Increases in Future AR Count and Size: Overview of the ARTMIP Tier 2 CMIP5/6 Experiment

Travis Allen O'Brien[1], Michael F Wehner[2], Ashley E. Payne[3], Christine A Shields[4], Jonathan J. Rutz[5], L. Ruby Leung[6], F. Martin Ralph[7], Allison B. Marquardt Collow[8], Bin Guan[9], Juan Manuel Lora[10], Elizabeth McClenny[11], Kyle M. Nardi[12], Alexandre M. Ramos[13], Ricardo Tomé[13], Chandan Sarangi[14], Eric Jay Shearer[15], Paul Ullrich[16], Colin M. Zarzycki[17], Burlen Loring[18], Huanping Huang[19], Héctor Alejandro Inda Díaz[16], Alan M. Rhoades[20], and Yang Zhou[20]

[1]Indiana University Bloomington
[2]Lawrence Berkeley National Laboratory (DOE)
[3]University of Michigan-Ann Arbor
[4]National Center for Atmospheric Research (UCAR)
[5]National Weather Service, National Oceanic and Atmospheric Administration
[6]PNNL
[7]National Oceanic and Atmospheric Administration (NOAA)
[8]Universities Space Research Association and NASA/GSFC
[9]University of California Los Angeles
[10]Yale University
[11]University of California, Davis
[12]Colorado State University
[13]Instituto Dom Luiz
[14]IIT Kanpur
[15]University of California, Irvine
[16]University of California Davis
[17]Pennsylvania State University
[18]Lawrence Berkeley Lab
[19]Dartmouth College
[20]Lawrence Berkeley National Laboratory

November 30, 2022

Abstract

The Atmospheric River (AR) Tracking Method Intercomparison Project (ARTMIP) is a community effort to systematically assess how the uncertainties from AR detectors (ARDTs) and climate models impact our scientific understanding of ARs. This study describes the ARTMIP Tier 2 experimental design and initial results using the Coupled Model Intercomparison Project (CMIP) Phases 5 and 6 multi-model ensembles. We show that AR statistics from a given ARDT in CMIP5/6 historical simulations compare remarkably well with reanalyses. In CMIP5/6 future simulations, most ARDTs project a global increase in AR frequency, counts, and sizes, especially along the western coastlines of the Pacific and Atlantic oceans. We find that the choice of ARDT is the dominant contributor to the uncertainty in projected AR frequency when compared with model choice.

These results imply that new projects investigating future changes in ARs should explicitly consider ARDT uncertainty as a core part of the experimental design.

Increases in Future AR Count and Size: Overview of the ARTMIP Tier 2 CMIP5/6 Experiment

T. A. O'Brien[1,2*], M. F. Wehner[3] and A. E. Payne[4] and C. A. Shields[5] and J. J. Rutz[6] and L.-R. Leung[7] and F. M. Ralph[8] and A. Collow[9,10] and B. Guan[11] and J. M. Lora[12] and E. McClenny[13] and K. M. Nardi[14] and A. M. Ramos[15] and R. Tomé[15] and C. Sarangi[7,17] and E. Shearer[16] and P. A. Ullrich[13] and C. Zarzycki[14] and B. Loring[3] and H. Huang[2] and H. A. Inda-Díaz[13,2] and A. M. Rhoades[2] and Y. Zhou[2]

[1]Dept. of Earth and Atmospheric Sciences, Indiana University, Bloomington, IN, USA
[2]Climate and Ecosystem Sciences Division, Lawrence Berkeley Lab, Berkeley, CA, USA
[3]Computational Research Division, Lawrence Berkeley Lab, Berkeley, CA, USA
[4]Dept. of Earth and Space Sciences, University of Michigan, Ann Arbor, MI, USA
[5]National Center for Atmospheric Research, Boulder, CO, USA
[6]National Weather Service, Western Region Headquarters, Science and Technology Infusion Division, Salt Lake City, UT, USA
[7]Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, WA, USA
[8]Center for Western Weather and Water Extremes, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, USA
[9]Universities Space Research Association, Columbia, MD, USA
[10]Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA
[11]Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, CA, USA
[12]Dept. of Earth and Planetary Sciences, Yale University, New Haven, CT, USA
[13]Dept. of Land, Air and Water Resources, University of California, Davis, Davis, CA, USA
[14]Dept. of Meteorology and Atmospheric Science, Penn State University, University Park, PA, USA
[15]Instituto Dom Luiz (IDL), Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal
[16]Center for Hydrometeorology and Remote Sensing, University of California, Irvine, Irvine, CA, USA
[17]Department of Civil Engineering, Indian Institute of Technology Madras, India

Key Points:

- Uncertainty associated with AR definition dominates model uncertainty for projections of Pacific and Atlantic landfalling ARs
- Most AR detection algorithms show an increase in AR frequency in future simulations
- AR statistics in CMIP 5-and-6 models compare remarkably well with reanalysis

*Dept. of Earth and Atmospheric Science, 1001 E. 10th St, Bloomington, IN, 47408

Corresponding author: T.A. O'Brien, `obrienta@iu.edu`

Abstract

The Atmospheric River (AR) Tracking Method Intercomparison Project (ARTMIP) is a community effort to systematically assess how the uncertainties from AR detectors (ARDTs) and climate models impact our scientific understanding of ARs. This study describes the ARTMIP Tier 2 experimental design and initial results using the Coupled Model Intercomparison Project (CMIP) Phases 5 and 6 multi-model ensembles. We show that AR statistics from a given ARDT in CMIP5/6 historical simulations compare remarkably well with reanalyses. In CMIP5/6 future simulations, most ARDTs project a global increase in AR frequency, counts, and sizes, especially along the western coastlines of the Pacific and Atlantic oceans. We find that the choice of ARDT is the dominant contributor to the uncertainty in projected AR frequency when compared with model choice. These results imply that new projects investigating future changes in ARs should explicitly consider ARDT uncertainty as a core part of the experimental design.

Plain Language Summary

Atmospheric rivers (ARs) are a type of weather pattern known to be important for moving water from the warm, moist tropics to the cool, dry polar regions; when they reach midlatitudes in the winter time, they are commonly associated with heavy rainfall. Recent studies that assess the impacts of global climate change on ARs tend to agree that there will be more ARs in a warmer climate, and that ARs will tend to be more extreme. However, it has been increasingly recognized by the AR research community that these results may depend on the method used to identify ARs and the choice of climate model. This study reports results from a controlled experiment, involving an international research community, designed to show how different AR identification methods and climate models might impact our scientific understanding of ARs in the future. This experiment shows that there will likely be more ARs in the future, and that ARs will generally have a larger spatial footprint. This experiment also shows that uncertainty in these results are large, with the uncertainty from AR identification methods outweighing that of climate models. Future efforts to better understand the physics of ARs may help us reduce this uncertainty.

1 Introduction

Over the past 30 years, research on atmospheric rivers (ARs), filamentary bands of intense water vapor transport, has increasingly demonstrated their importance for the global hydrological cycle (Newell et al., 1992; Zhu & Newell, 1998; Ralph et al., 2017) and regional energy and water cycles (Newell & Zhu, 1994; Neiman, Ralph, Wick, Kuo, et al., 2008; Ralph et al., 2005; Dettinger et al., 2011; Gimeno et al., 2016; Gershunov et al., 2017; Shields, Rosenbloom, et al., 2019). ARs are a main source of precipitation and hydroclimatological impacts in the midlatitude western margins of North America (Neiman et al., 2002; Ralph et al., 2004, 2005; Neiman, Ralph, Wick, Kuo, et al., 2008; Leung & Qian, 2009; Guan et al., 2010; Warner et al., 2012; Neiman et al., 2013; Ralph et al., 2013; Rutz et al., 2014), South America (Viale & Nuñez, 2011; Gimeno et al., 2016), Europe (Stohl et al., 2008; Lavers et al., 2012; Lavers & Villarini, 2013; Ramos et al., 2015; Gimeno et al., 2016), and South Africa (Blamey et al., 2018; Ramos et al., 2019). AR impacts on surface heat and water mass balance in polar regions are increasingly evident (Newell & Zhu, 1994; Gorodetskaya et al., 2014; Wille et al., 2019). Increased understanding of ARs has led to improvements in flood forecasting (Lavers, Waliser, et al., 2016; Lavers, Pappenberger, et al., 2016) and in communication of flood-related risks when intense ARs are imminent (Ralph, Rutz, et al., 2019).

Numerous recent studies have analyzed ARs in future climate scenarios (e.g., Warner et al., 2015; Lavers et al., 2015; Gao et al., 2015a, 2016; Shields & Kiehl, 2016b, 2016a;

Polade et al., 2017; Espinoza et al., 2018; Gershunov et al., 2019; Rhoades et al., n.d.) (see Payne et al. (2020) and references therein). Payne et al. (2020) reviews the related studies over the past 10 years and shows that (1) studies generally agree that global increases in atmospheric moisture will increase the intensity of ARs, and that (2) there is wide uncertainty in the results conveyed in the literature: especially in areas outside the well-studied U.S. west coast. Existing studies generally agree that the frequency and intensity of ARs will increase, and some studies indicate poleward shifts of the AR tracks (Sousa et al., 2020). Gershunov et al. (2019) show that intermodel differences in future projections of precipitation are much lower when considering precipitation due to ARs than those when considering changes in bulk precipitation. Given that precipitation is produced by a variety of meteorological phenomena, and that there is no guarantee that the relative proportions of precipitation from various phenomena are the same in models as they are in observations, Gershunov et al. (2019) highlight the importance in using a phenomenon-focused study of precipitation in future climate simulations.

Essentially all of the studies of ARs and future climate (and past climate, e.g., Lora et al., 2017; Skinner et al., 2020) rely on objective, quantitative methods to discriminate ARs from the background: AR detectors (ARDTs). At present, ARs have a qualitative definition (Ralph et al., 2018), which leaves researchers with the task of implementing a quantitative definition of ARs in specific ARDTs. ARDTs typically consist of a set of heuristic rules (e.g., thresholds and filters) that focus on identifying anomalously high moisture or moisture transport that occurs in contiguous, filamentary structures. The design of ARDTs is guided by understanding gained through decades of observational and model studies (Browning & Pardoe, 1973; McGuirk et al., 1987; Newell et al., 1992; Zhu & Newell, 1998; Lackmann & Gyakum, 1999; Neiman et al., 2002; Ralph et al., 2004, 2005; Bao et al., 2006; Neiman, Ralph, Wick, Kuo, et al., 2008; Neiman, Ralph, Wick, Lundquist, & Dettinger, 2008; Waliser et al., 2012). The number of ARDT algorithms has grown with the number of ARDT studies over the past decade, with new ARDTs often being developed for specialized purposes: e.g., ARDTs for understanding the global hydrological cycle (Zhu & Newell, 1998; Guan & Waliser, 2015), observed hydrometeorological extremes (Neiman, Ralph, Wick, Lundquist, & Dettinger, 2008; Rutz et al., 2014), the cryosphere (Gorodetskaya et al., 2014), and regional hydroclimate variability (Gershunov et al., 2017). Even though ARDTs are often initially designed with different purposes in mind, Payne et al. (2020) demonstrate that there is overlap in what they are ultimately used to study. The community has recently started to recognize that uncertainty associated with the numerical definition of ARs may have important implications for our understanding of ARs (Newman et al., 2012; Huning et al., 2017; Shields et al., 2018; Guan et al., 2018; Rutz et al., 2019; Ralph, Wilson, et al., 2019; Shields, Rutz, et al., 2019; Shields, Rosenbloom, et al., 2019; O'Brien, Payne, et al., 2020; O'Brien, Risser, et al., 2020, Lora et al., In Review)

The Atmospheric River Tracking Method Intercomparison Project (ARTMIP) was launched by members of the AR research community in order to systematically assess the impact of this uncertainty on our scientific understanding (Shields et al., 2018). The First ARTMIP Workshop (Shields, Rutz, et al., 2019) defined a multi-tier experimental design focusing on uncertainty in the observational record (Tier 1; Rutz et al., 2019), and uncertainty in AR variability and change (Tier 2). Two Tier 2 experiments were launched at the Second ARTMIP Workshop (Shields, Rutz, et al., 2019): the Tier 2 C20C+ experiment and the Tier 2 CMIP5/6 experiment. Both experiments are designed to elucidate the effect of uncertainty associated with ARDTs on our understanding of ARs, with the former focusing on uncertainty in regional impacts in a single high-resolution global model, and the latter focusing on the relative roles of model and ARDT-associated uncertainty. This manuscript overviews the Tier 2 CMIP5/6 experiment.

2 Data and Methods

We use data from the ARTMIP Tier 1 experiment (Shields et al., 2018; Rutz et al., 2019), which provides atmospheric river detections from multiple ARDT algorithms. All Tier 1 ARDTs run on a common set of atmospheric fields (e.g., integrated vapor transport) derived from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2; Gelaro et al., 2017). A subset of the Tier 1 algorithms have also been run on the Tier 2 input dataset described further on. The subset of algorithms run was determined by the subset of ARTMIP participants who volunteered to run their algorithms on the Tier 2 dataset; these algorithms include `ARCONNECT_v2`, `Guan_Waliser_v2`, `IDL_rel_future`, `IDL_rel_hist`, `Lora_v2`, `Mundhenk_v3`, `PNNL_v1`, and `TECA_BARD_v1.0` (see Table S1).

For the Tier 2 input dataset for ARDTs, we derive integrated water vapor (IWV), and the components of the integrated vapor transport (IVT) vector from outputs from atmosphere-ocean general circulation models associated with the Coupled Model Intercomparison Project (CMIP) 5 (Taylor et al., 2012) and 6 (Eyring et al., 2016; O'Neill et al., 2016) multi-model ensembles (hereafter referred to as CMIP5/6 when both ensembles are jointly discussed). We utilize model output from the historical simulations in both CMIP5 and CMIP6, and we utilize output from the representative concentration pathway 8.5 (RCP8.5, CMIP5) and shared socioeconomic pathways 5-8.5 experiments (SSP5-8.5, CMIP6). We utilize models that provided specific humidity q (`hus`) and wind $\overrightarrow{u}$ (`ua` and `va`) at 6-hourly intervals on native model levels (the `6hrLev` table); we further restrict the set of models to those which provide model output from the same ensemble member for both the historical and future (RCP8.5 and SSP5-8.5) simulations. At the time that the Tier 1 input dataset was constructed (in Summer 2019), we were able to access 6 models from CMIP5 (CCSM4, CSIRO-Mk3-6, CanESM2, IPSL-CM5A-LR, IPSL-CM5B-L, and NorESM1-M) and 3 models from CMIP6 (BCC-CSM2-MR, IPSL-CM6A-LR, MRI-ESM2-0; Xin et al., 2019; Yukimoto et al., 2019; Boucher et al., 2019) that satisfied these constraints (see Table S1) We focus on two 30-year time periods for assessing changes in ARs: 1981-2010 and 2070-2099. We calculate trends over the 1950-2099 period (some data are missing due to data availability and corruption issues, and years with these issues are not included in calculations; see Text S3). The models selected represent a range of horizontal resolutions (ranging from approximately 100 km to 300 km), and the RCP8.5 and SSP5-8.5 scenarios represent aggressive emission trajectories with large amounts of radiative forcing (nominally 8.5 W/m^2) by end-of-century.

