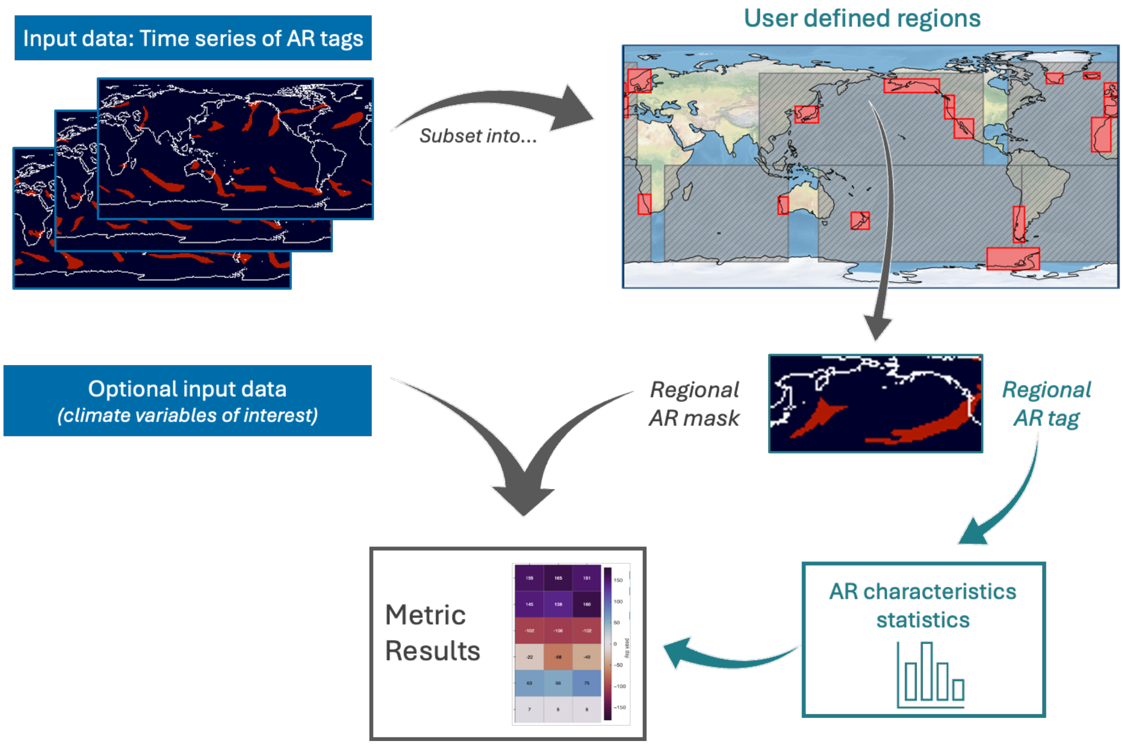


Fig. 1. AR metric tool workflow. Input data include time slices of AR tags from ARDTs of user choice, and optional climate data associated with ARs. The data are then subset into user-defined rectangular domains (blue boxes for ocean basins, red boxes for landfall regions) for regional tags and masks. User preferred statistical tools are applied on the regional AR tags to obtain AR characteristics. Finally, AR characteristics and AR masked climate data are presented as metric results.

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

Geoscientific Model Development Discussions

Table 1. List of AR metrics and diagnostics in this study. Numbers in the table indicate the number of regions where the metrics are applied. Each column is one AR property. Underscored items are model evaluation metrics, items in italic form are diagnostics of AR properties.

	ARs over Ocean Basins						Landfalling ARs		
metrics/ diagnostics	frequency	central latitude	central longitude	size	width	length	counts (frequency)	peak day	precipitation
mean bias	5	5	5	5	5	5	16	16	16
spatial correlation	5								
IoU							16		
spatial distribution	5						16		
sampling histogram		5	5	5	5	5			
monthly climatology histogram							16		

