Author Contributions:
Conceptualization: C. A. Shields, A. Payne, L. R. Leung, T. A. O'Brien, F. M. Ralph, J. Rutz, P. A. Ullrich, M. Wehner
Data curation: B. Guan, S. Kim, J. M. Lora, E. E. McClenny, K. Nardi, A. Payne, K. Reid, E. J. Shearer, R. Tomé, J. D. Wille, P. A. Ullrich
Formal analysis: C. A. Shields, K. Reid, E. J. Shearer, J. D. Wille, A. M. Ramos, I. V. Gorodetskaya

An Overview of ARTMIP's Tier 2 Reanalysis Intercomparison: Uncertainty in the Detection of Atmospheric Rivers and Their Associated Precipitation

A. B. Marquardt Collow[1,2,3] [ID], C. A. Shields[4] [ID], B. Guan[5,6] [ID], S. Kim[7], J. M. Lora[8] [ID], E. E. McClenny[9] [ID], K. Nardi[10], A. Payne[11,12], K. Reid[13] [ID], E. J. Shearer[14] [ID], R. Tomé[15] [ID], J. D. Wille[16] [ID], A. M. Ramos[15] [ID], I. V. Gorodetskaya[17], L. R. Leung[18] [ID], T. A. O'Brien[19,20] [ID], F. M. Ralph[21] [ID], J. Rutz[22], P. A. Ullrich[9] [ID], and M. Wehner[23] [ID]

[1]Universities Space Research Association, Columbia, MD, USA, [2]Now at University of Maryland Baltimore County, Baltimore, MD, USA, [3]Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA, [4]Climate and Global Dynamics Lab, National Center for Atmospheric Research, Boulder, CO, USA, [5]Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, Los Angeles, CA, USA, [6]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, [7]Department of Geography, University of California Berkeley, Berkeley, CA, USA, [8]Department of Earth and Planetary Sciences, Yale University, New Haven, CT, USA, [9]University of California, Davis, Davis, CA, USA, [10]Department of Meteorology and Atmospheric Science, Pennsylvania State University, University Park, PA, USA, [11]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA, [12]Tomorrow.io, Boston, MA, USA, [13]School of Geography, Earth, and Atmospheric Sciences and ARC Centre of Excellence for Climate Extremes, The University of Melbourne, Melbourne, VIC, Australia, [14]Center for Hydrometeorology and Remote Sensing (CHRS), Department of Civil and Environmental Engineering, Henry Samueli School of Engineering, University of California, Irvine, Irvine, CA, USA, [15]Instituto Dom Luiz, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal, [16]CNRS/UGA/IRD/G-INP, Institut des Géosciences de l'Environnement, Saint Martin d'Hères, Grenoble, France, [17]Department of Physics, CESAM – Centre for Environmental and Marine Studies, University of Aveiro, Aveiro, Portugal, [18]Earth Systems Analysis and Modeling, Pacific Northwest National Laboratory, Richland, WA, USA, [19]Department of Earth and Atmospheric Sciences, Indiana University Bloomington, Bloomington, IN, USA, [20]Climate and Ecosystem Sciences Division, Lawrence Berkeley National Lab, Berkeley, CA, USA, [21]Center for Western Weather and Water Extremes, Scripps Institution of Oceanography, La Jolla, CA, USA, [22]Science and Technology Infusion Division, National Weather Service Western Region Headquarters, National Oceanic and Atmospheric Administration, Salt Lake City, UT, USA, [23]Applied Mathematics and Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Abstract Atmospheric rivers, or long but narrow regions of enhanced water vapor transport, are an important component of the hydrologic cycle as they are responsible for much of the poleward transport of water vapor and result in precipitation, sometimes extreme in intensity. Despite their importance, much uncertainty remains in the detection of atmospheric rivers in large datasets such as reanalyses and century scale climate simulations. To understand this uncertainty, the Atmospheric River Tracking Method Intercomparison Project (ARTMIP) developed tiered experiments, including the Tier 2 Reanalysis Intercomparison that is presented here. Eleven detection algorithms submitted hourly tags--binary fields indicating the presence or absence of atmospheric rivers--of detected atmospheric rivers in the Modern Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2) and European Centre for Medium-Range Weather Forecasts' Reanalysis Version 5 (ERA5) as well as six-hourly tags in the Japanese 55-year Reanalysis (JRA-55). Due to a higher climatological mean for integrated water vapor transport in MERRA-2, atmospheric rivers were detected more frequently relative to the other two reanalyses, particularly in algorithms that use a fixed threshold for water vapor transport. The finer horizontal resolution of ERA5 resulted in narrower atmospheric rivers and an ability to detect atmospheric rivers along resolved coastlines. The fraction of hemispheric area covered by ARs varies throughout the year in all three reanalyses, with different atmospheric river detection tools having different seasonal cycles.