The mass-weighted vertical integrals of water vapor (ρq) and water vapor transport ($\rho \overrightarrow{u} q$) are calculated from the CMIP5/6 output as:

$$\mathrm{IWV} = -\frac{1}{g}\sum_{k=1}^{N} q_k \Delta p_k \quad (1)$$

$$\overrightarrow{\mathrm{IVT}} = -\frac{1}{g}\langle \sum_{k=1}^{N} u_k q_k \Delta p_k, \sum_{k=1}^{N} v_k q_k \Delta p_k \rangle, \quad (2)$$

where index k corresponds to model levels going from the surface ($k = 1$) to the top of the model atmosphere ($k = N$), and Δp_k is the difference in level pressures, estimated at level k. The total vapor transport is calculated as the vector magnitude: $\mathrm{IVT} = \left|\overrightarrow{\mathrm{IVT}}\right|$.

These ARDTs consist of a mixture of algorithms that detect ARs globally (global algorithms) and algorithms designed for specific regions (regional algorithms); see Table S1. We focus most of the analysis in this manuscript on the location of the AR tracks, changes in these tracks, and uncertainty therein. We therefore focus the bulk of the discussion on the global subset of algorithms; the full set of algorithms is discussed in Section 3.3 when comparing the relative magnitudes of uncertainty related to ARDT design and model choice.

2.1 Tier 2 CMIP5/6 Experiment Overview

All Tier 2 CMIP5/6 ARDT contributions use the common dataset of IWV, IVT, and $\overrightarrow{\mathrm{IVT}}$ described in Section 2, which come from 9 models in the CMIP5 and CMIP6 multi-model ensembles. ARDT outputs are regridded to a common 4°x5° analysis grid. We assess the CMIP5/6 models by comparing December-January-February (DJF) spatial patterns of AR frequency between the Tier 1 and Tier 2 experiments, for each detection scheme independently: focusing on spatial pattern correlation and spatial variability. Given the 6-hourly frequency of the dataset, we report frequency as 'equivalent' AR days, which we define as 0.25 times the total number of timesteps with AR conditions. We provide details about Tier 2-specific modifications to ARDTs in Text S1 and details about missing data in Text S3.

Grouping algorithms by the type of criteria applied (relative versus absolute thresholds) and degree of restrictiveness (magnitude of thresholds employed, number of criteria involved) can reduce the spread associated with ARDTs (Rutz et al., 2019; Ralph, Wilson, et al., 2019). Here, we group ARDTs into three categories, based on their treatment of thresholds: *absolute* (`ARCONNECT_v2`, `PNNL_v1`, and `Lora_v2`), *fixed relative* (`Guan_Waliser_v2`, `IDL_rel_future`, `IDL_rel_hist`, and `Mundhenk_v3`), and *relative* (`Tempest` and `TECA_BARD_v1.0`). The categorizations are described and justified in Text S2. A key motivation for this categorization is aggregating ARDTs by their sensitivity to thermodynamic changes in IVT, with the assumption that ARDTs employing absolute thresholds to moisture fields will be the most sensitive, and ARDTs employing time-dependent thresholds will be least sensitive.

3 Results

3.1 Evaluation of Historical Simulations

We show maps of DJF average AR frequency from the Tier 1 (MERRA-2) experiments for the 6 global ARDT algorithms in the top row of Figure 1. The ARDTs show broad consistency in the spatial patterns of DJF ARs. All ARDTs identify well-known AR tracks, with distinct maxima in the Pacific and the Atlantic, and with a circumglobal maximum in the Southern Ocean. The ARDTs also identify significant areas with little or no AR activity: the tropics, northeastern Asia, Africa, and the subtropical eastern Pacific (near the cold tongue region). The ARDTs differ significantly in the relative frequency of AR conditions. Some of the ARDTs identify AR conditions occurring upwards of 30 days per season (approximately one third of the time) in the main AR tracks, and other ARDTs identify AR conditions occurring fewer than 10 days per season. These results are consistent with previous ARDT comparisons, indicating a wide range of restrictiveness across ARDTs (Ralph, Wilson, et al., 2019; Rutz et al., 2019). The algorithms also differ in the degree to which the AR tracks penetrate inland, with the `Guan_Waliser_v2` algorithm commonly identifying ARs in continental interiors, and `TECA_BARD_v1.0` rarely identifying ARs in continental interiors. The average frequency of ARs (the top-right panel in Figure 1) exhibits a similar spatial pattern to the various ARDTs, with ARs occurring approximately 10 days per season in the core AR track.

Simulated ARs in the Tier 2 CMIP5/6 experiment are remarkably consistent with those in the Tier 1 MERRA-2 experiment. Results from an arbitrary model–MRI-ESM-2-0 from the CMIP6 multimodel ensemble–are shown in the second row of Figure 1, and a similar plot showing results from all possible model-ARDT pairs is shown in Figure S1. The placement of the AR tracks (and opposing gaps in ARs) are very similar when comparing spatial maps for a given ARDT. The algorithm-mean AR frequencies (last column) show very little difference between Tier 1 and 2; this is true for all models analyzed (see Figure S1).

Each ARDT has idiosyncratic spatial patterns that are expressed in both Tier 1 and Tier 2. This suggests that the spatial pattern maps are an emergent property of each ARDT, and that these spatial patterns are relatively insensitive to significant changes in the representation of the underlying atmospheric dynamics. For example, the diffuse spatial pattern associated with the `Guan_Waliser_v2` (GW) algorithm is evident in Tier 1 and in all Tier 2 simulations (Figures S1 and S2), and the multi-model mean for the GW algorithm exhibits a similar spatial pattern. This suggests that there is much more variability in AR frequency across ARDT algorithms than there is across simulations; we quantify this in Section 3.3.

Figure 2a quantitatively shows that simulations compare well with observations when compared within a single ARDT. Spatial correlation coefficients between the DJF AR frequency maps in individual Tier 2 simulations and the corresponding Tier 1 map are consistently around $r \approx 0.95$, and the ratio of spatial standard deviations of DJF AR frequency (Tier 2 divided by Tier 1) is close to 1 for many ARDT-model pairs. The Taylor skill scores (Taylor, 2001) are above 0.95 for most ARDT-model pairs. Variability exists, with some ARDT-model pairs reaching as high as $r \approx 0.97$ and only 5 ARDT-model pairs with correlation coefficients between 0.8 and 0.9 (and skill scores below 0.85); likewise, one combination (`ARCONNECT_v2` and CMIP5 IPSL-CM5A-LR) has variability that is too low by approximately 25%, and one combination (`Tempest` and CMIP5 IPSL-CM5B-LR) has variability that is about 50% too high. Overall, this emphasizes the high degree of similarity between simulated ARs and ARs in MERRA-2, when comparing results using a single ARDT.

Altogether, the various ARDTs portray a similar assessment of model skill, with essentially all of the models analyzed appearing to be 'fit for purpose'. This is true even for the lowest resolution simulations (e.g., CMIP5 CanESM2 with a nominal 310 km horizontal resolution in the tropics; see Table S1), which have some of the highest correlation coefficients. A survey of the literature (Gao et al., 2015b; Hagos et al., 2015; Shields & Kiehl, 2016b; Guan & Waliser, 2017; Payne et al., 2020; Rhoades et al., 2020) indicates a mix of possible resolution effects, with some indication that the effect of resolution may depend on the experimental setup (e.g., coupled vs uncoupled; Guan & Waliser, 2017). We hypothesize that resolution effects may depend on the ARDT used; these effects could be studied more systematically by applying multiple ARDTs to the CMIP6 HighResMIP experiment (Haarsma et al., 2016). The ARTMIP community has discussed the possibility of coordinating a Tier 2 Resolution experiment (O'Brien, Payne, et al., 2020) to explore this more systematically.

Results associated with the `Tempest` algorithm are a somewhat notable exception: five of the models evaluated with `Tempest` have high spatial variability relative to MERRA-2, and relatively low spatial correlations. This may be related to some differences in the implementation of `Tempest` between the Tier 1 and Tier 2 experiments (see Text S1).

3.2 Projected Changes in AR Frequency, Count, and Size

When the ARDTs are applied to the various future simulations described in Section 2, they project a variety of trends in AR frequency relative to the historical counterparts evaluated in Section 3.1. Figure 1 (third row) shows that most ARDTs applied to the MRI-ESM2-0 simulation indicate increases in AR frequency in the main AR tracks. Within each algorithm, the trends from the MRI-ESM2-0 simulation are quantitatively and qualitatively similar to trends from other simulations (see Figure S3), as indicated by the similarity between the MRI-ESM2-0 trends and the multi-model trends shown in the bottom row of Figure 1. The average trend across all model-ARDT combinations (lower right panel of Figure 1) likewise indicates an increase in AR frequency in the midlatitude storm tracks, with increases of $\sim$5 AR days per century (an approximate 50%

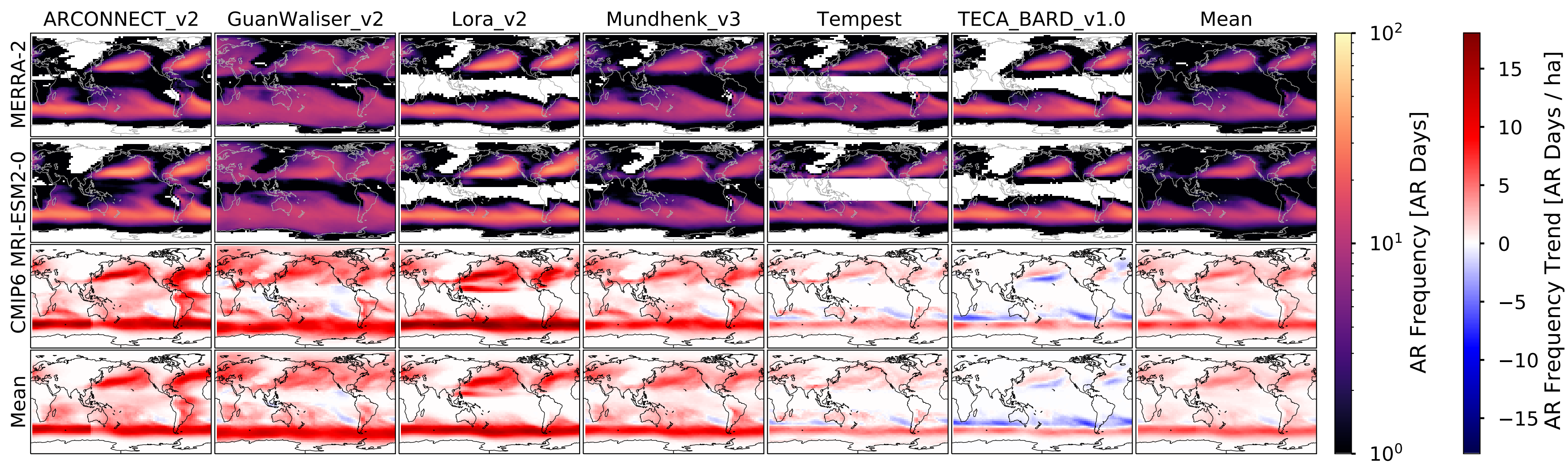

Figure 1. (first and second rows) Maps of AR frequency (shown as average number of days with AR conditions) in DJF for the 1981-2010 period. Each column corresponds to a global AR detection algorithm, and the last column represents the average across all AR detection algorithms. The top row corresponds to AR detections on the MERRA-2 dataset (the Tier 1 ARTMIP experiment) and the second row corresponds to AR detections on the CMIP6 MRI-ESM-2-1 simulation. White indicates areas where average AR occurrence is fewer than 1 day. (third and fourth rows) Maps of trends in DJF AR frequency in the MRI-ESM-2-1 simulation (third row) and all models (fourth row), organized by detection algorithm (columns).

increase). There are essentially no areas where the model-ARDT mean indicates a decrease in AR frequency.

The climatological pattern of AR frequency is primarily controlled by changes in AR size, AR occurrence (count), and AR location. Two ARDTs (`TECA_BARD_v1.0` and to a lesser extent `Tempest`) suggest poleward shifts in AR location (Figure 1, bottom row, and Figure S3), whereas `ARCONNECT_v2`, `GuanWaliser_v2`, `Lora_v2`, and `Mundhenk_v3` indicate quasi-global increases in AR frequency. We discuss why differences in the quantitative definition of ARs may cause different behavior in future climate simulations and its implications in Section 4.

We decompose the changes in AR frequency by changes in AR area A and AR count N; Figure 2b shows the median size of AR objects versus the median number of AR objects counted at any given time. In the historical simulations, the ARDTs appear to cluster along a continuum, with ARDTs typically detecting 5–20 ARs, which is consistent with manual counts of ARs in synoptic maps (Zhu & Newell, 1998; O'Brien, Risser, et al., 2020). `Tempest` is a notable exception, with AR counts ranging from 20–50. In order to aid in interpreting the continuum along which the ARDTs lay in Figure 2b, we add lines of constant global area $A_\oplus$ percentage (calculated as $100\% \cdot A \cdot N / A_\oplus$). These show that algorithms typically detect ARs such that approximately 5% of the Earth's surface is covered in AR objects in the historical simulations. Therefore, we can interpret the relative location of ARDTs in Figure 2b as an indicator of the relative spatial coherence of AR objects: ARDTs on the left detect few, large AR objects and ARDTs on the right detect many small AR objects. This grouping along lines of constant global area fraction is an emergent collective behavior of the ARDTs, and we speculate that it is associated with the tuning process for each algorithm. AR coherence might make a useful measure for objective grouping of AR results in future ARTMIP studies.

Figure 2b shows that four of the ARDTs (except `Tempest` and `TECA_BARD_v1.0`) tend to detect more ARs and larger ARs in the future simulations. These changes result in increases in the global area coverage of AR objects: changing from $\sim$5% global area to $\sim$7% global area. The global count of AR objects does not change in the `TECA_BARD_v1.0` algorithm, though there are slight increases in AR area in some simulations. In contrast, the `Tempest` algorithm indicates increases in global AR count, with very little change in AR area.

There is an indication that the resolution of the underlying model may affect the characteristics of detected ARs for some ARDTs. The CMIP6 BCC-CSM2-MR, CMIP6 MRI-ESM2-0, and CMIP5 CCSM4 simulations–which are the three highest resolution simulations analyzed (Table S1)–tend to occur on the right side of each ARDT cluster: ARs in these simulations are systematically less coherent. However, the model resolution does not appear to affect the climate change signal evident in Figure 2b. Further, the CMIP5/6 simulations analyzed here do not attempt to control for model resolution; the CMIP6 HighResMIP experiment (Haarsma et al., 2016) could provide a way to examine resolution effects more systematically.

3.3 Sources of Uncertainty in End-of-Century Projections of ARs

The results in Figure 1 indicate that there may be substantial uncertainty in future AR frequency associated with choice of ARDT. Further, it is not clear from the spatial maps in Figure 1 whether the trends in AR frequency evident over the ocean (e.g., the decrease in the southeastern Atlantic) extend to the coastal areas where AR presence matters for western-coastal water cycles and hydrometeorological impacts. We quantify these changes and their uncertainty in Figure 2c,d, which show the mean trend in AR frequency for the Pacific (Figure 2c) and Atlantic west coasts (Figure 2d) from 1980-2099. Figure 2c,d shows trends for all ARDTs listed in Table S1: both regional and global ARDTs.