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

Geoscientific Model Development
Discussions
Open Access
EGU

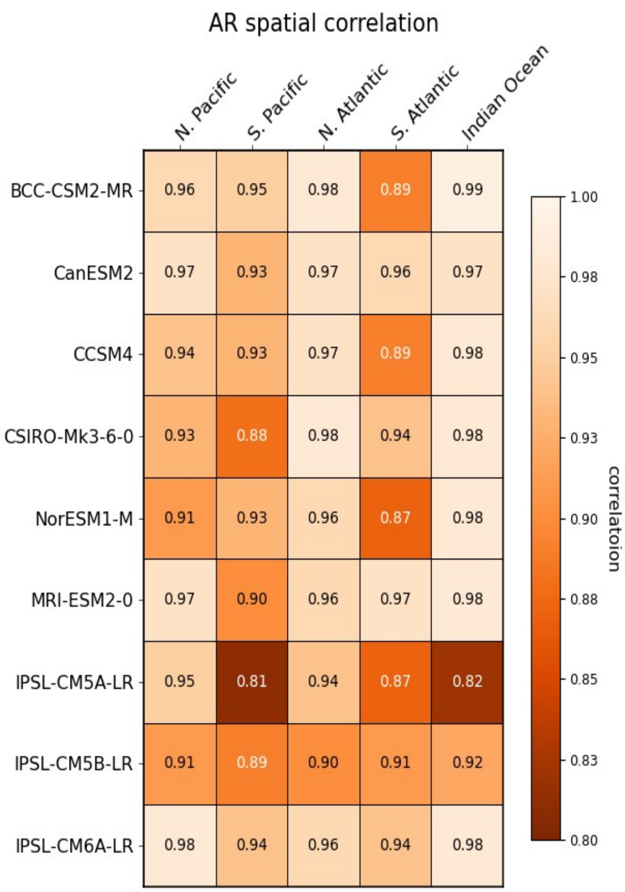


Fig. 2. Spatial pattern correlation of AR frequency for the period 1979-1989 between ERA5 and climate models for major ocean basins.

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

Geoscientific Model Development
Discussions
Open Access

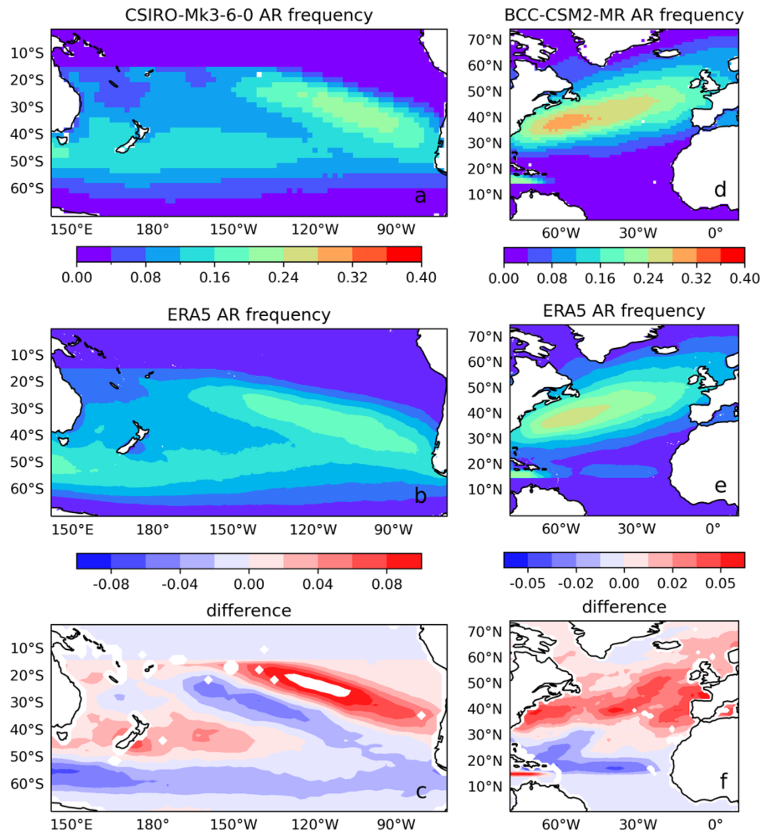


Fig. 3. AR frequency in the South Pacific for (a) CSIRO-MK3-6-0, (b) ERA5 and their difference (c) as (a) - (b). AR frequency in the North Atlantic for (d) BCCCSM2-MR, (e) ERA5 and their difference (f) as (d) - (e)