1. Introduction

Atmospheric Rivers (ARs), defined by the AMS Glossary of Meteorology as "a long, narrow, and transient corridor of strong horizontal water vapor transport that is typically associated with a low-level jet stream ahead of the cold front of an extratropical cyclone" (Ralph et al., 2018), play a large role in the Earth's hydrologic

cycle by accounting for nearly all of the poleward transport of water vapor (Zhu & Newell, 1998). ARs can range from being beneficial, relieving droughts and providing essential freshwater resources (Dettinger, 2013; Dettinger et al., 2011), to having catastrophic societal impacts through extreme precipitation events, flooding, and landslides (Collow et al., 2020; Guan et al., 2010; Neiman et al., 2011; Ralph et al., 2006). Numerous studies have contributed to a growing understanding of ARs, including their underlying processes and lifecycle (Guan & Waliser, 2019; Kim & Chiang, 2021; Shearer et al., 2020; Zhou & Kim, 2019), interaction with other meteorological features such as blocking (Benedict et al., 2019; Wille et al., 2021), Rossby wave breaking (Ryoo et al., 2013; Zavadoff & Kirtman, 2020), and monsoonal systems (Park et al., 2021), relationship to winds (Waliser & Guan, 2017), jets (Shields & Kiehl, 2016a), precipitation (Arabzadeh et al., 2020; Collow et al., 2020; Gorodetskaya et al., 2014; Ramos et al., 2015; Wille et al., 2021), flooding (Ionita et al., 2020; Lavers et al., 2011; Neiman et al., 2011; Ralph et al., 2006), and snowpack (Chen et al., 2019; Hu & Nolin, 2019; Wille et al., 2019, 2021), contribution to the global energy budget (Shields et al., 2019), and predictability at numerical weather prediction (Chen et al., 2018), seasonal (Huang et al., 2021), and climate (Massoud et al., 2019; McClenny et al., 2020; Payne et al., 2020; Rhoades et al., 2020) timescales. Due to the Clausius-Clapeyron relationship, ARs are projected to contain more moisture in the future. However, knowledge gaps and uncertainties exist, such as how changes in the large-scale circulation may impact the size, duration, and frequency of ARs in a changing climate (Payne et al., 2020 and references within).

Uncertainties arise in our understanding of ARs and how they may change in the future due to the vast array of techniques used to detect ARs in large datasets (Lora et al., 2020; O'Brien et al., 2020; Zhou et al., 2021). In an effort to reduce ambiguities and inform best practices for future studies in AR science, the Atmospheric River Tracking Method Intercomparison Project (ARTMIP) devised a set of goals and tiered intercomparisons to quantify the characteristics and impacts of ARs and how they vary due to the methodology used for AR detection (Ralph et al., 2019; Rutz et al., 2019; Shields et al., 2018). ARTMIP began with a Tier 1 intercomparison of AR detection in a common data set constrained through the assimilation of observations, NASA's Modern Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2; Rutz et al., 2019). The Tier 2 intercomparisons build upon the work completed in Tier 1 by aiming to answer more specific science questions. Other Tier 2 intercomparisons are focused on the high-resolution climate catalogs of C20C (Payne et al., 2020) and ARs in a changing climate using CMIP-5 and -6 multimodel ensembles (O'Brien et al., 2021). Here, we present an overview of the Tier 2 reanalysis intercomparison.