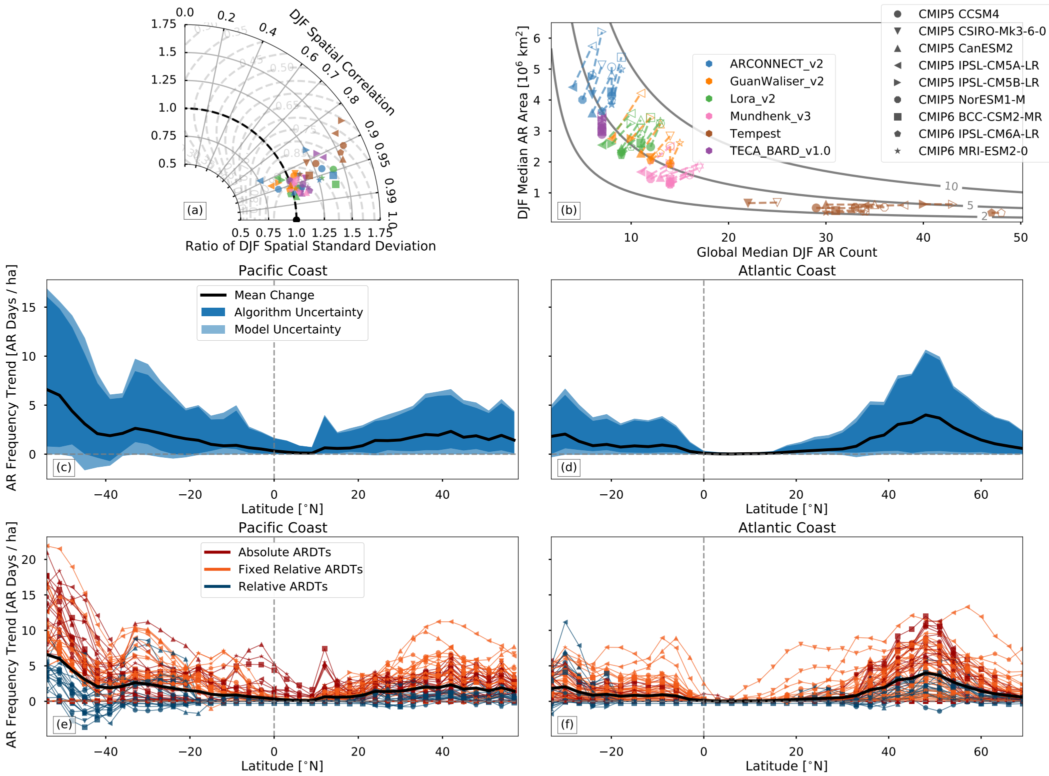

Figure 2. (a) A Taylor diagram comparing the spatial correlation (azimuthal axis) and spatial variability (radial axis) of DJF AR frequency between CMIP5 and 6 simulations (denoted by different symbols) and the MERRA-2 reanalysis. Colors indicate different AR detection algorithms (legend in panel b). Gray dashed lines show lines of constant skill score (Taylor, 2001). (b) Median AR area vs global median AR count for all available combinations of ARDTs (marker colors) and simulations (marker symbols). Filled symbols indicate calculations performed on the 1981-2010 period of each simulation, and open symbols indicate calculations on the 2070-2099 period (two exceptions noted in Text S3). Gray contours show lines of constant fractional areal coverage of ARs (shown as a percentage of Earth's area), calculated as the product of AR area and AR count, divided by Earth's area. (c and d) Trends in AR frequency (black curve) and associated total range of uncertainty (blue and light blue shading) for the west-facing (c) Pacific coastline and (d) Atlantic coastline. Dark blue shading indicates the portion of uncertainty associated with AR detection and the light blue shading indicates the portion of the spread associated with models (across both CMIP5 and CMIP6). The area of dark blue shading is proportional to $\sigma_A^2/\sigma_T^2 \cdot$ (max $-$ min), where 'max' and 'min' are the minimum and maximum trend at each latitude. (e and f) as in (c and d), but showing indivdual ARDT-model combinations. Markers indicate simulations (legend in panel b) and colors indicate the ARDT classification.

Figure 2c,d shows that coastal areas in both the Pacific and Atlantic show increasing trends in AR frequency (+2–5 AR days per century in the midlatitudes), and the full spread of the blue and light blue shading in Figure 2c,d shows the full range of trends from all ARDTs and all models. There are two areas where `TECA_BARD_v1.0` indicates weakly decreasing trends (Figure S4 shows the trends by model and by algorithm): southern Chile, near 40°S, and near the entrance of the Mediterranean Sea, at about 35°N. Otherwise all model-ARDT combinations indicate increasing trends in landfalling AR frequency for both Pacific and Atlantic ARs in both hemispheres.

Large uncertainty appears in the magnitude of the trends, which ranges from just below 0 days/ha to over 15 days/ha, depending on location. There are two main components of uncertainty in these trends: uncertainty associated with choice of model simulation, and uncertainty associated with choice of ARDT. We decompose the uncertainty as $\sigma_T^2 \approx \sigma_A^2 + \sigma_M^2$, where σ_T^2 is the total variance, σ_A^2 is the variance across ARDTs of each ARDT's multi-model mean, and σ_M^2 is the variance across models for each model's multi-ARDT mean. These variances can equivalently be viewed as the variance down the rightmost column in Figure S3 (σ_M^2) and the variance across the bottommost row in Figure S3 (σ_A^2), (excluding the multi-model/multi-ARDT mean in the bottom right corner of Figure S3 and excluding trends from MERRA-2).

This decomposition shows that uncertainty associated with choice of ARDT accounts for most of the spread in the climate change signal across all latitudes in both the Pacific and Atlantic coasts. In essence, uncertainty associated with the numerical definition of ARs dominates the combined uncertainty associated with choice of model and choice of model epoch (CMIP5 vs CMIP6).

4 Discussion and Conclusions

The ARTMIP Tier 2 experiments show that most ARDTs project an increase in AR frequency, with mean trends of approximately +2-5 AR days per century along the coastlines of North America, South America, Southern Africa, and Europe (Figure 2c,d). However, there is considerable spread in the magnitude, with some ARDT-model combinations indicating negative trends (southern Chile and southern Spain) and other ARDT-model combinations indicating trends of up to ∼15 AR days per century. Given that we have no *a priori* reason to prefer one ARDT over another, we therefore cannot rule out the possibility that there will be decreases or no changes in AR frequency in a future climate.

Globally, all ARDTs indicate either an increase in the total number of ARs, an increase in the areal extent of ARs, or both (Figure 2b). In the historical simulations, the AR area vs size relationship for all ARDTs approximately falls along a line of constant global coverage, with ARDTs in the historical simulations detecting ARs that cover approximately 5% of the global area. This number is somewhat smaller than the 10% figure indicated by Zhu and Newell (1998), which is likely because we are considering the total global coverage, including the tropics, rather than the fraction of zonal circumference in the midlatitudes. But it is qualitatively consistent in the sense that areas of anomalously high moisture transport occupy a small fraction of the global area. The global areal coverage increases in the future simulations to some degree in all ARDT algorithms, with most indicating a several percent increase in the areal extent of ARs due to increases in both AR size and count.

These results further show that future changes in AR frequency can qualitatively differ depending on the type of ARDT used. We aggregate trends by AR classification (see Sections 2 and Text S2) in Figures 2e,f. This aggregation shows that use of any absolute thresholds (*absolute ARDTs*) and time-independent relative thresholds (*fixed relative ARDTs*) tend to produce increases in AR frequency, whereas use of time-dependent relative thresholds (*relative ARDTs*) tend to produce patterns more indicative of a poleward shift. *Absolute ARDTs* and *fixed relative ARDTs*, with thresholds that do not change in time, would be expected to increase the frequency of exceedence of regions above the historical thresholds: more detected AR days in a warmer climate. Such ARDTs are designed to detected increases in occurrence of regions with high IVT, which are important for AR impacts. In contrast, *relative ARDTs* (e.g., `TECA_BARD_v1.0`) are designed to only account for dynamical changes in ARs.

The literature documents two major modes of AR change associated with climate change: (1) a quasi-global increase in IVT associated with Clausius-Clapeyron scaling (thermodynamic; Payne et al., 2020), and (2) a poleward shift in ARs (dynamic; Payne et al., 2020) associated with the poleward shift in the midlatitude storm tracks (Chang et al., 2012). Poleward shift patterns appear to co-exist to some extent with quasi-global increases in AR frequency in some simulations (e.g., the CMIP5 CSIRO-MK3-6-0 simulation; see Figure S3) for all ARDTs. We argue that *absolute ARDTs* and *fixed relative ARDTs* are more sensitive to thermodynamic changes than *relative ARDTs*. It is worth noting here that trend patterns in the MERRA-2 reanalysis are remarkably similar across ARDTs (Figure S3), with all ARDTs indicating a poleward shift in ARs. This might suggest that the observed poleward shift in the storm tracks (Fyfe, 2003) dominates over quasi-global increases in IVT in the historical record. This should be investigated further as part of the Tier 2 Reanalysis experiment.

The algorithm-wise validation of simulated ARs (Figure 2a) shows that models replicate AR statistics from reanalysis remarkably well. This is a noteworthy result in the context of the ARDT uncertainty shown here. If only one algorithm is used in a study, such validation could give false confidence in the robustness of results. It therefore seems important to explicitly include ARDT uncertainty as part of evaluation of a model's ability to represent ARs, which, relatedly, points to the utility of appropriate ensemble weighting strategies to help reduce such uncertainty (e.g., Massoud et al., 2019). It also highlights the value of AR-related, but not ARDT-dependent, evaluations of models (e.g., Payne & Magnusdottir, 2015).

Recent work involving manual identification of ARs by experts (Prabhat et al., 2020; O'Brien, Risser, et al., 2020) suggests that the spread in AR algorithm behavior may be linked to differences in opinion about what does and does not constitute an AR. O'Brien, Risser, et al. (2020) show that this spread in subjective opinion projects directly on to quantitative differences in the sign of the correlation coefficient between an El Niño index and global AR count. Such differences in subjective opinion likely also play a role in the design of the quantitative choices made by various ARDT designers.

Somewhat relatedly, the ARTMIP project has established that different AR detectors are designed with different–and equally legitimate–purposes (Shields et al., 2018; Rutz et al., 2019; Ralph, Wilson, et al., 2019). Some ARDTs intentionally choose to discriminate ARs from the background based on absolute thresholds in IVT (e.g., Rutz et al., 2014), since it is well-established that coastal orographic precipitation is directly linked to IVT magnitude (Neiman et al., 2002; Ralph et al., 2004, 2005; Neiman, Ralph, Wick, Kuo, et al., 2008; Ralph, Rutz, et al., 2019); such a design choice makes it easy to relate ARDT results directly to hydrometeorological impacts. Other algorithms (e.g., Shields & Kiehl, 2016b; O'Brien, Risser, et al., 2020) intentionally use relative thresholds in order to avoid increases in AR detection due to long-term increases in atmospheric water vapor. Both are valid for the purposes for which they were designed: absolute methods detect areas that will likely lead to hydrometeorological impacts–which will increase in a warmer climate–and relative methods seek to focus on the core of regions associated with anomalous vapor transport.

These results suggest that new projects investigating future changes in the statistics and characteristics of ARs should explicitly consider ARDT uncertainty as a core part of the experimental design. This study makes it clear that ARDT design choices can have a major impact on the results of climate change studies. The Bayesian, multi-ARDT approach of O'Brien, Risser, et al. (2020) can quantify parametric uncertainty associated with a single ARDT, but it is not yet clear how parametric uncertainty compares to structural uncertainty (i.e., choices in what heuristic rules to employ in the ARDT). There are at least four ARDT codes that are now in the public domain (`Mundhenk_v1`, `Guan_Waliser_v2`, `Tempest`, and `TECA_BARD_v1.0`), and we encourage current and future ARDT designers to likewise enter their codes into the public domain in order to facilitate such uncertainty exploration in future studies.

Ralph et al. (2018) provide a concise, qualitative definition of ARs, and this has been a major benefit to the AR research community. They intentionally chose to "leave specifications of how the boundaries of an AR are to be quantified open for future and specialized developments." The results in this manuscript demonstrate that the choice of how to define AR boundaries–the fundamental job of an ARDT–have a demonstrably large control on the statistics of ARs detected in future climate simulations. These results suggest that the AR research community would further benefit from studies that aim to quantitatively constrain the definition of ARs; e.g., with first-principles analyses that constrain AR properties like size, count, etc. Such constraints could help reduce uncertainty associated with ARDT design choice (and parameter choice), and by extension they could constrain results concerning ARs and future climate change. That said, given that different experiments motivate different ARDT design choices (e.g., absolute vs relative thresholds), it seems unavoidable that some of this uncertainty is irreducible. It is clear, however, that it is imperative for studies to explore and understand the implications of this uncertainty.

This study focuses on a bulk, global perspective of uncertainty associated with ARDTs and simulations in the Tier 2 CMIP5/6 experiment. There are many other types of more detailed analyses that others could take on, and we encourage others in the research community to utilize this dataset for research on future ARs and climate change (see data availability statement in Acknowledgements). In particular, it seems valuable to revisit past studies of ARs and future climate change in the context of ARDT uncertainty. Payne et al. (2020) review the numerous results concerning the future of ARs that have appeared in the literature in the last decade. There are almost as many ARDTs as there are such results, which makes intercomparison of the results challenging. The Tier 2 CMIP5/6 dataset provides a way to revisit many–if not all–of these previous results within a uniform experimental framework.

In summary, this initial analysis of the Tier 2 CMIP5/6 experiment shows that most ARDTs and simulations indicate an increasing trend in AR frequency, size, and number in future simulations with strong radiative forcing. It also shows the critical importance of understanding the implications of this uncertainty for AR-related research.

Acknowledgments

This research was supported by the Director, Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy Regional and Global Climate Modeling Program (RGMA) and used resources of the National Energy Research Scientific Computing Center (NERSC), also supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, 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 who support CMIP6 and ESGF. We thank DOE's RGMA program area, the Data Management program, and NERSC for making this coordinated CMIP6 analysis activity possible.