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

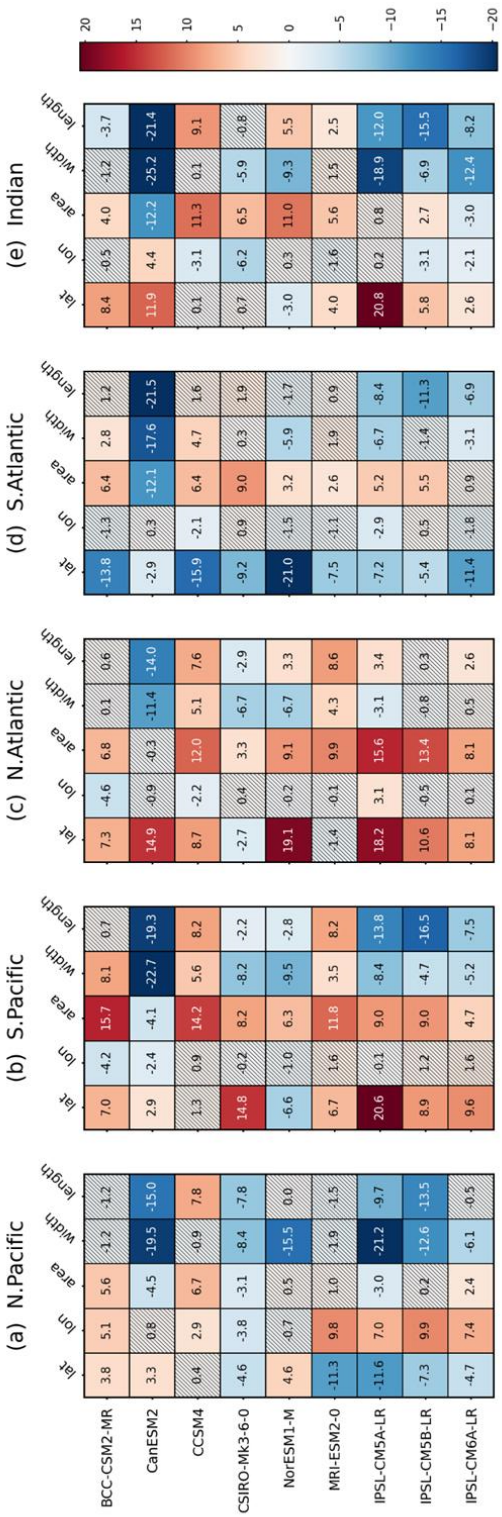
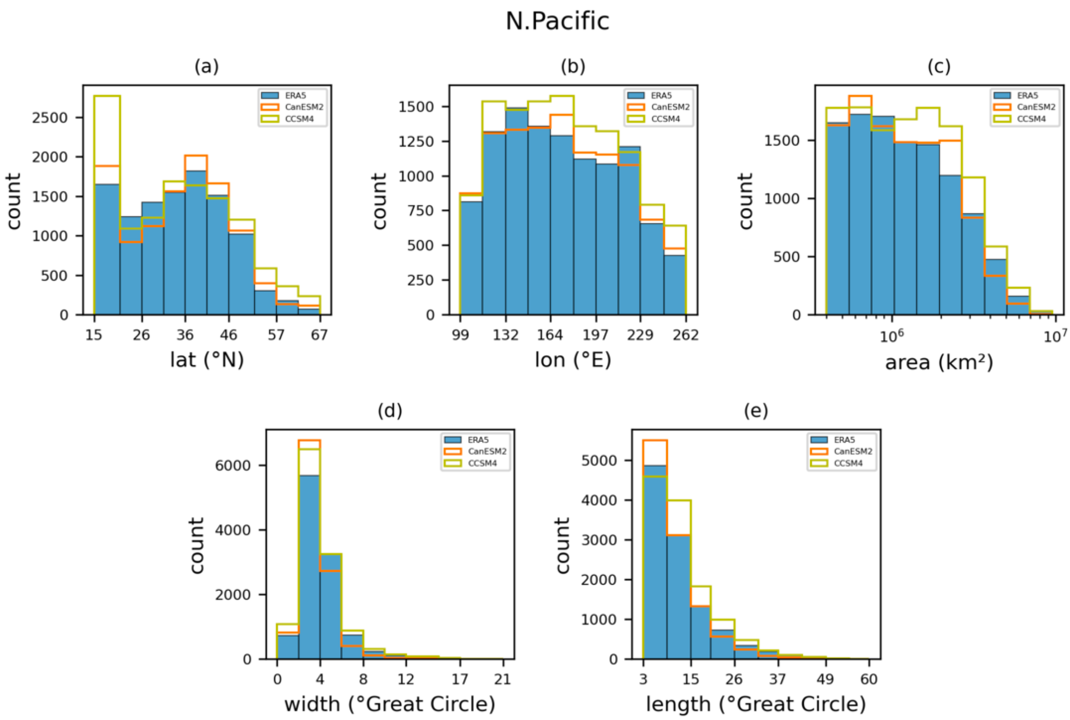


Fig. 4 AR characteristics bias (normalized as Z-score) in climate models for major ocean basins. Hatching indicates that the differences are statistically insignificant.

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

Geoscientific Model Development
Discussions

771



772

773 Fig. 5 North Pacific AR characteristics distribution for (a) central latitude, (b)
774 central longitude, (c) area, (d) width and (e) length, in ERA5 reanalysis,
775 CanESM2 and CCSM4 model