A reanalysis blends a constant version of a general circulation model and data assimilation system with conventional and satellite-based observations of the atmosphere to give a spatially and temporally consistent view of the atmosphere over the course of decades. Reanalysis products should be considered as a proxy for actual observations. They are produced by a highly constrained numerical model subject to the biases in the model's formulations. Not all components of the Earth system are adequately sampled by the observing system throughout the forty plus years included in many modern reanalyses and it is up to the underlying model to fill in these gaps. Each reanalysis has its own model biases and contains unique features that lend a given reanalysis for use in specific research studies. For example, MERRA-2 includes the assimilation of aerosol optical depth as well as a dry mass constraint to balance the hydrologic cycle (Gelaro et al., 2017), the European Centre for Medium-Range Weather Forecasts' (ECMWF's) ERA5 uses an ensemble approach for data assimilation which allows for a measure of uncertainty (Hersbach et al., 2020), and the Japan Meteorological Agency's JRA-55 incorporates a consistent analysis for sea ice concentration and sea surface temperature, reducing jumps associated with changes in boundary forcing (Kobayashi et al., 2015).

All three of these reanalyses have been used to study ARs; however, ARTMIP's Tier 1 experiment only included 3-hourly data from MERRA-2. Tier 1 results are being used to evaluate climate models without proper investigation of the suitability for this purpose (O'Brien et al., 2021). As part of the Tier 2 reanalysis intercomparison, ARTMIP developed the following set of AR science questions:

1. What are the appropriate metrics for investigating ARs and how have/will they change(d) in time?
1a. What is the best way to handle regional differences in ARs in a global algorithm?
1b. How are diagnostic reanalysis variables, such as precipitation, impacted by AR detection?
2. How does the uncertainty in AR frequency among reanalyses compare to the uncertainty among tracking algorithms?
3. How suitable are century scale reanalyses for studying ARs?

Table 1
Reanalysis Products Used in This Study

Reanalysis	Spatial resolution	Temporal resolution	Temporal coverage used	Data assimilation
MERRA-2	$0.625° \times 0.5°$	1 Hour	1980–2019	6-hour 3D-Var
ERA5	$0.25° \times 0.25°$	1 Hour	1980–2019 or 2000–2019	12-hour 4D-Var ensemble
JRA-55	$1.25° \times 1.25°$	6 Hours	1980–2019	6-hour 4D-Var
JRA-55C	$1.25° \times 1.25°$	6 Hours	1980–2012	Conventional-only 6-hour 4D-Var

Note. Temporal resolution and coverage indicate the product versions used here and may not reflect the original reanalysis output.

4. Does a specific reanalysis lend itself to determining whether a climate model is suitable for studying ARs?

Reanalysis intercomparisons have been performed in the past for individual AR detection algorithms and regions (Guan et al., 2018; Guan & Waliser, 2017; Huning et al., 2019; Lavers et al., 2012; Wille et al., 2019, 2021). The Tier 2 reanalysis intercomparison expands upon the previous work by taking a global approach to compare and contrast AR detection in multiple algorithms. Here we present an overview of the Tier 2 reanalysis intercomparison and share preliminary results as the AR community works together in answering the specific science questions. This paper addresses science questions 1b and 2, and it is anticipated that future work will not only expand on answering these science questions but will also address the science questions not covered here, particularly the usefulness of century scale reanalyses and the suitability of reanalyses to evaluate climate models. Section 2 details the methodology for the intercomparison and participating AR detection methods. An intercomparison of hydrological fields in the reanalyses is presented in Section 3 while results comparing the detection of ARs and their impact on precipitation can be found in Section 4.

2. Data and Methods

2.1. Reanalysis Products

Three global reanalyses, 1-hourly MERRA-2 from the US National Aeronautics and Space Administration (NASA; Gelaro et al., 2017), 1-hourly ERA5 from the European Centre for Medium-range Weather Forecasts (ECMWF; Hersbach et al., 2020), and 6-hourly JRA-55 from the Japan Meteorological Agency (JMA; Kobayashi et al., 2015) were selected for an intercomparison following an assessment of the spatial and temporal resolutions, data availability, and use in published studies on ARs among the array of currently available global reanalysis products. JRA-55C, a counterpart to JRA-55 that only assimilates conventional surface and upper air observations, was also selected to evaluate the usefulness of reanalyses prior to the satellite era and to assess the suitability of century scale reanalyses for studying trends in ARs. While JRA-55C is introduced here as part of the overall methodology for the Tier 2 Reanalysis Intercomparison, it is not included in the results in Section 4. Additional details on each reanalysis can be found in Table 1.