Alexandre M. Ramos is supported by the Fundação para a Ciência e Tecnologia (FCT, Portugal) project "Weather Extremes in the Euro Atlantic Region: Assessment and Impacts-WEx-Atlantic" (PTDC/CTA-MET/29233/2017). Alexandre M.Ramos also acknowledges the Scientific Employment Stimulus 2017 from FCT (CEECIND/00027/2017)

ARTMIP Tier 2 CMIP5/6 catalogues can be found on the Climate Data Gateway: doi.org/10.26024/s4p7-pf13 (`https://go.iu.edu/3ci2`)

References

Bao, J. W., Michelson, S. A., Neiman, P. J., Ralph, F. M., & Wilczak, J. M. (2006). Interpretation of enhanced integrated water vapor bands associated with extratropical cyclones: Their formation and connection to tropical moisture. *Monthly Weather Review*, *134*(4), 1063–1080. doi: 10.1175/MWR3123.1

Blamey, R. C., Ramos, A. M., Trigo, R. M., Tomé, R., & Reason, C. J. (2018). The influence of atmospheric rivers over the South Atlantic on winter rainfall in South Africa. *Journal of Hydrometeorology*, *19*(1), 127–142. doi: 10.1175/JHM-D-17-0111.1

Boucher, O., Denvil, S., Caubel, A., & Foujols, M. A. (2019). *IPSL IPSL-CM6A-LR model output prepared for CMIP6 ScenarioMIP ssp585.* Earth System Grid Federation. Retrieved from `https://doi.org/10.22033/ESGF/CMIP6.5271` (Version 20180803) doi: 10.22033/ESGF/CMIP6.5271

Browning, K. A., & Pardoe, C. W. (1973). Structure of low-level jet streams ahead of mid-latitude cold fronts. *Quarterly Journal of the Royal Meteorological Society*, *99*(422), 619–638. doi: 10.1002/qj.49709942204

Chang, E. K., Guo, Y., & Xia, X. (2012). CMIP5 multimodel ensemble projection of storm track change under global warming. *Journal of Geophysical Research Atmospheres*, *117*(23), 1–19. doi: 10.1029/2012JD018578

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. doi: 10.3390/w3020445

Espinoza, V., Waliser, D. E., Guan, B., Lavers, D. A., & Ralph, F. M. (2018, may). Global Analysis of Climate Change Projection Effects on Atmospheric Rivers. *Geophysical Research Letters*, *45*(9), 4299–4308. Retrieved from `http://doi.wiley.com/10.1029/2017GL076968` doi: 10.1029/2017GL076968

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

Fyfe, J. C. (2003). Extratropical Southern Hemisphere cyclones: Harbingers of climate change? *Journal of Climate*, *16*(17), 2802–2805. doi: 10.1175/1520-0442(2003)016⟨2802:ESHCHO⟩2.0.CO;2

Gao, Y., Lu, J., & Leung, L. R. (2016, 09). Uncertainties in Projecting Future Changes in Atmospheric Rivers and Their Impacts on Heavy Precipitation over Europe. *Journal of Climate*, *29*(18), 6711-6726. Retrieved from `https://doi.org/10.1175/JCLI-D-16-0088.1` doi: 10.1175/JCLI-D-16-0088.1

Gao, Y., Lu, J., Leung, L. R., Yang, Q., Hagos, S., & Qian, Y. (2015a). Dynamical and thermodynamical modulations on future changes of landfalling atmospheric rivers over western north america. *Geophysical Research Letters*, *42*(17), 7179-7186. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015GL065435` doi: 10.1002/2015GL065435

Gao, Y., Lu, J., Leung, L. R., Yang, Q., Hagos, S., & Qian, Y. (2015b, sep). Dynamical and thermodynamical modulations on future changes of landfalling atmospheric rivers over western North America. *Geophysical Research Letters*, *42*(17), 7179–7186. Retrieved from http://doi.wiley.com/10.1002/2015GL065435 doi: 10.1002/2015GL065435

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., ... Zhao, B. (2017, jul). The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. Retrieved from http://journals.ametsoc.org/doi/10.1175/JCLI-D-16-0758.1 doi: 10.1175/JCLI-D-16-0758.1

Gershunov, A., Shulgina, T., Clemesha, R. E. S., Guirguis, K., Pierce, D. W., Dettinger, M. D., ... Ralph, F. M. (2019, dec). Precipitation regime change in Western North America: The role of Atmospheric Rivers. *Scientific Reports*, *9*(1), 9944. Retrieved from http://www.nature.com/articles/s41598-019-46169-w doi: 10.1038/s41598-019-46169-w

Gershunov, A., Shulgina, T., Ralph, F. M., Lavers, D. A., & Rutz, J. J. (2017, aug). Assessing the climate-scale variability of atmospheric rivers affecting western North America. *Geophysical Research Letters*, *44*(15), 7900–7908. Retrieved from http://doi.wiley.com/10.1002/2017GL074175 doi: 10.1002/2017GL074175

Gimeno, L., Dominguez, F., Nieto, R., Trigo, R., Drumond, A., Reason, C. J., ... Marengo, J. (2016). Major Mechanisms of Atmospheric Moisture Transport and Their Role in Extreme Precipitation Events. *Annual Review of Environment and Resources*, *41*(1), 117–141. doi: 10.1146/annurev-environ-110615-085558

Gorodetskaya, I. V., Tsukernik, M., Claes, K., Ralph, M. F., Neff, W. D., & Van Lipzig, N. P. M. (2014, sep). The role of atmospheric rivers in anomalous snow accumulation in East Antarctica. *Geophysical Research Letters*, *41*(17), 6199–6206. Retrieved from http://doi.wiley.com/10.1002/2014GL060881 doi: 10.1002/2014GL060881

Guan, B., Molotch, N. P., Waliser, D. E., Fetzer, E. J., & Neiman, P. J. (2010, oct). Extreme snowfall events linked to atmospheric rivers and surface air temperature via satellite measurements. *Geophysical Research Letters*, *37*(20), n/a–n/a. Retrieved from http://doi.wiley.com/10.1029/2010GL044696 doi: 10.1029/2010GL044696

Guan, B., & Waliser, D. E. (2015, dec). Detection of atmospheric rivers: Evaluation and application of an algorithm for global studies. *Journal of Geophysical Research: Atmospheres*, *120*(24), 12514–12535. Retrieved from http://doi.wiley.com/10.1002/2015JD024257 doi: 10.1002/2015JD024257

Guan, B., & Waliser, D. E. (2017). Atmospheric rivers in 20 year weather and climate simulations: A multimodel, global evaluation. *Journal of Geophysical Research*, *122*(11), 5556–5581. doi: 10.1002/2016JD026174

Guan, B., Waliser, D. E., & Ralph, F. M. (2018, 02). An Intercomparison between Reanalysis and Dropsonde Observations of the Total Water Vapor Transport in Individual Atmospheric Rivers. *Journal of Hydrometeorology*, *19*(2), 321-337. Retrieved from https://doi.org/10.1175/JHM-D-17-0114.1 doi: 10.1175/JHM-D-17-0114.1

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., ... von Storch, J.-S. (2016, nov). High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6. *Geoscientific Model Development*, *9*(11), 4185–4208. Retrieved from http://www.geosci-model-dev.net/9/4185/2016/https://gmd.copernicus.org/articles/9/4185/2016/ doi: 10.5194/gmd-9-4185-2016

Hagos, S., Ruby Leung, L., Yang, Q., Zhao, C., & Lu, J. (2015). Resolution and dynamical core dependence of atmospheric river frequency in global model simulations. *Journal of Climate*, *28*(7), 2764–2776. doi:

10.1175/JCLI-D-14-00567.1

Huning, L. S., Margulis, S. A., Guan, B., Waliser, D. E., & Neiman, P. J. (2017). Implications of Detection Methods on Characterizing Atmospheric River Contribution to Seasonal Snowfall Across Sierra Nevada, USA. *Geophysical Research Letters*, *44*(20), 10,445–10,453. doi: 10.1002/2017GL075201

Lackmann, G. M., & Gyakum, J. R. (1999). Heavy cold-season precipitation in the northwestern United States: Synoptic climatology and an analysis of the flood of 17-18 January 1986. *Weather and Forecasting*, *14*(5), 687–700. doi: 10.1175/1520-0434(1999)014⟨0687:HCSPIT⟩2.0.CO;2

Lavers, D. A., Pappenberger, F., Richardson, D. S., & Zsoter, E. (2016). ECMWF Extreme Forecast Index for water vapor transport: A forecast tool for atmospheric rivers and extreme precipitation. *Geophysical Research Letters*, *43*(22), 11,852–11,858. doi: 10.1002/2016GL071320

Lavers, D. A., Ralph, F. M., Waliser, D. E., Gershunov, A., & Dettinger, M. D. (2015, jul). Climate change intensification of horizontal water vapor transport in CMIP5. *Geophysical Research Letters*, *42*(13), 5617–5625. Retrieved from `http://doi.wiley.com/10.1002/2015GL064672` doi: 10.1002/2015GL064672

Lavers, D. A., & Villarini, G. (2013). The nexus between atmospheric rivers and extreme precipitation across Europe. *Geophysical Research Letters*, *40*(12), 3259–3264. doi: 10.1002/grl.50636

Lavers, D. A., Villarini, G., Allan, R. P., Wood, E. F., & Wade, A. J. (2012, oct). The detection of atmospheric rivers in atmospheric reanalyses and their links to British winter floods and the large-scale climatic circulation. *Journal of Geophysical Research Atmospheres*, *117*(20), 1–13. Retrieved from `http://doi.wiley.com/10.1029/2012JD018027` doi: 10.1029/2012JD018027

Lavers, D. A., Waliser, D. E., Ralph, F. M., & Dettinger, M. D. (2016). Predictability of horizontal water vapor transport relative to precipitation: Enhancing situational awareness for forecasting western U.S. extreme precipitation and flooding. *Geophysical Research Letters*, *43*(5), 2275–2282. doi: 10.1002/2016GL067765

Leung, L.-R., & Qian, Y. (2009). Atmospheric rivers induced heavy precipitation and flooding in the western U.S. simulated by the WRF regional climate model. *Geophysical Research Letters*, *36*(3), 1–6. doi: 10.1029/2008GL036445

Lora, J. M., Mitchell, J. L., Risi, C., & Tripati, A. E. (2017, jan). North Pacific atmospheric rivers and their influence on western North America at the Last Glacial Maximum. *Geophysical Research Letters*, *44*(2), 1051–1059. Retrieved from `http://doi.wiley.com/10.1002/2016GL071541https://onlinelibrary.wiley.com/doi/abs/10.1002/2016GL071541` doi: 10.1002/2016GL071541

Massoud, E., Espinoza, V., Guan, B., & Waliser, D. (2019). Global climate model ensemble approaches for future projections of atmospheric rivers. *Earth's Future*, *7*(10), 1136-1151. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019EF001249` doi: 10.1029/2019EF001249

McGuirk, J. P., Thompson, A. H., & Smith, N. R. (1987, apr). Moisture bursts over the tropical Pacific Ocean. *Monthly Weather Review*, *115*(4), 787–798. Retrieved from `http://journals.ametsoc.org/doi/abs/10.1175/1520-0493{\%}281987{\%}29115{\%}3C0787{\%}3AMBOTTP{\%}3E2.0.CO{\%}3B2` doi: 10.1175/1520-0493(1987)115⟨0787:MBOTTP⟩2.0.CO;2

Neiman, P. J., Ralph, F. M., White, A. B., Kingsmill, D. E., & Persson, P. O. (2002). The statistical relationship between upslope flow and rainfall in California's coastal mountains: Observations during CALJET. *Monthly Weather Review*, *130*(6), 1468–1492. doi: 10.1175/1520-0493(2002)130⟨1468:TSRBUF⟩2.0.CO;2

Neiman, P. J., Ralph, F. M., Wick, G. A., Kuo, Y. H., Wee, T. K., Ma, Z., … Dettinger, M. D. (2008). Diagnosis of an intense atmospheric river impacting the pacific northwest: Storm summary and offshore vertical structure observed with COSMIC satellite retrievals. *Monthly Weather Review*, *136*(11), 4398–4420. doi: 10.1175/2008MWR2550.1

Neiman, P. J., Ralph, F. M., Wick, G. A., Lundquist, J. D., & Dettinger, M. D. (2008). Meteorological characteristics and overland precipitation impacts of atmospheric rivers affecting the West coast of North America based on eight years of SSM/I satellite observations. *Journal of Hydrometeorology*, *9*(1), 22–47. doi: 10.1175/2007JHM855.1

Neiman, P. J., Ralph, M. F., 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. doi: 10.1175/JHM-D-12-0101.1

Newell, R. E., Newell, N. E., Zhu, Y., & Scott, C. (1992, dec). Tropospheric rivers? - A pilot study. *Geophysical Research Letters*, *19*(24), 2401–2404. Retrieved from `http://doi.wiley.com/10.1029/92GL02916` doi: 10.1029/92GL02916

Newell, R. E., & Zhu, Y. (1994, jan). Tropospheric rivers: A one-year record and a possible application to ice core data. *Geophysical Research Letters*, *21*(2), 113–116. Retrieved from `http://doi.wiley.com/10.1029/93GL03113` doi: 10.1029/93GL03113

Newman, M., Kiladis, G. N., Weickmann, K. M., Ralph, M. F., & Sardeshmukh, P. D. (2012). Relative contributions of synoptic and low-frequency eddies to time-mean atmospheric moisture transport, including the role of atmospheric rivers. *Journal of Climate*, *25*(21), 7341–7361. doi: 10.1175/JCLI-D-11-00665.1

O'Brien, T. A., Payne, A. E., Shields, C. A., Rutz, J., Brands, S., Castellano, C., … Zhou, Y. (2020, jun). Detection Uncertainty Matters for Understanding Atmospheric Rivers. *Bulletin of the American Meteorological Society*, *101*(6), E790–E796. Retrieved from `https://doi.org/10.31223/osf.io/ftwgmhttp://journals.ametsoc.org/doi/10.1175/BAMS-D-19-0348.1https://journals.ametsoc.org/bams/article/101/6/E790/345615/Detection-Uncertainty-Matters-for-Understanding` doi: 10.1175/BAMS-D-19-0348.1

O'Brien, T. A., Risser, M. D., Loring, B., Elbashandy, A. A., Krishnan, H., Johnson, J., … Collins, W. D. (2020). Detection of Atmospheric Rivers with Inline Uncertainty Quantification: TECA-BARD v1.0. *Geoscientific Model Development Discussions*, *In Review*(April). Retrieved from `https://www.geosci-model-dev-discuss.net/gmd-2020-55/{\#}discussion` doi: 10.5194/gmd-2020-55

O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., … Sanderson, B. M. (2016). The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*(9), 3461–3482. doi: 10.5194/gmd-9-3461-2016

Payne, A. E., Demory, M.-E., Leung, L. R., Ramos, A. M., Shields, C. A., Rutz, J. J., … Ralph, F. M. (2020, mar). Responses and impacts of atmospheric rivers to climate change. *Nature Reviews Earth & Environment*, *1*(3), 143–157. Retrieved from `http://dx.doi.org/10.1038/s43017-020-0030-5http://www.nature.com/articles/s43017-020-0030-5` doi: 10.1038/s43017-020-0030-5

Payne, A. E., & Magnusdottir, G. (2015, nov). An evaluation of atmospheric rivers over the North Pacific in CMIP5 and their response to warming under RCP 8.5. *Journal of Geophysical Research: Atmospheres*, *120*(21), 11,173–11,190. Retrieved from `http://doi.wiley.com/10.1002/2015JD023586` doi: 10.1002/2015JD023586

Polade, S. D., Gershunov, A., Cayan, D. R., Dettinger, M. D., & Pierce, D. W. (2017, dec). Precipitation in a warming world: Assessing projected hydroclimate changes in California and other Mediterranean climate regions. *Scientific Reports*, *7*(1), 10783. Retrieved from `http://dx.doi.org/10.1038/s41598-017-11285-yhttp://www.nature.com/articles/s41598-017-11285-y` doi: 10.1038/s41598-017-11285-y

Prabhat, P., Kashinath, K., Mudigonda, M., Kim, S., Kapp-Schwoerer, L., Graubner, A., … Collins, W. (2020). ClimateNet: an expert-labelled open dataset and Deep Learning architecture for enabling high-precision analyses of extreme weather. *Geoscientific Model Development Discussions*, *In Review*(April). Retrieved from `https://doi.org/10.5194/gmd-2020-72?utm{\_}source=researcher{\_}app{\&}utm{\_}medium=referral{\&}utm{\_}campaign=RESR{\_}MRKT{\_}Researcher{\_}inbound` doi: 10.5194/gmd-2020-72