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

Geoscientific Model Development
Discussions
Open Access

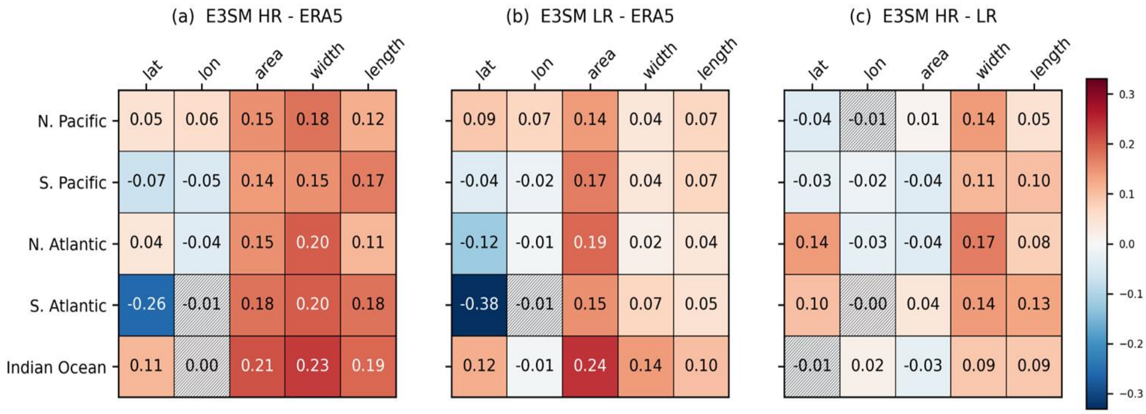


Fig. 6. AR characteristics bias in E3SM (a) HR and (b) LR simulations. (c) is the difference between HR and LR. Hatching indicate that the differences are statistically insignificant.