2.2. AR Detection Algorithms

A total of eleven algorithms submitted AR detection tags for participation in the reanalysis intercomparison. The algorithms used a wide variety of methods to detect an AR including relative and absolute thresholds (Rutz et al., 2019) for atmospheric moisture, winds, and integrated water vapor transport (IVT), geometry thresholds, deep learning (Prabhat et al., 2021), and object identification (Guan & Waliser, 2019; Shearer et al., 2020). While most algorithms are global, four are designed for AR detection in specific regions (Table 2). Additional information for the algorithms, including a reference describing the technique used for AR detection, can be found in Table 2.

2.3. Methodology

Each participating AR detection tool (ARDT) was required to submit binary tags of AR detection for the entire 40-year period of MERRA-2 and the last 20 years, encompassing 2000 through 2019, for ERA5. The decision regarding the temporal requirement for ERA5 was made to ease the burden for those with limited computational

Table 2
AR Detection Algorithms Included in the Reanalysis Intercomparison

| Algorithm	Participant	Region	Reference
ARCONNECT_v2[†]	Eric Shearer	Global	Shearer et al., 2020
ClimateNet_DL	Sol Kim	Global	Prabhat et al., 2021
GuanWaliser_v2*[†]	Bin Guan	Global	Guan & Waliser, 2015; Guan et al., 2018
IDL_v2b*[†]	Ricardo Tomé	Western Europe, South Africa	Ramos et al., 2016; Blamey et al., 2018
Lora_v2	Juan Lora	Global	Lora et al., 2017; Skinner et al., 2020
Mundhenk_v3*[†]	Kyle Nardi	Global	Mundhenk et al., 2016
Payne*[†]	Ashley Payne	Western US	Payne & Magnusdottir, 2014; Payne & Magnusdottir, 2015
Reid500*[†]	Kimberley Reid	Global	Reid et al., 2020
Shields_v1*	Christine Shields	Western US, Western Europe	Shields & Kiehl, 2016a; Shields & Kiehl, 2016b
TempestLR*[†]	Beth McClenny	Global, except 15 S to 15 N	McClenny et al., 2020
Wille_v2.4*	Jonathan Wille	Arctic, Antarctica	Wille et al., 2021

Note. Algorithms that were tuned separately for each reanalysis are italicized. *Included full ERA5 period, [†]Included JRA-55 and JRA-55C.

resources as the 0.25°, 1-hourly resolution of ERA5 translated to a yearly file size of 144 GB. Participants also had the option to submit AR tags for 1980 through 1999 in ERA5, 1980 through 2019 in JRA-55, and 1980 through 2012 in JRA-55C. The contributions of each algorithm are denoted by the asterisks and crosses in Table 2. ARs were detected at the native spatial and temporal resolution for each reanalysis specified in Table 1. It is important to note that AR tags in MERRA-2 can differ from the tags associated with ARTMIP's Tier 1 effort (Rutz et al., 2019; Shields et al., 2018). Model derived vertically integrated water vapor fluxes were used to derive IVT for the reanalysis intercomparison (GMAO, 2015a), while the previous Tier 1 studies computed IVT based on the vertical profile of specific humidity and horizontal winds (GMAO, 2015b; Shields et al., 2018). This change in methodology allowed for the use of hourly AR tags as opposed to three-hourly and eliminated any concern of differing IVT in the reanalyses associated with vertical resolution. As demonstrated by Reid et al. (2020), AR detection can be sensitive to interpolation and regridding of the input data set. A brief assessment of the MERRA-2 model derived IVT compared to the IVT used for the Tier 1 effort can be found in the supplemental document.

3. Reanalysis Intercomparison of AR Related Fields

Prior to evaluating the detection of ARs in the reanalyses, it is essential to compare fields commonly used for the detection and analysis of ARs in the reanalysis products. Each reanalysis uses a unique underlying general circulation model, data assimilation scheme, and observing system. While the model parameterizations and data assimilation scheme are held fixed in a reanalysis, the observing system and boundary forcing changes with time. Reanalyses can be prone to model biases during earlier time periods such as prior to the launch of NASA's EOS satellites or in data sparse regions like the Arctic. As a result, reanalyses can differ from one another in space and time despite their observational constraints.