Ralph, F. M., Coleman, T., Neiman, P. J., Zamora, R. J., & Dettinger, M. D. (2013, apr). 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. Retrieved from `https://journals.ametsoc.org/jhm/article/14/2/443/5819/Observed-Impacts-of-Duration-and-Seasonality-of` doi: 10.1175/JHM-D-12-076.1

Ralph, F. M., Dettinger, M., Lavers, D., Gorodetskaya, I. V., Martin, A., Viale, M., … Cordeira, J. (2017). Atmospheric rivers emerge as a global science and applications focus. *Bulletin of the American Meteorological Society*, *98*(9), 1969–1973. doi: 10.1175/BAMS-D-16-0262.1

Ralph, F. M., Dettinger, M. D., Cairns, M. M., Galarneau, T. J., & Eylander, J. (2018, apr). Defining "Atmospheric River": How the Glossary of Meteorology Helped Resolve a Debate. *Bulletin of the American Meteorological Society*, *99*(4), 837–839. Retrieved from `http://journals.ametsoc.org/doi/10.1175/BAMS-D-17-0157.1` doi: 10.1175/BAMS-D-17-0157.1

Ralph, F. M., Neiman, P. J., & Rotunno, R. (2005). Dropsonde observations in low-level jets over the northeastern Pacific Ocean from CALJET-1998 and PACJET-2001: Mean vertical-profile and atmospheric-river characteristics. *Monthly Weather Review*, *133*(4), 889–910. doi: 10.1175/MWR2896.1

Ralph, F. M., Neiman, P. J., & Wick, G. A. (2004). Satellite and CALJET aircraft observations of atmospheric rivers over the Eastern North Pacific Ocean during the winter of 1997/98. *Monthly Weather Review*, *132*(7), 1721–1745. doi: 10.1175/1520-0493(2004)132⟨1721:SACAOO⟩2.0.CO;2

Ralph, F. M., Rutz, J. J., Cordeira, J. M., Dettinger, M., Anderson, M., Reynolds, D., … Smallcomb, C. (2019, feb). A Scale to Characterize the Strength and Impacts of Atmospheric Rivers. *Bulletin of the American Meteorological Society*, *100*(2), 269–289. Retrieved from `https://journals.ametsoc.org/bams/article/100/2/269/69196/A-Scale-to-Characterize-the-Strength-and-Impacts` doi: 10.1175/BAMS-D-18-0023.1

Ralph, F. M., Wilson, A. M., Shulgina, T., Kawzenuk, B., Sellars, S., Rutz, J. J., … Wick, G. A. (2019, apr). ARTMIP-early start comparison of atmospheric river detection tools: how many atmospheric rivers hit northern California's Russian River watershed? *Climate Dynamics*, *52*(7-8), 4973–4994. Retrieved from `https://doi.org/10.1007/s00382-018-4427-5.,http://link.springer.com/10.1007/s00382-018-4427-5` doi: 10.1007/s00382-018-4427-5

Ramos, A. M., Blamey, R. C., Algarra, I., Nieto, R., Gimeno, L., Tomé, R., … Trigo, R. M. (2019). From amazonia to southern africa: atmospheric moisture transport through low-level jets and atmospheric rivers. *Annals of the New York Academy of Sciences*, *1436*(1), 217-230. Retrieved from `https://nyaspubs.onlinelibrary.wiley.com/doi/abs/10.1111/nyas.13960` doi: 10.1111/nyas.13960

Ramos, A. M., Trigo, R. M., Liberato, M. L. R., & Tomé, R. (2015, apr). Daily

Precipitation Extreme Events in the Iberian Peninsula and Its Association with Atmospheric Rivers*. *Journal of Hydrometeorology*, *16*(2), 579–597. Retrieved from https://journals.ametsoc.org/jhm/article/16/2/579/6106/Daily-Precipitation-Extreme-Events-in-the-Iberian doi: 10.1175/JHM-D-14-0103.1

Rhoades, A. M., Jones, A. D., O'Brien, T. A., O'Brien, J. P., Ullrich, P. A., & Zarzycki, C. M. (2020). Influences of North Pacific Ocean Domain Extent on the Western U.S. Winter Hydroclimatology in Variable-Resolution CESM. *Journal of Geophysical Research: Atmospheres*, *125*(14), e2019JD031977. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JD031977 (e2019JD031977 10.1029/2019JD031977) doi: 10.1029/2019JD031977

Rhoades, A. M., Jones, A. D., Srivastava, A., Huang, H., O'Brien, T. A., Patricola, C. M., ... Zhou, Y. (n.d.). The Shifting Scales of Western US Landfalling Atmospheric Rivers Under Climate Change. *Geophysical Research Letters*, *n/a*(n/a), e2020GL089096. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020GL089096 (e2020GL089096 2020GL089096) doi: 10.1029/2020GL089096

Rutz, J. J., James Steenburgh, W., & Martin Ralph, F. (2014, feb). Climatological characteristics of atmospheric rivers and their inland penetration over the western united states. *Monthly Weather Review*, *142*(2), 905–921. Retrieved from https://journals.ametsoc.org/mwr/article/142/2/905/71947/Climatological-Characteristics-of-Atmospheric doi: 10.1175/MWR-D-13-00168.1

Rutz, J. J., Shields, C. A., Lora, J. M., Payne, A. E., Guan, B., Ullrich, P., ... Viale, M. (2019, dec). The Atmospheric River Tracking Method Intercomparison Project (ARTMIP): Quantifying Uncertainties in Atmospheric River Climatology. *Journal of Geophysical Research: Atmospheres*, *124*(24), 13777–13802. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1029/2019JD030936 doi: 10.1029/2019JD030936

Shields, C. A., & Kiehl, J. T. (2016a, aug). Atmospheric river landfall-latitude changes in future climate simulations. *Geophysical Research Letters*, *43*(16), 8775–8782. Retrieved from http://doi.wiley.com/10.1002/2016GL070470 doi: 10.1002/2016GL070470

Shields, C. A., & Kiehl, J. T. (2016b, jul). Simulating the Pineapple Express in the half degree Community Climate System Model, CCSM4. *Geophysical Research Letters*, *43*(14), 7767–7773. Retrieved from http://doi.wiley.com/10.1002/2016GL069476 doi: 10.1002/2016GL069476

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. doi: 10.1029/2019GL085565

Shields, C. A., Rutz, J. J., Leung, L. R., Ralph, F. M., Wehner, M., O'Brien, T., & Pierce, R. (2019, feb). Defining Uncertainties through Comparison of Atmospheric River Tracking Methods. *Bulletin of the American Meteorological Society*, *100*(2), ES93–ES96. Retrieved from http://journals.ametsoc.org/doi/10.1175/BAMS-D-18-0200.1 doi: 10.1175/BAMS-D-18-0200.1

Shields, C. A., Rutz, J. J., Leung, L.-Y., Ralph, F. M., Wehner, M., Kawzenuk, B., ... Nguyen, P. (2018, jun). Atmospheric River Tracking Method Intercomparison Project (ARTMIP): project goals and experimental design. *Geoscientific Model Development*, *11*(6), 2455–2474. Retrieved from https://www.geosci-model-dev.net/11/2455/2018/ doi: 10.5194/gmd-11-2455-2018

Skinner, C. B., Lora, J. M., Payne, A. E., & Poulsen, C. J. (2020). Atmospheric river changes shaped mid-latitude hydroclimate since the midholocene. *Earth and Planetary Science Letters*, *541*, 116293. Retrieved from http://www.sciencedirect.com/science/article/pii/S0012821X20302363 doi: https://doi.org/10.1016/j.epsl.2020.116293

Sousa, P. M., Ramos, A. M., Raible, C. C., Messmer, M., Tomé, R., Pinto, J. G., & Trigo, R. M. (2020, 12). North Atlantic Integrated Water Vapor Transport—From 850 to 2100 CE: Impacts on Western European Rainfall. *Journal of Climate*, *33*(1), 263-279. Retrieved from https://doi.org/10.1175/JCLI-D-19-0348.1 doi: 10.1175/JCLI-D-19-0348.1

Stohl, A., Forster, C., & Sodemann, H. (2008). Remote sources of water vapor forming precipitation on the Norwegian west coast at 60°N - A tale of hurricanes and an atmospheric river. *Journal of Geophysical Research Atmospheres*, *113*(5), 1–13. Retrieved from http://dx.doi.org/10.1029/2007JD009006 doi: 10.1029/2007JD009006

Taylor, K. E. (2001, apr). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, *106*(D7), 7183–7192. Retrieved from http://doi.wiley.com/10.1029/2000JD900719 doi: 10.1029/2000JD900719

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498. doi: 10.1175/BAMS-D-11-00094.1

Viale, M., & Nuñez, M. N. (2011). Climatology of winter orographic precipitation over the subtropical central Andes and associated synoptic and regional characteristics. *Journal of Hydrometeorology*, *12*(4), 481–507. doi: 10.1175/2010JHM1284.1

Waliser, D. E., Moncrieff, M. W., Burridge, D., Fink, A. H., Gochis, D., Goswami, B. N., ... Yuter, S. (2012, aug). The "Year" of Tropical Convection (May 2008–April 2010): Climate Variability and Weather Highlights. *Bulletin of the American Meteorological Society*, *93*(8), 1189–1218. Retrieved from http://journals.ametsoc.org/doi/abs/10.1175/2011BAMS3095.1 doi: 10.1175/2011BAMS3095.1

Warner, M. D., Mass, C. F., & Salathé, E. P. (2015). Changes in Winter Atmospheric Rivers along the North American West Coast in CMIP5 Climate Models. *Journal of Hydrometeorology*, *16*(1), 118–128. Retrieved from http://journals.ametsoc.org/doi/10.1175/JHM-D-14-0080.1 doi: 10.1175/JHM-D-14-0080.1

Warner, M. D., Mass, C. F., & Salatheé, E. P. (2012). Wintertime extreme precipitation events along the Pacific Northwest Coast: Climatology and synoptic evolution. *Monthly Weather Review*, *140*(7), 2021–2043. doi: 10.1175/MWR-D-11-00197.1

Wille, J. D., Favier, V., Dufour, A., Gorodetskaya, I. V., Turner, J., Agosta, C., & Codron, F. (2019). West Antarctic surface melt triggered by atmospheric rivers. *Nature Geoscience*, *12*(11), 911–916. Retrieved from http://dx.doi.org/10.1038/s41561-019-0460-1 doi: 10.1038/s41561-019-0460-1

Xin, X., Wu, T., Shi, X., Zhang, F., Li, J., Chu, M., ... Wei, M. (2019). *BCC BCC-CSM2MR model output prepared for CMIP6 ScenarioMIP ssp585.* Earth System Grid Federation. Retrieved from https://doi.org/10.22033/ESGF/CMIP6.3050 (Version 20190125) doi: 10.22033/ESGF/CMIP6.3050

Yukimoto, S., Koshiro, T., Kawai, H., Oshima, N., Yoshida, K., Urakawa, S., ... Adachi, Y. (2019). *MRI MRI-ESM2.0 model output prepared for CMIP6 ScenarioMIP ssp585.* Earth System Grid Federation. Retrieved from https://doi.org/10.22033/ESGF/CMIP6.6929 (Version 20190625) doi: 10.22033/ESGF/CMIP6.6929

Zhu, Y., & Newell, R. E. (1998, mar). A Proposed Algorithm for Moisture Fluxes from Atmospheric Rivers. *Monthly Weather Review*, *126*(3), 725–735. Retrieved from http://journals.ametsoc.org/doi/abs/10.1175/1520

-0493{\%}281998{\%}29126{\%}3C0725{\%}3AAPAFMF{\%}3E2.0.CO{\%}3B2
doi: 10.1175/1520-0493(1998)126⟨0725:APAFMF⟩2.0.CO;2

Supporting Information for "Increases in Future AR Count and Size: Overview of the ARTMIP Tier 2 CMIP5/6 Experiment"

T. A. O'Brien[1,2] *, M. F. Wehner[3] and A. E. Payne[4] and C. A. Shields[5] and J. J. Rutz[6] and L.-R. Leung[7] and F. M. Ralph[8] and A. Collow[9,10] and B. Guan[11] and J. M. Lora[12] and E. McClenny[13] and K. M. Nardi[14] and A. M. Ramos[15] and R. Tomé[15] and C. Sarangi[7] and E. Shearer[16] and P. A. Ullrich[13] and C. Zarzycki[14] and B. Loring[3] and H. Huang[2] and H. A. Inda-Díaz[13,2] and A. M. Rhoades[2] and Y. Zhou[2]

[1]Dept. of Earth and Atmospheric Sciences, Indiana University, Bloomington, IN, USA

[2]Climate and Ecosystem Sciences Division, Lawrence Berkeley Lab, Berkeley, CA, USA

[3]Computational Research Division, Lawrence Berkeley Lab, Berkeley, CA, USA

[4]Dept. of Earth and Space Sciences, University of Michigan, Ann Arbor, MI, USA

[5]National Center for Atmospheric Research, Boulder, CO, USA

[6]National Weather Service, Western Region Headquarters, Science and Technology Infusion Division, Salt Lake City, UT, USA

[7]Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, WA, USA

[8]Center for Western Weather and Water Extremes, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, USA

[9]Universities Space Research Association, Columbia, MD, USA

[10]Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA

[11]Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, CA, USA

[12]Dept. of Earth and Planetary Sciences, Yale University, New Haven, CT, USA

[13]Dept. of Land, Air and Water Resources, University of California, Davis, Davis, CA, USA

[14]Dept. of Meteorology and Atmospheric Science, Penn State University, University Park, PA, USA

[15]Instituto Dom Luiz (IDL), Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

[16]Center for Hydrometeorology and Remote Sensing, University of California, Irvine, Irvine, CA, USA

Contents of this file

1. Text S1 to S4
2. Figures S1 to S3
3. Table S1

Introduction

This supplementatal information provides additional useful details on ARDTs, their treatment of thresholds, and our grouping of ARDTs into categories. The supplemental figures expand on figures in the main text to show all ARDT-simulation combinations.

Text S1.

Treatment of Thresholds

Corresponding author: T. A. O'Brien, Department of Earth and Atmospheric Sciences, Indiana University, 1001 E. 10th Street, Bloomington, IN 47408, USA (obrienta@iu.edu)

*Dept. of Earth and Atmospheric Science,

1001 E. 10th St, Bloomington, IN, 47408

We document here choices/specializations (if any) that ARDT contributors made in running their ARDTs on the Tier 2 CMIP5/6 simulations

- `ARCONNECT_v2`: only uses absolute threshold; no Tier 2-specific decisions needed
- `Guan_Waliser_v2`: uses 85$^{\text{th}}$ percentile from the historical simulation
- `IDL_rel_future`: uses 85$^{\text{th}}$ percentile calculated from the future simulation
- `IDL_rel_hist`: uses 85$^{\text{th}}$ percentile calculated from the historical simulation
- `Lora_v2`: uses a time-and-latitude dependent IVT threshold that asymptotes to 225 kg/m/s at the poles; the time/latitude dependence of the threshold is a function of the 30-day running mean and a zonal average of IWV, so no Tier 2-specific decisions are needed
- `Mundhenk_v3`: calculates the mean and seasonal cycle of IVT based on the historical simulation and removes this to determine the IVT anomaly relative to the historical period
- `PNNL_v1`: only uses absolute threshold; no Tier 2-specific decisions needed
- `TECA_BARD_v1.0`: uses threshold relative to spatial map of IVT at a given time; no Tier 2-specific decisions needed

The `Mundhenk_v3` algorithm differs from prior published versions (i.e., `Mundhenk_v1`, `Mundhenk_v2`) in its more reliable detection of AR objects that cross the boundary of the dataset's spatial domain.