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

Geoscientific Model Development Discussions

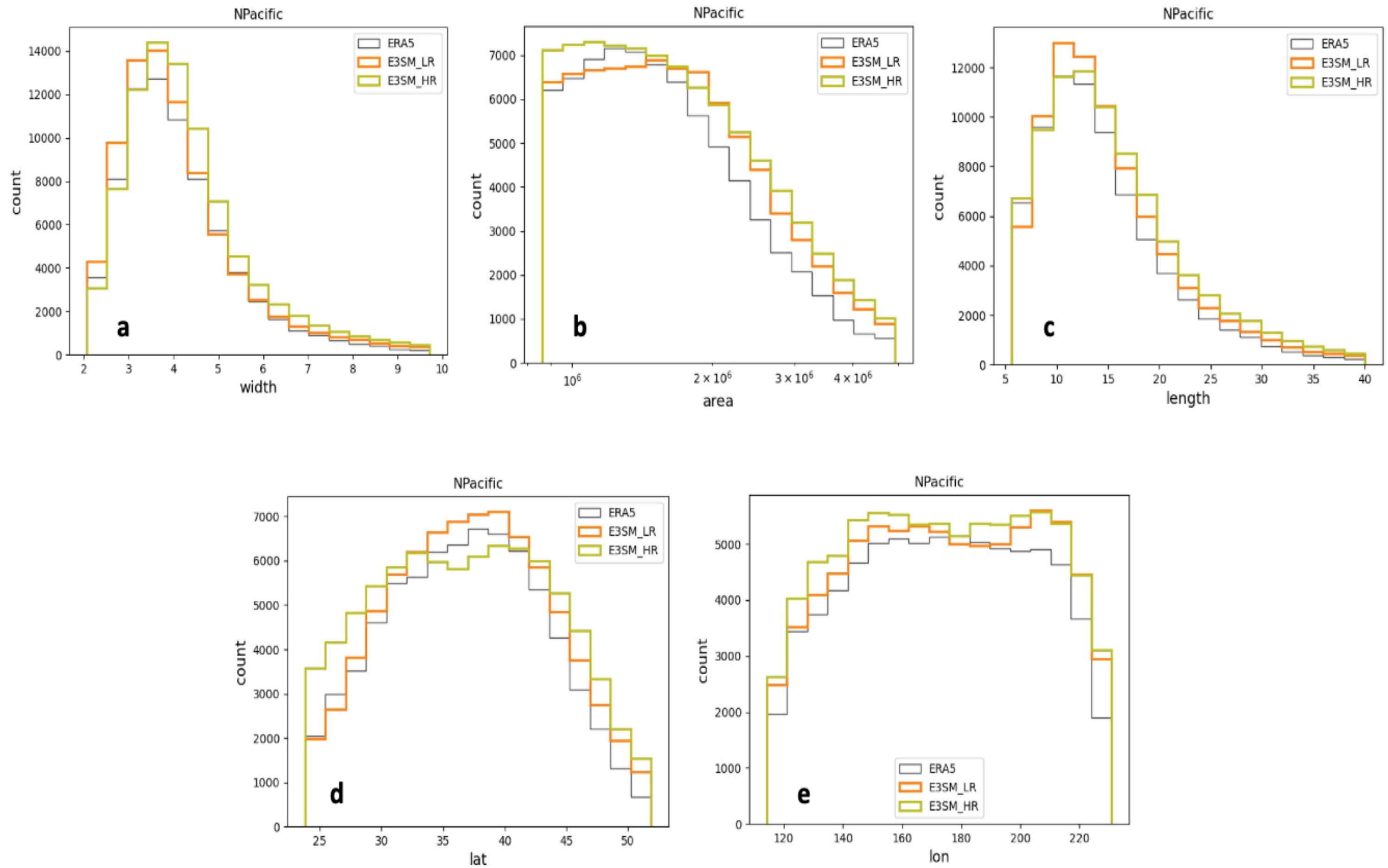


Fig. 7. AR characteristics distribution of (a) width (° great circle), (b) area (km^2), (c) length (°great circle), (d) central latitude (°N), and (e) central longitude (°E) in the North Pacific for ERA5, E3SM HR and LR simulations.

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

Geoscientific Model Development
Discussions
Open Access

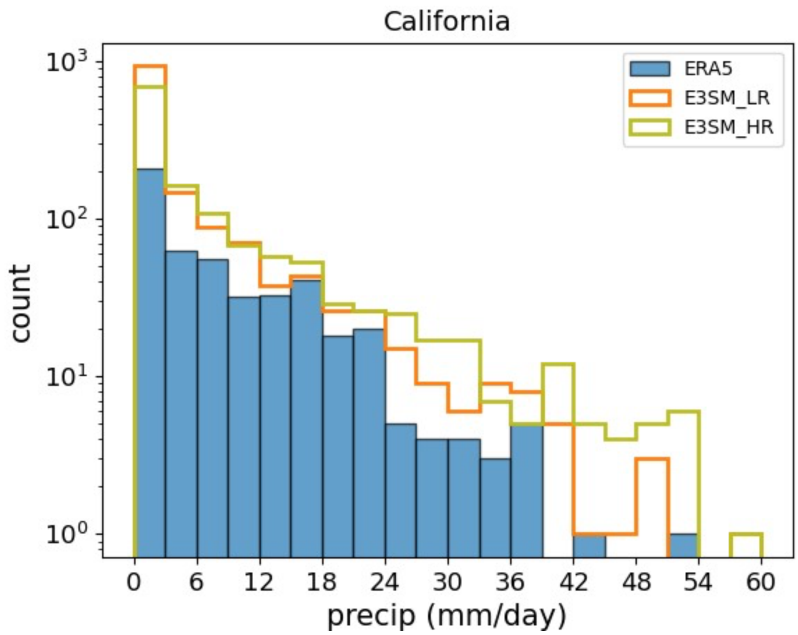


Fig. 8. Landfalling AR precipitation histogram in California from 1990-1999 in the ERA5 reanalysis, E3SM HR and LR simulations.

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

Geoscientific Model Development
Discussions

Table 2. AR landfall concurrence in Mundhenk and TE, normalized by total counts of AR landfalls detected in both ARDTs for different regions. Values are shown in percentage.

Region	California	S. America	N. Europe	Australia	S. Africa	Baja	Pacific Northwest	New Zealand
Concurrence (%)	56	68	82	62	51	30	72	77
Region	Alaska	UK	W. Europe	Iceland	Greenland	E. Asia	Antarctica	New England
Concurrence (%)	81	84	74	77	72	56	69	83