Figure 1 shows a time series of monthly, global mean fields smoothed with a 12-month running mean in MERRA-2, ERA5, and JRA-55. All three reanalyses are well correlated with one another, with peaks in total precipitable water (TPW) during El Niño years and valleys in strong La Niña years (Figure 1a). However, the magnitudes and trends throughout the forty-year time series differ. The largest separation between the reanalysis products occurs prior to 1998 when the first advanced microwave sounders were assimilated (McCarty et al., 2016). Interestingly ERA5 and JRA55 are in excellent agreement for TPW between 1998 and 2000, but diverge for the remainder of the time series, likely due to an observation source used by one reanalysis but not the other in the second half of the time series. On average, the TPW in MERRA-2 is ~0.75 kg m^{-2} higher than ERA5 and ~1 kg m^{-2} higher than JRA55. Bosilovich et al. (2017) evaluated ocean-only TPW between 60°N and 60°S in MERRA-2 and demonstrated that MERRA-2 agrees well with SSM/I retrievals in the late 1990s and early 2000s. However, MERRA-2 suffers from a bias with respect to the expected increasing trend in TPW associated with the Clausius–Clapeyron
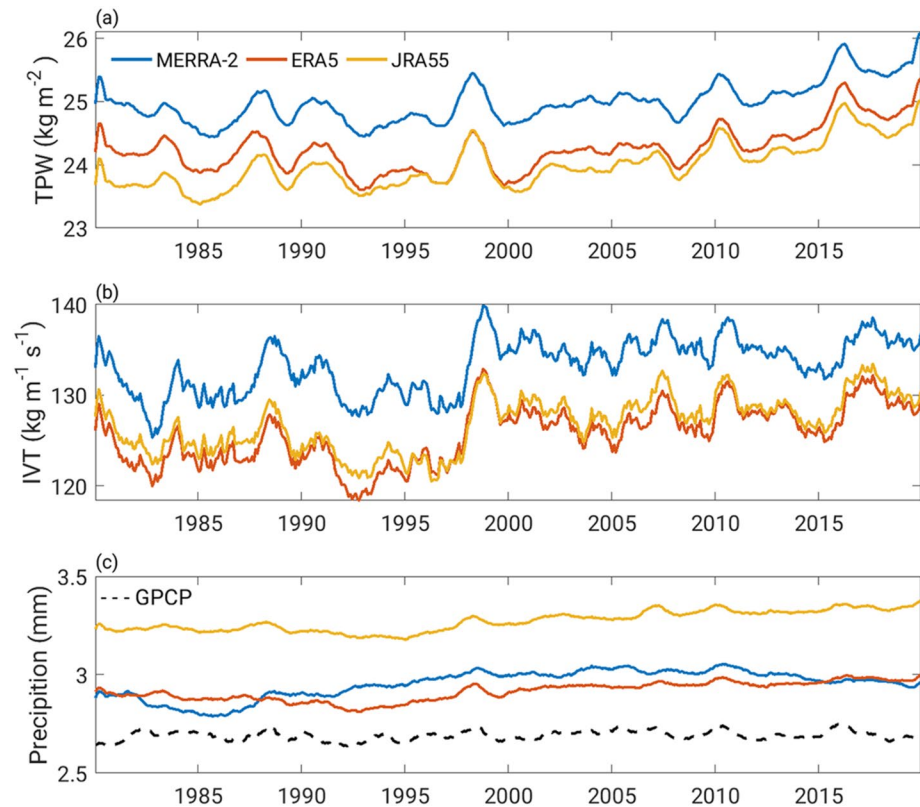
Figure 1. Monthly, global mean (a) total precipitable water vapor, (b) integrated water vapor transport, and (c) precipitation in MERRA-2, ERA5, and JRA-55 as well GPCPv2.3 (Adler et al., 2003). All fields are smoothed with a 12-month running mean.

rate such that TPW in MERRA-2 is biased high in the late 1980s and early 1990s and biased low in the 2010s. JRA-55 and ERA5 are also unable to capture the magnitude of the trend, and thus have a better agreement to the SSM/I retrievals than MERRA-2 in the late 1980s but a stronger negative bias in the 2010s (Xue et al., 2019).

Due to the larger water vapor content in MERRA-2, global mean IVT is higher in MERRA-2 relative to ERA5 and JRA-55. Like with the time series for TPW, there is a jump in the magnitude for IVT in the late 1990s within all three reanalyses (Figure 1b). An increase in the analysis increment for specific humidity begins to occur in the lower troposphere in MERRA-2 (Gelaro et al., 2017) and ERA5 (Hersbach et al., 2020) at this point in time. This is coupled with a doubling in the count of assimilated atmospheric motion vectors, at least in MERRA-2, due to the influx of information from the NOAA-15 and NOAA-16 satellites (McCarty et al., 2016). IVT is also a function of wind speed however wind speeds, especially in the extratropics, are well constrained with minimal differences between the three reanalyses (Figure S2 in Supporting Information S1).