The `Tempest` ARDT uses an absolute threshold for the laplacian of IVT. The Tier 1 version also utilized an absolute threshold of 250 kg/m/s of IVT, but it was later determined that this threshold had no effect on the ARDT results because regions that satisfied the Laplacian threshold also satisfied the IVT threshold. The minimum latitude for ARs was raised to 20°, from 15°, to filter easterly waves. The stencil radius and magnitude used for the Laplacian depends on the model grid, and this is held constant for the historical and future simulations.

Discussions with the `Tempest` contributing scientists indicate that the algorithm may benefit from further tuning of their method when applied to moderately low resolution data, and efforts are underway to provide a second version of their contribution to Tier 2. Such discoveries and improvements are a benefit of intercomparison projects.

Text S2.

Classification of ARDTs

Building on Rutz et al. (2019), we classify the Tier 2 CMIP5/6 ARDTs into three groups, based on their treatment of thresholds: *absolute*, *fixed relative*, and *relative*. These classifications are indicated as *abs.*, *fix. rel.*, and *rel.* in Table S1. A key motivation for this categorization is aggregating ARDTs by their sensitivity to thermodynamic changes in IVT, with the assumption that ARDTs employing absolute thresholds to moisture fields will be the most sensitive, and ARDTs employing time-dependent thresholds will be least sensitive.

Absolute ARDTs: We define *absolute ARDTs* as utilizing any fixed thresholds (e.g., in IVT) for discriminating ARs from the background. `ARCONNECT_v2` and `PNNL_v1` unambiguously fit in this category. `Lora_v2` uses an IVT threshold that varies with latitude and time, and the threshold asymptotes to 250 kg/m/s at mid-to-high latitudes (the threshold increases toward infinity approaching the tropics). This design effectively imposes an absolute threshold of at least 250 kg/m/s. Because of this, we classify `Lora_v2` as an *absolute ARDT*, while recognizing that this is not a perfect categorization.

Fixed relative ARDTs: We define *fixed relative ARDTs* as those that employ relative thresholds that do not vary with time. For example, `Guan_Waliser_v2` calculates the 85^{th} percentile of IVT from the historical simulations and discriminates ARs from the background where IVT is greater than the local, historical 85^{th} percentile; hence the threshold used in the `Guan_Waliser_v2` algorithm does not change in time. The `IDL_rel_hist` and `IDL_rel_future` ARDTs use a similar approach and are therefore also categorized as *fixed relative ARDTs*. `Mundhenk_v3` calculates IVT anomalies relative to the historical period and identifies ARs that are above the 94^{th} percentile of the historical simulation, so it also fits unambiguously in the *fixed relative* category.

Relative: We define *relative ARDTs* as those that employ relative thresholds that vary with time. `TECA_BARD_v1.0` unambiguously fits into this category, since ARs are identified where IVT is above a fixed percentile of IVT, where the percentile is calculated in space (in contrast to time, e.g., for `Guan_Waliser_v2`). `Tempest` uses an absolute threshold applied to the Laplacian of the IVT field, which might warrant its classification as an absolute ARDT. However, the use of the Laplacian removes the mean of the IVT field; therefore `Tempest` identifies areas of IVT that are high relative to nearby areas of IVT at the same timestep. We therefore classify `Tempest` as a *relative ARDT*.

Text S3.

Details on Missing Data All ARDTs detect ARs for the 1950-2099 period for the combined historical and future simulations for each CMIP5/6 model. We analyze output from the entire 1950-2099 timeperiod. There are some exceptions to this: output from the CMIP6 IPSL-CM6A-LR SSP5-8.5 simulation are only available through 2049, there are data corruption issues for the year 2006 in the CMIP5 CSIRO-Mk3-6-0 simulation, and there are data corruption issues for the year 2095-2099 for the `TECA_BARD_v1.0` output applied to the CMIP5 IPSL-CM5B-LR simulation. Years with data corruption issues are marked as missing, and trends and climatologies are only calculated considering non-missing data. The `Guan_Waliser_v2` algorithm did not supply ARDT catalogues for the NorESM1-M and BCC-CSM2-MR simulations due to technical issues at the time.

Table S1. (left) ARDT algorithms, and associated metadata, that contributed to the Tier 2 CMIP5/6 experiment. ARDT classifications ('Class.') are described in Text S2. (right) Details of CMIP5/6 models used in the Tier 2 experiment.

ARDTs				Models			
Algorithm ID	Contrib.	Class.	Region	MIP Era	Model Name	Inst.	~Res. [km]
ARCONNECT_v2	Shearer	abs.	Global	CMIP5	CCSM4	NCAR	120
GuanWaliser_v2	Guan	fix. rel.	Global	CMIP5	CSIRO-Mk3-6	CSIRO	207
IDL_rel_future	Ramos	fix. rel.	W. Europe, S. Africa	CMIP5	CanESM2	CCCMA	310
IDL_rel_hist	Ramos	fix. rel.	W. Europe, S. Africa	CMIP5	IPSL-CM5A-LR	IPSL	296
Lora_v2	Lora	abs.	Global	CMIP5	IPSL-CM5B-LR	IPSL	296
Mundhenk_v3	Nardi	fix. rel.	Global	CMIP5	NorESM1-M	NCC	242
PNNL_v1	Sarangi	abs.	W. U.S.	CMIP6	BCC-CSM2-MR	BCC	124
Tempest	McClenny	rel.	Global	CMIP6	IPSL-CM6A-LR	IPSL	198
TECA_BARD_v1.0	O'Brien	rel.	Global	CMIP6	MRI-ESM2-0	MRI	124

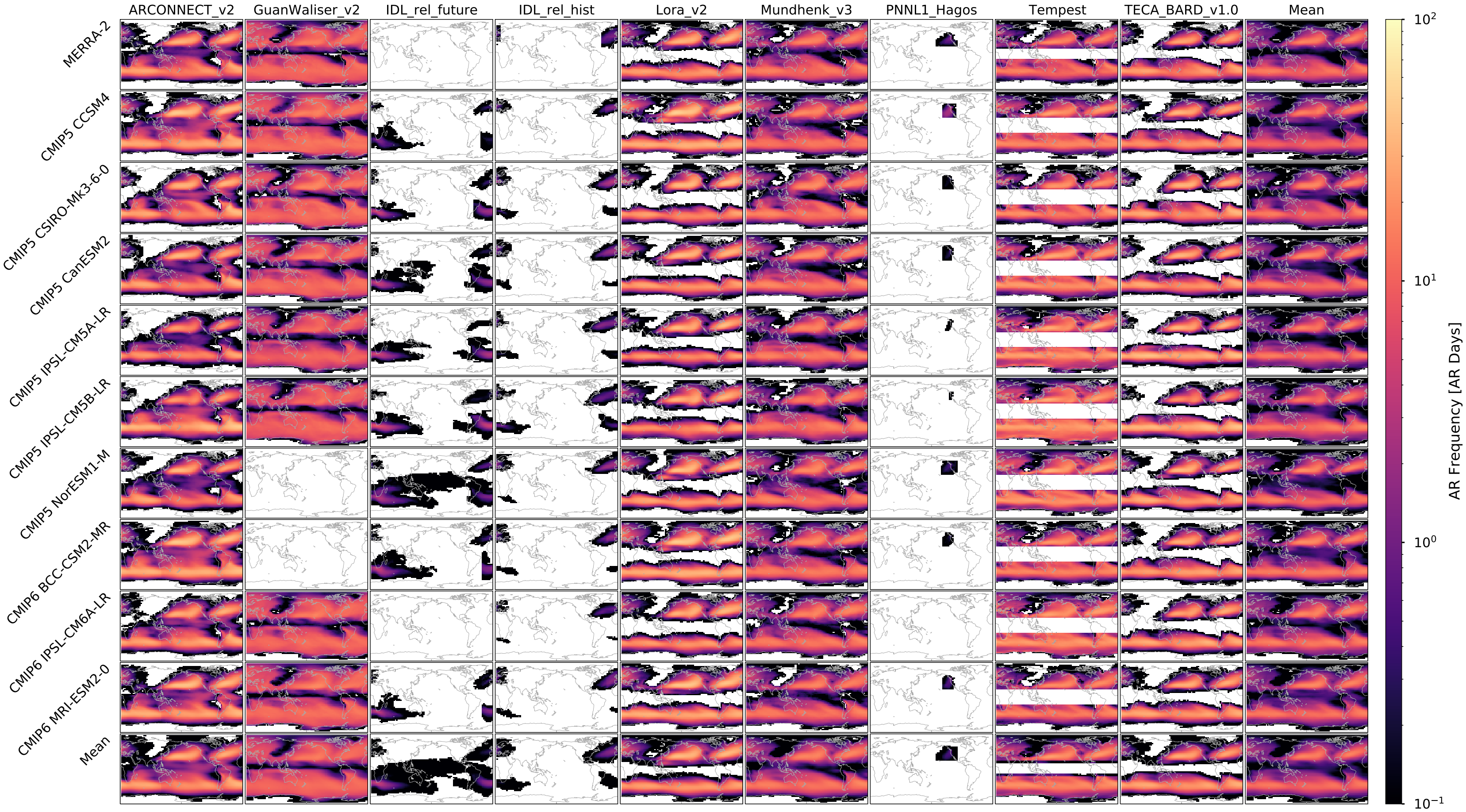

Figure S1. Maps of DJF AR frequency (AR days) from 1981-2010. Columns correspond to ARs detected by specific ARDTs, and rows correspond to input datasets (MERRA-2 for the first row and CMIP5/6 for other rows). The rightmost column shows the multi-ARDT mean frequency for each model. The bottom row shows the multi-model mean for each ARDT (excluding MERRA-2 from the mean). The bottom right panel shows the multi-model, multi-ARDT mean frequency (excluding MERRA-2 from the mean).

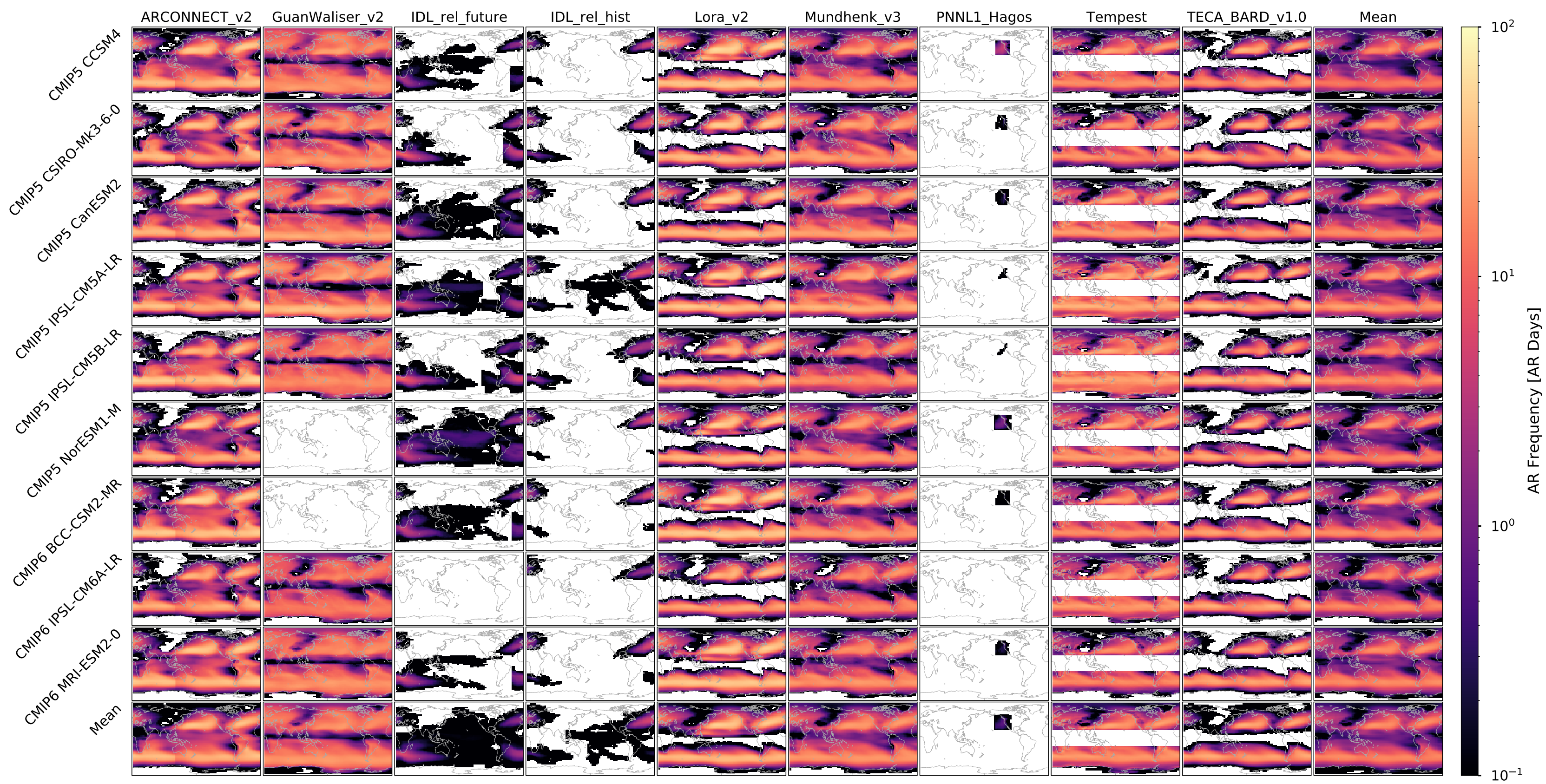

Figure S2. Maps of DJF AR frequency (AR days) from 2070-2099. Columns correspond to ARs detected by specific ARDTs, and rows correspond to CMIP5/6 models. The rightmost column shows the multi-ARDT mean frequency for each model. The bottom row shows the multi-model mean for each ARDT. The bottom right panel shows the multi-model, multi-ARDT mean. The CMIP6 IPSL-CM6A-LR row shows the frequency from the last 30 years of available simulation data (2020-2049).