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

Geoscientific Model Development
Discussions
Open Access

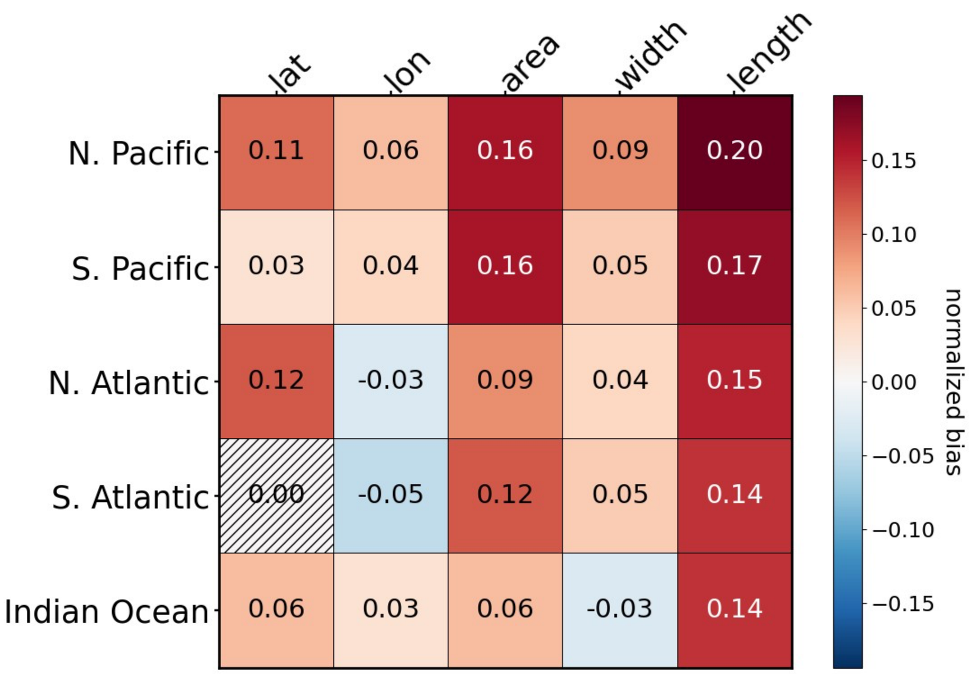


Fig. 9. AR characteristic difference between Mundhenk and TE in ERA5

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

Geoscientific Model Development
Discussions

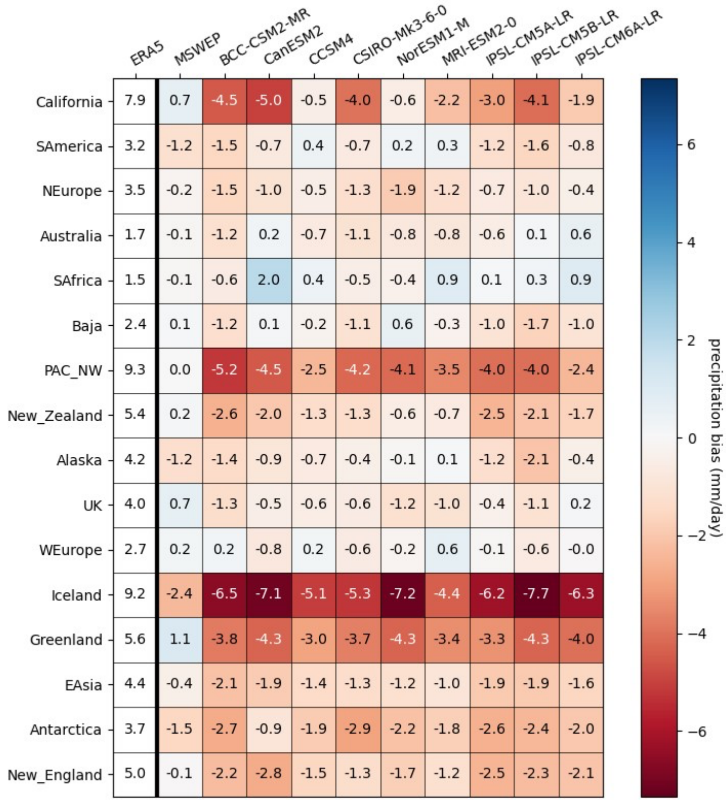


Fig. 10. Landfalling AR precipitation bias in climate models relative to ERA5 (the first column). The MSWEP data is also included in the second column as an additional reference data, showed as the difference between ERA5.