Precipitation is a fully model-generated field in the reanalyses, meaning that data assimilation only constrains variables that could influence precipitation and not the precipitation itself. All reanalyses are biased high with respect to gridded satellite products such as the Global Precipitation Climatology Project (GPCP; Gelaro et al., 2017; Hersbach et al., 2020); however, on average there is, at a global, monthly mean, 0.3 mm more precipitation in JRA-55 than the other two reanalyses (Figure 1c). Though not show here due to inconsistencies in the time series and a lack of data prior to 2000, the global mean precipitation from Integrated Multi-satellitE Retrievals for the Global Precipitation Measurement Mission (GPM IMERG) lies in between JRA-55 and the other two reanalyses, skewing toward MERRA-2 and ERA5 after the launch of the GPM Core Observatory satellite. JRA-55 and ERA5 have an increasing trend in global mean precipitation that exceeds any trend in GPCP, and this trend matches the timing of the trend in global mean TPW. While MERRA-2 has a positive trend in precipitation over the middle of the timeseries, there is a decreasing trend in global mean precipitation over the past decade. As a result,
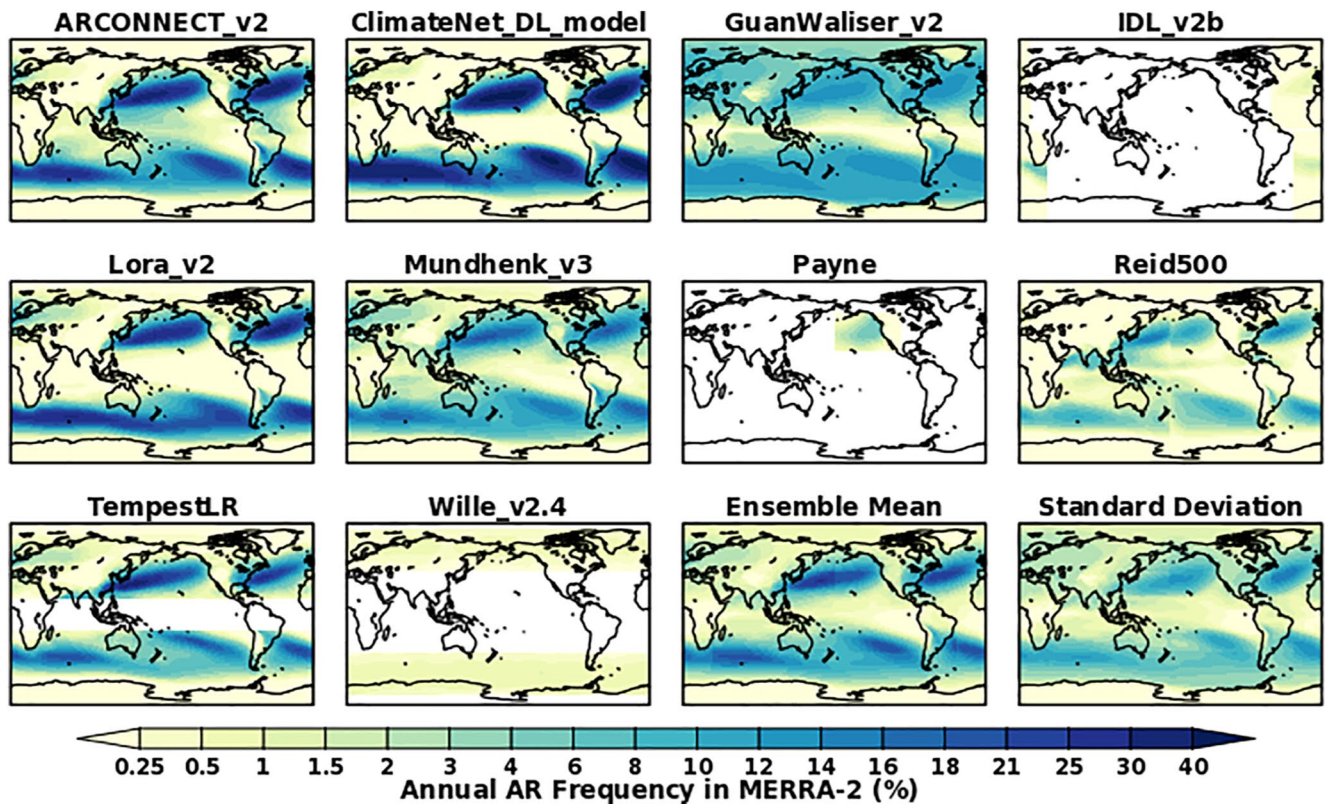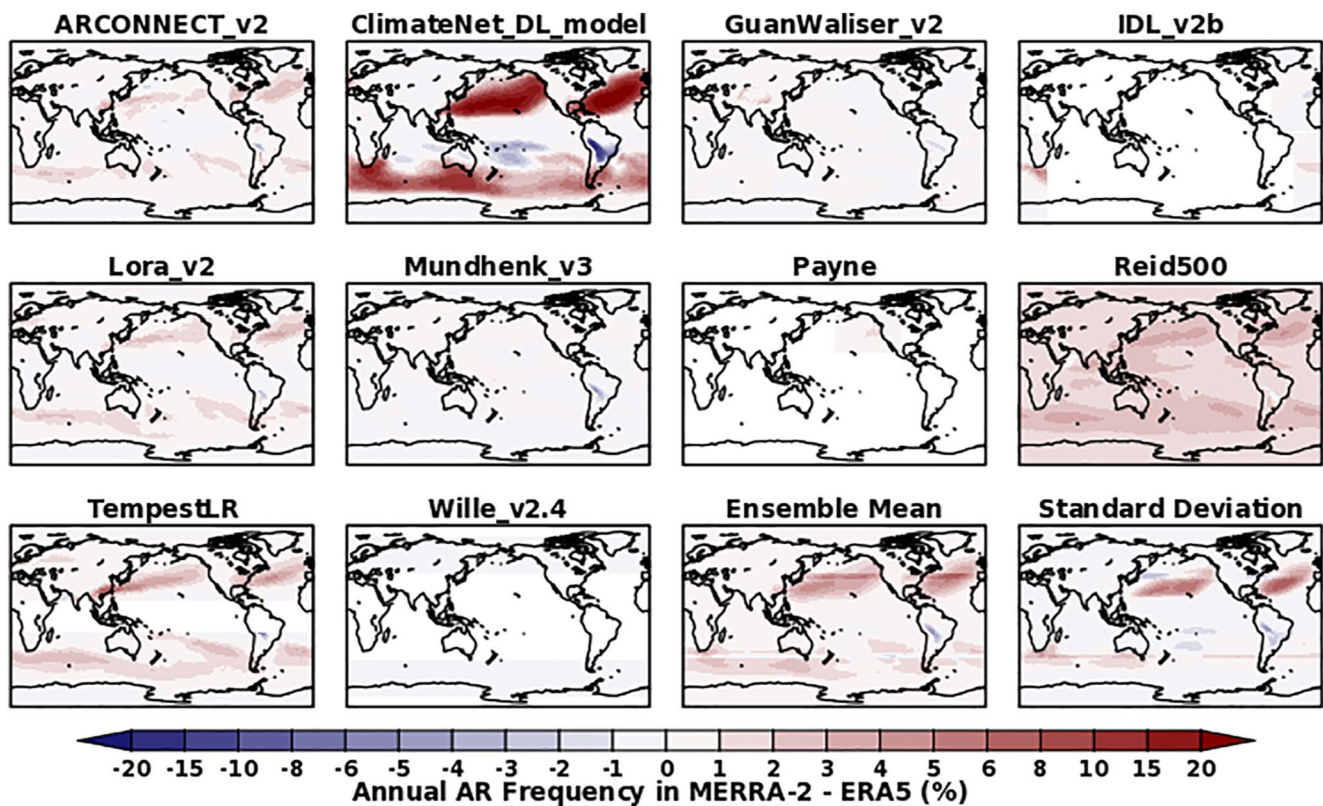
Figure 2. The annual AR frequency of occurrence in MERRA-2 as demonstrated by all contributing algorithms as well as the ensemble mean and standard deviation. White shading indicates no data areas for regional ARDTs (IDL_v2b, Payne, TempestLR, and Wille_v2.4).