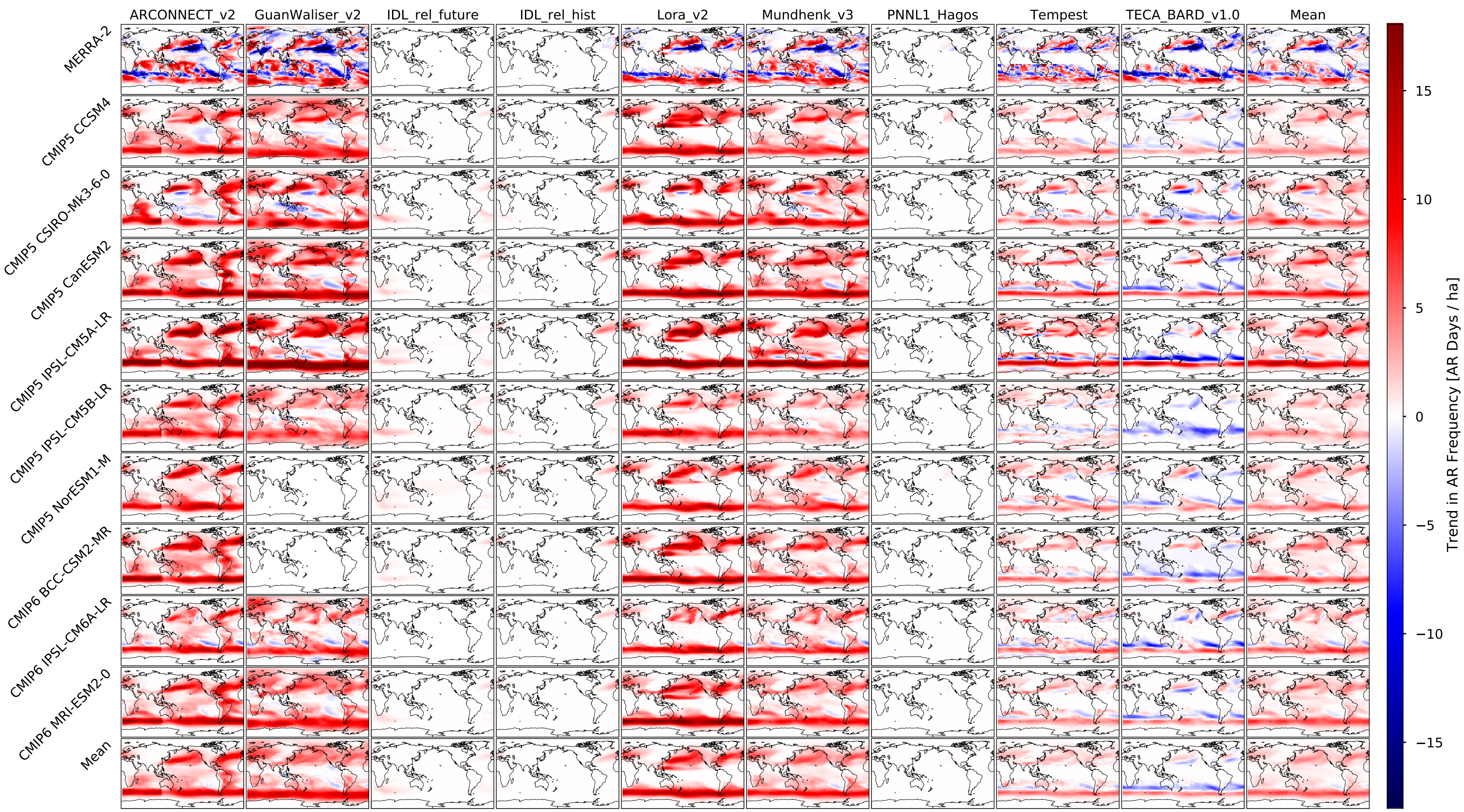

Figure S3. Maps of trends in DJF AR frequency (AR days / ha) from 1981-2100 (from 1981-2017 for MERRA2). Columns correspond to ARs detected by specific ARDTs, and rows correspond to input datasets (MERRA-2 for the first row and CMIP5/6 for other rows). The rightmost column shows the multi-ARDT mean trend for each model. The bottom row shows the multi-model mean for each ARDT (excluding MERRA-2 from the mean). The bottom right panel shows the multi-model, multi-ARDT mean frequency (excluding MERRA-2 from the mean). Trends for CMIP6 IPSL-CM6A-LR are calculated through 2049.

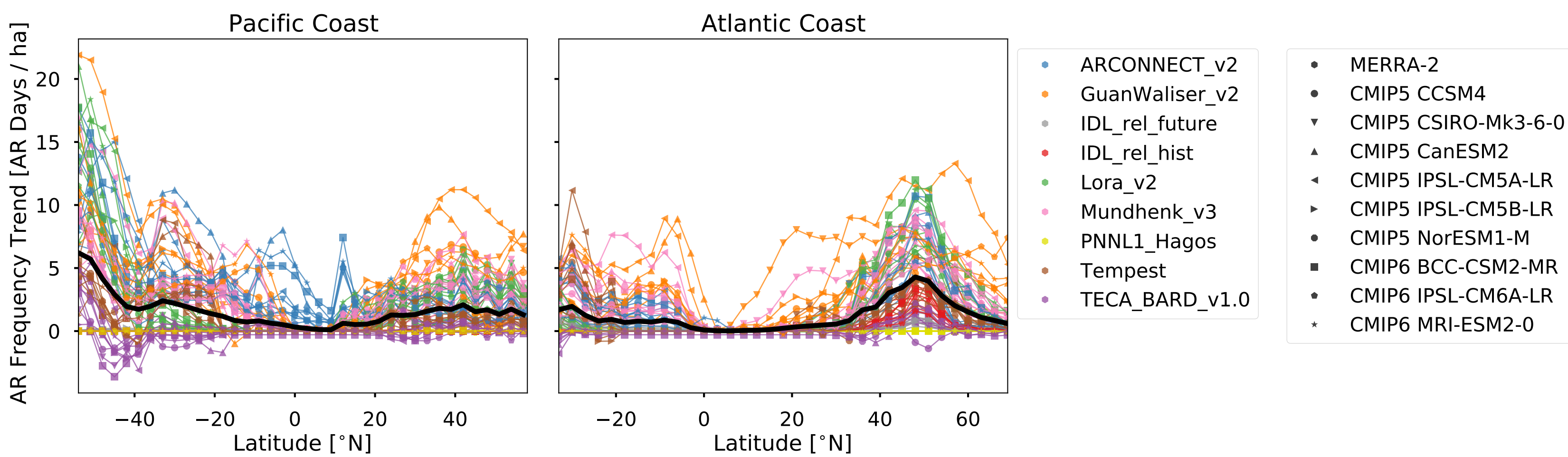

Figure S4. Trends in DJF AR frequency (AR days /ha) from 1981-2100 for western coastlines along the (a) Pacific and (b) Atlantic oceans. Colors correspond to ARDTs and markers correspond to simulations.

Supporting Information for "Increases in Future AR Count and Size: Overview of the ARTMIP Tier 2 CMIP5/6 Experiment"

T. A. O'Brien[1,2] *, M. F. Wehner[3] and A. E. Payne[4] and C. A. Shields[5] and J. J. Rutz[6] and L.-R. Leung[7] and F. M. Ralph[8] and A. Collow[9,10] and B. Guan[11] and J. M. Lora[12] and E. McClenny[13] and K. M. Nardi[14] and A. M. Ramos[15] and R. Tomé[15] and C. Sarangi[7] and E. Shearer[16] and P. A. Ullrich[13] and C. Zarzycki[14] and B. Loring[3] and H. Huang[2] and H. A. Inda-Díaz[13,2] and A. M. Rhoades[2] and Y. Zhou[2]

[1]Dept. of Earth and Atmospheric Sciences, Indiana University, Bloomington, IN, USA

[2]Climate and Ecosystem Sciences Division, Lawrence Berkeley Lab, Berkeley, CA, USA

[3]Computational Research Division, Lawrence Berkeley Lab, Berkeley, CA, USA

[4]Dept. of Earth and Space Sciences, University of Michigan, Ann Arbor, MI, USA

[5]National Center for Atmospheric Research, Boulder, CO, USA

[6]National Weather Service, Western Region Headquarters, Science and Technology Infusion Division, Salt Lake City, UT, USA

[7]Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, WA, USA

[8]Center for Western Weather and Water Extremes, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, USA

[9]Universities Space Research Association, Columbia, MD, USA

[10]Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA

[11]Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, CA, USA

[12]Dept. of Earth and Planetary Sciences, Yale University, New Haven, CT, USA

[13]Dept. of Land, Air and Water Resources, University of California, Davis, Davis, CA, USA

[14]Dept. of Meteorology and Atmospheric Science, Penn State University, University Park, PA, USA

[15]Instituto Dom Luiz (IDL), Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

[16]Center for Hydrometeorology and Remote Sensing, University of California, Irvine, Irvine, CA, USA

Contents of this file

Introduction

This supplementatal information provides additional useful details on ARDTs, their treatment of thresholds, and our grouping of ARDTs into categories. The supplemental figures expand on figures in the main text to show all ARDT-simulation combinations.

Text S1.

Treatment of Thresholds

Corresponding author: T. A. O'Brien, Department of Earth and Atmospheric Sciences, Indiana University, 1001 E. 10th Street, Bloomington, IN 47408, USA (obrienta@iu.edu)

*Dept. of Earth and Atmospheric Science,

1001 E. 10th St, Bloomington, IN, 47408

We document here choices/specializations (if any) that ARDT contributors made in running their ARDTs on the Tier 2 CMIP5/6 simulations

- `ARCONNECT_v2`: only uses absolute threshold; no Tier 2-specific decisions needed
- `Guan_Waliser_v2`: uses 85^{th} percentile from the historical simulation
- `IDL_rel_future`: uses 85^{th} percentile calculated from the future simulation
- `IDL_rel_hist`: uses 85^{th} percentile calculated from the historical simulation
- `Lora_v2`: uses a time-and-latitude dependent IVT threshold that asymptotes to 225 kg/m/s at the poles; the time/latitude dependence of the threshold is a function of the 30-day running mean and a zonal average of IWV, so no Tier 2-specific decisions are needed
- `Mundhenk_v3`: calculates the mean and seasonal cycle of IVT based on the historical simulation and removes this to determine the IVT anomaly relative to the historical period
- `PNNL_v1`: only uses absolute threshold; no Tier 2-specific decisions needed
- `TECA_BARD_v1.0`: uses threshold relative to spatial map of IVT at a given time; no Tier 2-specific decisions needed

The `Mundhenk_v3` algorithm differs from prior published versions (i.e., `Mundhenk_v1`, `Mundhenk_v2`) in its more reliable detection of AR objects that cross the boundary of the dataset's spatial domain.

The `Tempest` ARDT uses an absolute threshold for the laplacian of IVT. The Tier 1 version also utilized an absolute threshold of 250 kg/m/s of IVT, but it was later determined that this threshold had no effect on the ARDT results because regions that satisfied the Laplacian threshold also satisfied the IVT threshold. The minimum latitude for ARs was raised to 20°, from 15°, to filter easterly waves. The stencil radius and magnitude used for the Laplacian depends on the model grid, and this is held constant for the historical and future simulations.

Discussions with the `Tempest` contributing scientists indicate that the algorithm may benefit from further tuning of their method when applied to moderately low resolution data, and efforts are underway to provide a second version of their contribution to Tier 2. Such discoveries and improvements are a benefit of intercomparison projects.

Text S2.

Classification of ARDTs

Building on Rutz et al. (2019), we classify the Tier 2 CMIP5/6 ARDTs into three groups, based on their treatment of thresholds: *absolute*, *fixed relative*, and *relative*. These classifications are indicated as *abs.*, *fix. rel.*, and *rel.* in Table S1. A key motivation for this categorization is aggregating ARDTs by their sensitivity to thermodynamic changes in IVT, with the assumption that ARDTs employing absolute thresholds to moisture fields will be the most sensitive, and ARDTs employing time-dependent thresholds will be least sensitive.

Absolute ARDTs: We define *absolute ARDTs* as utilizing any fixed thresholds (e.g., in IVT) for discriminating ARs from the background. `ARCONNECT_v2` and `PNNL_v1` unambiguously fit in this category. `Lora_v2` uses an IVT threshold that varies with latitude and time, and the threshold asymptotes to 250 kg/m/s at mid-to-high latitudes (the threshold increases toward infinity approaching the tropics). This design effectively imposes an absolute threshold of at least 250 kg/m/s. Because of this, we classify `Lora_v2` as an *absolute ARDT*, while recognizing that this is not a perfect categorization.

Fixed relative ARDTs: We define *fixed relative ARDTs* as those that employ relative thresholds that do not vary with time. For example, `Guan_Waliser_v2` calculates the 85^{th} percentile of IVT from the historical simulations and discriminates ARs from the background where IVT is greater than the local, historical 85^{th} percentile; hence the threshold used in the `Guan_Waliser_v2` algorithm does not change in time. The `IDL_rel_hist` and `IDL_rel_future` ARDTs use a similar approach and are therefore also categorized as *fixed relative ARDTs*. `Mundhenk_v3` calculates IVT anomalies relative to the historical period and identifies ARs that are above the 94^{th} percentile of the historical simulation, so it also fits unambiguously in the *fixed relative* category.

Relative: We define *relative ARDTs* as those that employ relative thresholds that vary with time. `TECA_BARD_v1.0` unambiguously fits into this category, since ARs are identified where IVT is above a fixed percentile of IVT, where the percentile is calculated in space (in contrast to time, e.g., for `Guan_Waliser_v2`). `Tempest` uses an absolute threshold applied to the Laplacian of the IVT field, which might warrant its classification as an absolute ARDT. However, the use of the Laplacian removes the mean of the IVT field; therefore `Tempest` identifies areas of IVT that are high relative to nearby areas of IVT at the same timestep. We therefore classify `Tempest` as a *relative ARDT*.

Text S3.

Details on Missing Data All ARDTs detect ARs for the 1950-2099 period for the combined historical and future simulations for each CMIP5/6 model. We analyze output from the entire 1950-2099 timeperiod. There are some exceptions to this: output from the CMIP6 IPSL-CM6A-LR SSP5-8.5 simulation are only available through 2049, there are data corruption issues for the year 2006 in the CMIP5 CSIRO-Mk3-6-0 simulation, and there are data corruption issues for the year 2095-2099 for the `TECA_BARD_v1.0` output applied to the CMIP5 IPSL-CM5B-LR simulation. Years with data corruption issues are marked as missing, and trends and climatologies are only calculated considering non-missing data. The `Guan_Waliser_v2` algorithm did not supply ARDT catalogues for the NorESM1-M and BCC-CSM2-MR simulations due to technical issues at the time.

Table S1. (left) ARDT algorithms, and associated metadata, that contributed to the Tier 2 CMIP5/6 experiment. ARDT classifications ('Class.') are described in Text S2. (right) Details of CMIP5/6 models used in the Tier 2 experiment.