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

Geoscientific Model Development
Discussions
Open Access

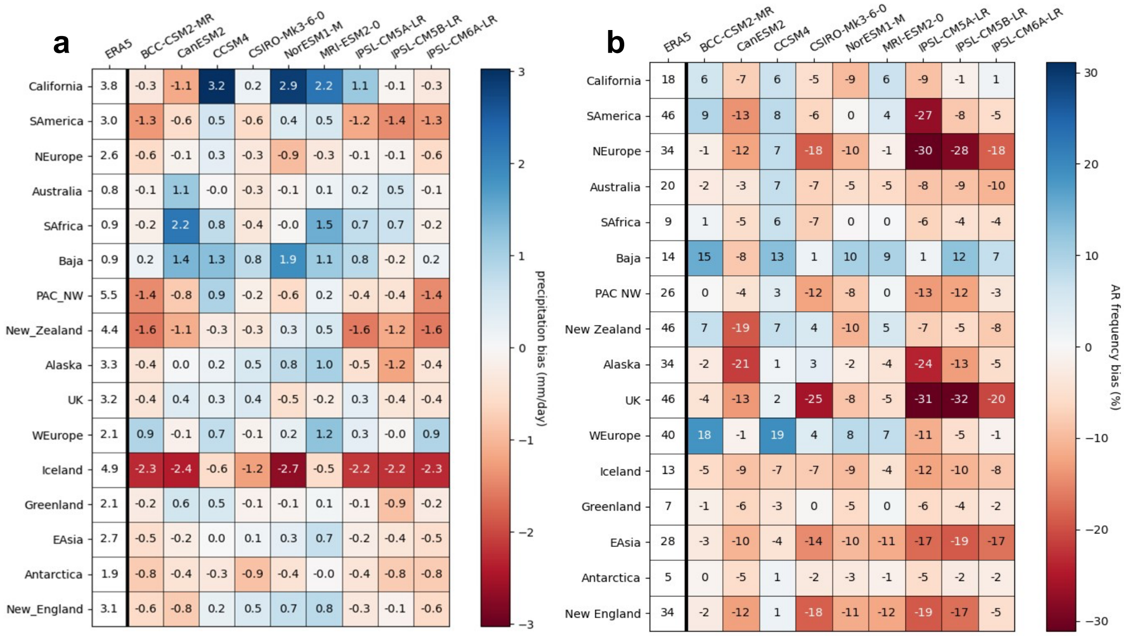


Fig. 11. (a) Total precipitation bias and (b) landfalling AR frequency bias

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

Geoscientific Model Development
Discussions
Open Access

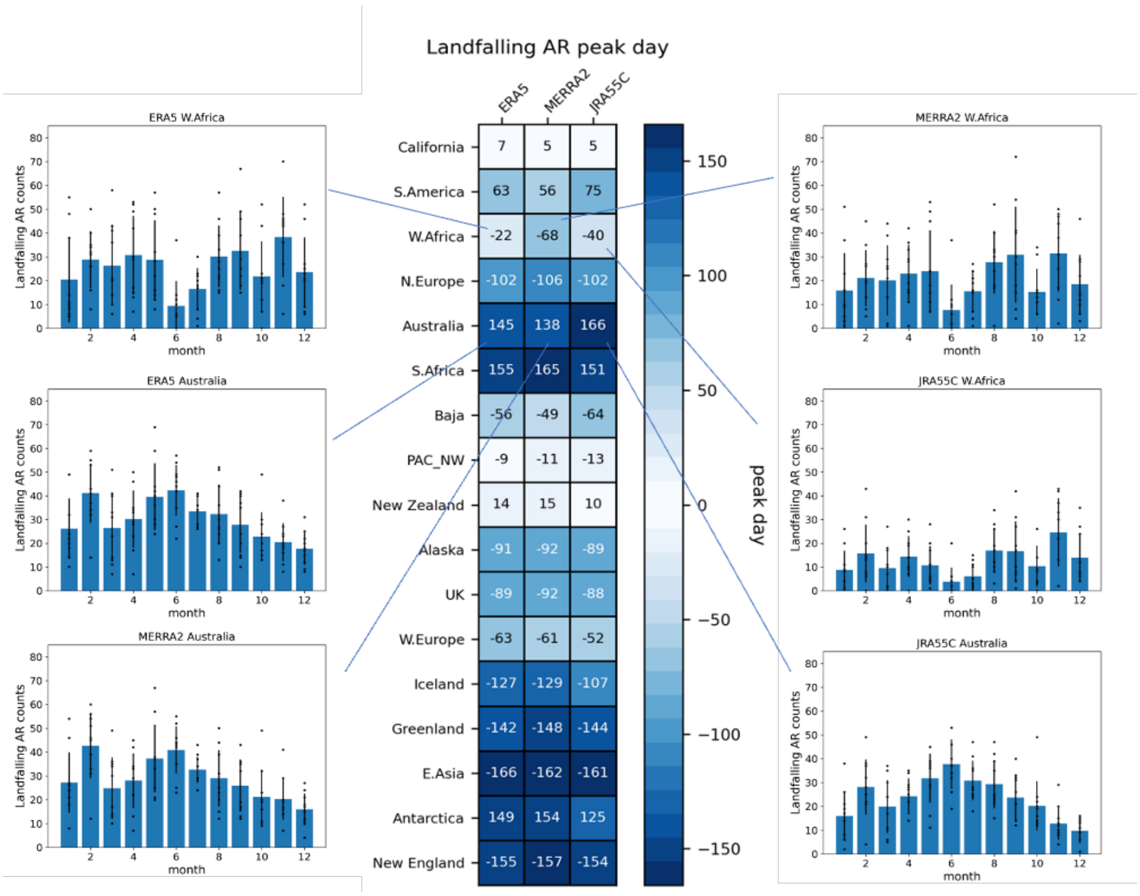


Fig. 12 (a) Landfalling AR peak day in ERA5, MERRA2, and JRA55C reanalysis. (b-g) show examples of probability distribution. Height of the blue bars indicate the time mean counts. Black dots represent peak day for each individual year, and vertical bars are the standard deviation range in the 10-year data from 1979-1988