ARDTs				Models			
Algorithm ID	Contrib.	Class.	Region	MIP Era	Model Name	Inst.	∼Res. [km]
ARCONNECT_v2	Shearer	abs.	Global	CMIP5	CCSM4	NCAR	120
GuanWaliser_v2	Guan	fix. rel.	Global	CMIP5	CSIRO-Mk3-6	CSIRO	207
IDL_rel_future	Ramos	fix. rel.	W. Europe, S. Africa	CMIP5	CanESM2	CCCMA	310
IDL_rel_hist	Ramos	fix. rel.	W. Europe, S. Africa	CMIP5	IPSL-CM5A-LR	IPSL	296
Lora_v2	Lora	abs.	Global	CMIP5	IPSL-CM5B-LR	IPSL	296
Mundhenk_v3	Nardi	fix. rel.	Global	CMIP5	NorESM1-M	NCC	242
PNNL_v1	Sarangi	abs.	W. U.S.	CMIP6	BCC-CSM2-MR	BCC	124
Tempest	McClenny	rel.	Global	CMIP6	IPSL-CM6A-LR	IPSL	198
TECA_BARD_v1.0	O'Brien	rel.	Global	CMIP6	MRI-ESM2-0	MRI	124

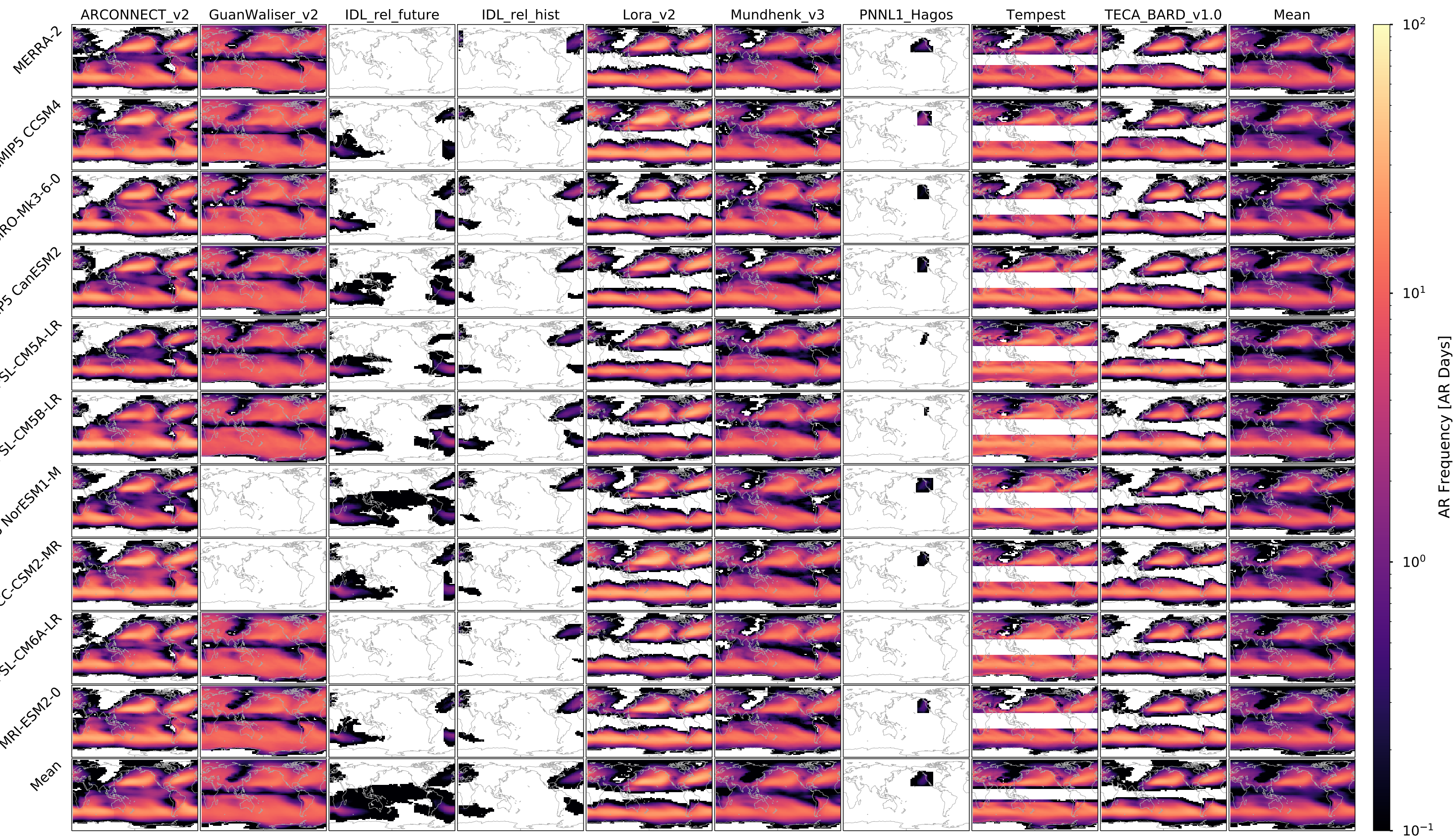

Figure S1. Maps of DJF AR frequency (AR days) from 1981-2010. Columns correspond to ARs detected by specific ARDTs, and rows correspond to input datasets (MERRA-2 for the first row and CMIP5/6 for other rows). The rightmost column shows the multi-ARDT mean frequency for each model. The bottom row shows the multi-model mean for each ARDT (excluding MERRA-2 from the mean). The bottom right panel shows the multi-model, multi-ARDT mean frequency (excluding MERRA-2 from the mean).

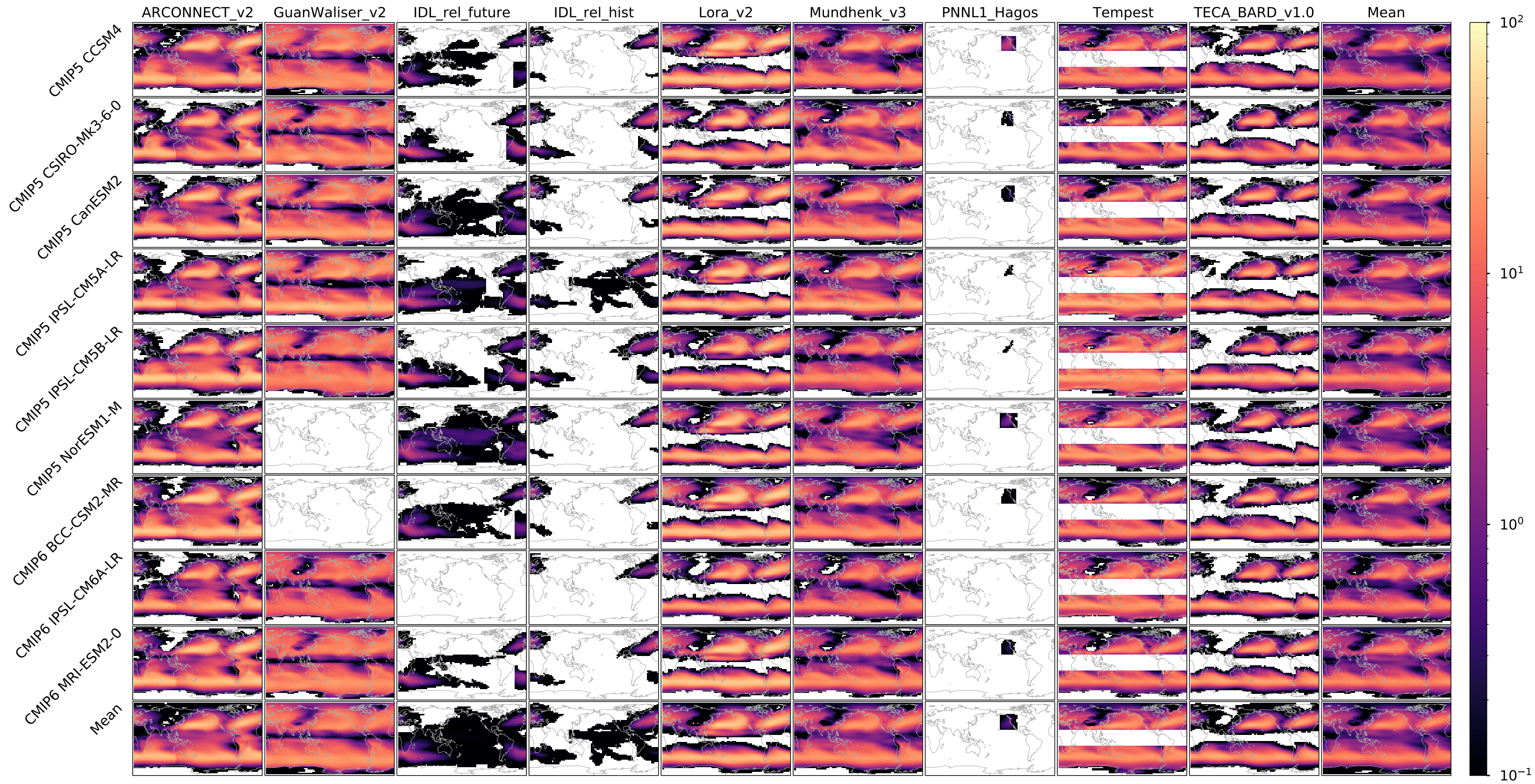

Figure S2. Maps of DJF AR frequency (AR days) from 2070-2099. Columns correspond to ARs detected by specific ARDTs, and rows correspond to CMIP5/6 models. The rightmost column shows the multi-ARDT mean frequency for each model. The bottom row shows the multi-model mean for each ARDT. The bottom right panel shows the multi-model, multi-ARDT mean. The CMIP6 IPSL-CM6A-LR row shows the frequency from the last 30 years of available simulation data (2020-2049).

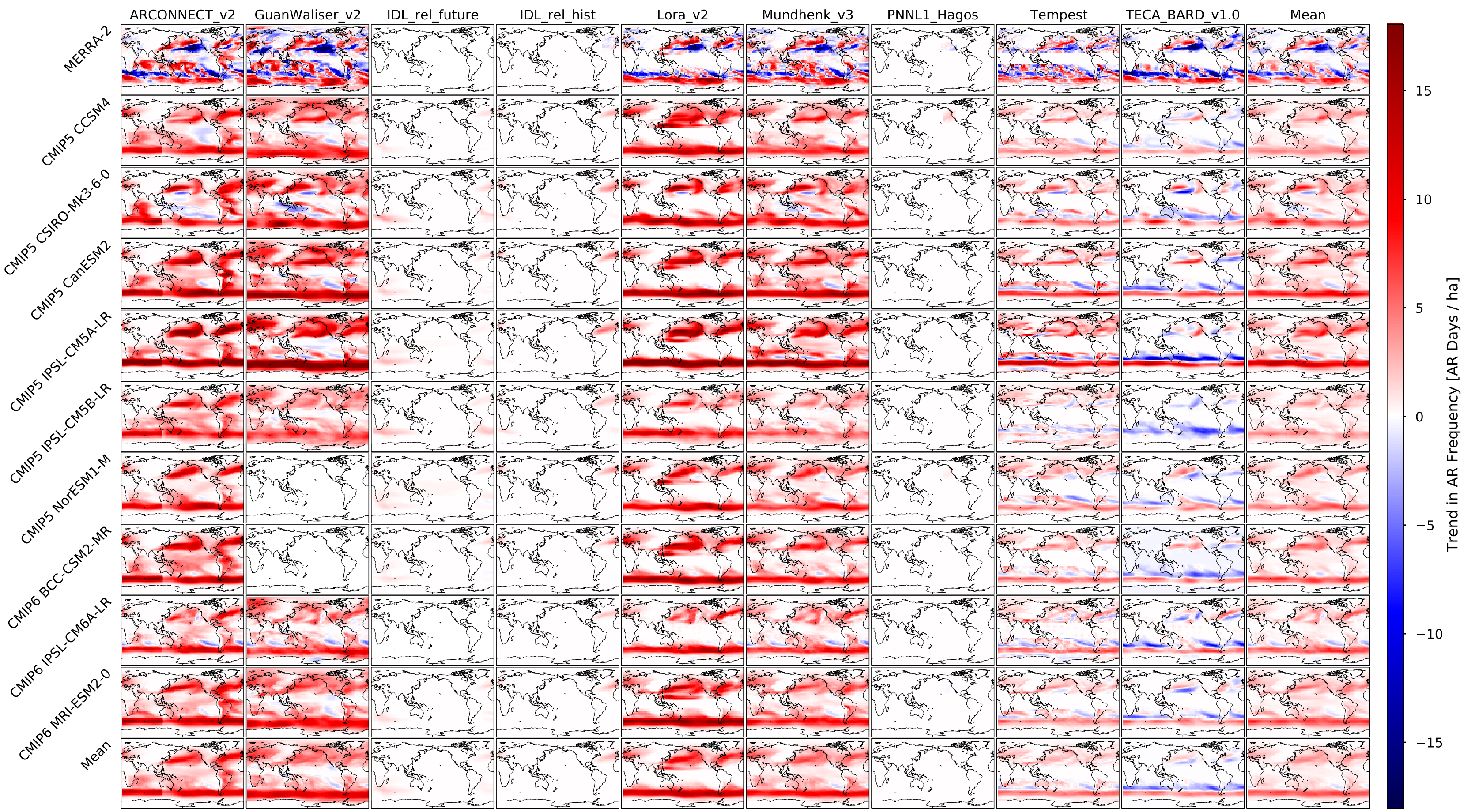

Figure S3. Maps of trends in DJF AR frequency (AR days / ha) from 1981-2100 (from 1981-2017 for MERRA2). Columns correspond to ARs detected by specific ARDTs, and rows correspond to input datasets (MERRA-2 for the first row and CMIP5/6 for other rows). The rightmost column shows the multi-ARDT mean trend for each model. The bottom row shows the multi-model mean for each ARDT (excluding MERRA-2 from the mean). The bottom right panel shows the multi-model, multi-ARDT mean frequency (excluding MERRA-2 from the mean). Trends for CMIP6 IPSL-CM6A-LR are calculated through 2049.

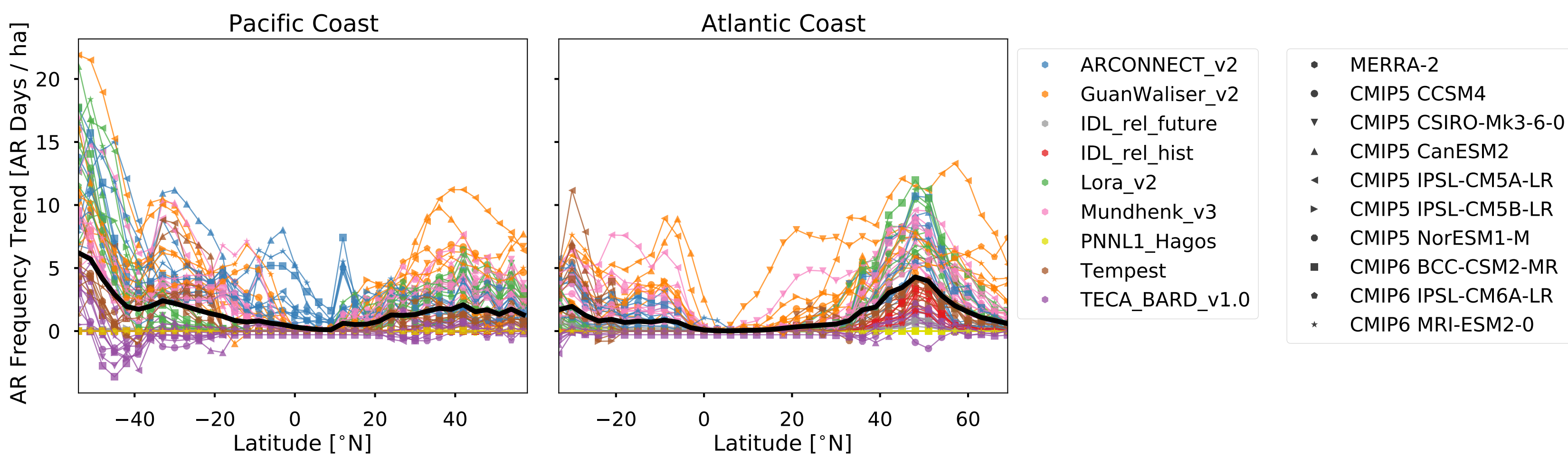

Figure S4. Trends in DJF AR frequency (AR days /ha) from 1981-2100 for western coastlines along the (a) Pacific and (b) Atlantic oceans. Colors correspond to ARDTs and markers correspond to simulations.