Geoscientific Model Development
Discussions
Open Access

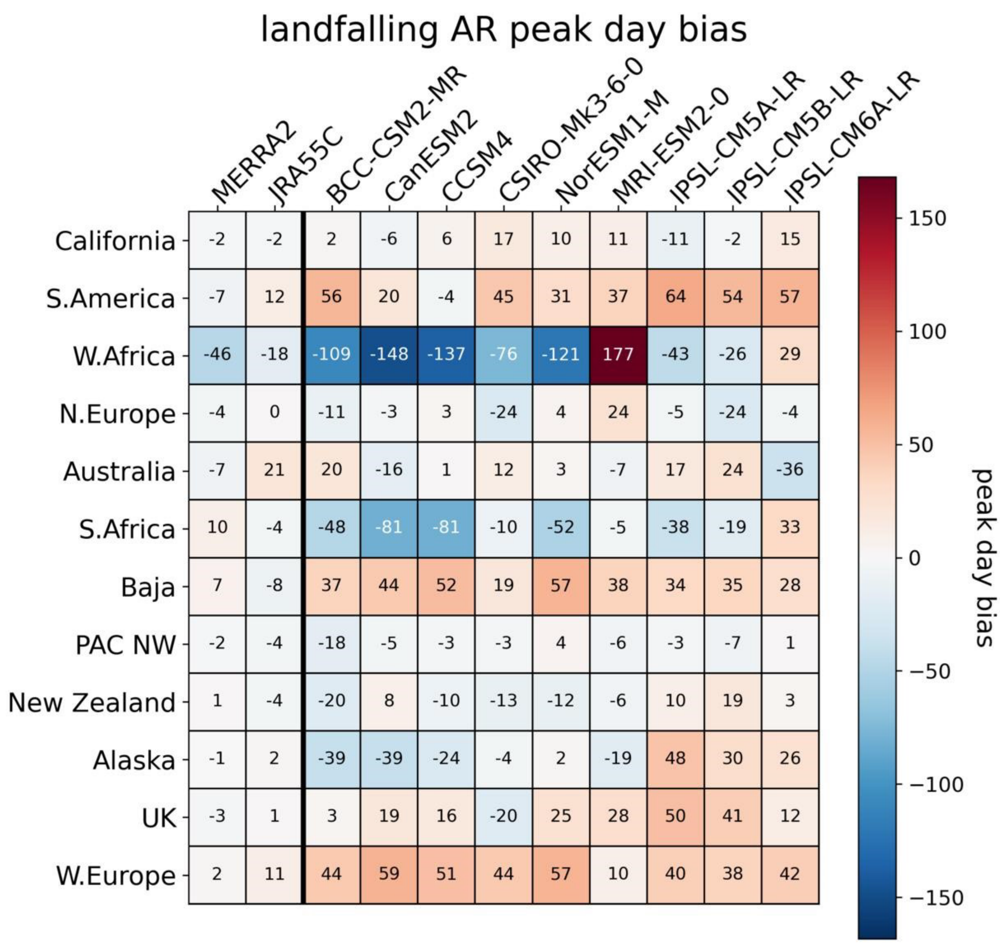


Fig. 13. Landfalling AR peak day bias in reanalyses and models compared with ERA5.

https://doi.org/10.5194/gmd-2024-142
Preprint. Discussion started: 20 August 2024

Geoscientific Model Development
Discussions

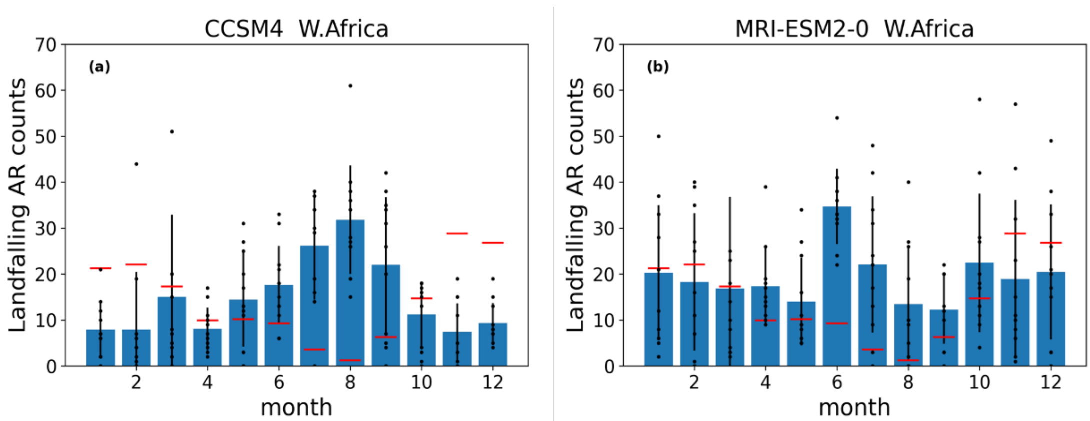


Fig. 14. Landfalling AR counts in (a) CCSM4 and (b) MRI-ESM2-0 for western Africa region. Height of the blue bars indicate the time mean counts. Vertical lines represent the standard deviations. Black dots represent counts for each individual year. Red bars show ERA5 values as the reference.