MERRA-2 has had more precipitation than ERA5 for the bulk of the 40-year record, but the opposite is true in the mid-1980s and there is excellent agreement between the two beginning in 2015.

4. Results

4.1. Global AR Detection

For a broad, initial comparison, global maps of annual AR frequency of occurrence in each ARDT are assessed. We begin by looking at the bulk frequency of ARs detected by ARDTs in MERRA-2 in Figure 2, and then the difference of the two other reanalyses relative to MERRA-2 in Figures 4 and 5. Spatial maps of annual AR frequency for ERA5 and JRA-55 can be found in the supplemental material. The ensemble mean AR frequency and standard deviation were computed using all ARDTs. Similar spatial patterns of enhanced frequency of ARs emerge in all the ARDTs, with an enhanced frequency of ARs over the northern and southern hemisphere extratropics. While it is tempting to relate these regions of enhanced AR frequency to the extratropical storm tracks, these "AR tracks" should really be considered their own entity as they are distinct features due to the dependence on moisture. Most evident is the heightened frequency of ARs over the ocean, which is to be expected, as well as the poleward tracks in the northern Atlantic and Pacific Oceans and equatorward tracks in the southern Atlantic and Pacific Oceans. Therefore, despite the inclusion of different ARDTs, the ensemble mean is very similar to what is presented in Figure 1a of Lora et al. (2020) using the Tier 1 catalog of ARDTs. The one exception to this is the reduced frequency of ARs along the equatorial Pacific Ocean in the Tier 2 ensemble mean, which is only picked up by ARCONNECT_v2 and Reid500. This disagreement is likely related to the sample of ARDTs included in the analysis. Figure 1a of Lora et al. (2020) discarded the four ARDTs with the highest threshold for IVT, but their supplemental document demonstrated that the frequency of AR occurrence differs in the Pacific ITCZ region depending on whether the complete set of ARDTs is included in the ensemble mean.

Figure 3. The annual AR frequency of occurrence in ERA5 relative to MERRA-2 as demonstrated by all contributing algorithms as well as the ensemble mean and standard deviation.

The expected spatial pattern is present in every global ARDT, at varying magnitudes and gradients. The most ARs are detected by ClimateNet_DL_model, an ARDT that uses deep learning, while the least are detected by Reid 500, which uses a stricter absolute threshold of 500 kg m^{-1} s^{-1}, relative to other ARDTs. GuanWaliser_v2 has the weakest equator to pole gradient in the frequency of ARs, with noticeably more ARs at the poles and over land, and fewer ARs in the "AR tracks". Unlike other algorithms that use a single threshold for IVT to detect ARs, GuanWaliser_v2 incorporates a globally fixed minimum threshold of 100 kg m^{-1} s^{-1} in addition to a regionally dependent threshold of the 85th percentile. Though not to the extent of GuanWalier_v2, Mundhenk_v3 also detects slightly more ARs over land compared to other ARDTs because of the mutual use of the 85th percentile as a relative threshold for IVT.

Regional ARDTs, specifically IDL_v2b, Payne, and Wille_v2.4 detect noticeably fewer ARs than the global ARDTs. The reason for this is different for each algorithm. IDL_v2b seeks to detect only the most extreme ARs, Payne only detects landfalling ARs, and Wille_v2.4 was designed for polar ARs. This results in artifacts in the ensemble means and standard deviations that are unrealistic and is a caveat of combining global and regional algorithms in data sets with varying resolutions. Both can be seen in the ensemble mean and standard deviation for MERRA-2 (Figure 2) but are also noticeable in the difference plots (Figures 3 and 4). A noticeable discontinuity is present at 180° longitude in the northern hemisphere, which is the westernmost extent of Payne. However, Reid500 follows a grid from −180° to 180°, not allowing for ARs to extend through the barrier of 180° longitude, a feature that will be fixed in the next version of the algorithm. As a result, the frequency of ARs to the east of 180° longitude is suppressed for that ARDT. Also evident is the artifact corresponding to IDL_v2b in the east Atlantic and south Indian Oceans. The discontinuities with respect to latitude correspond to Wille_v2.4, with ARs not detected between 42°N and 37°S in MERRA-2 and between 42.25°N and 37.25°S in ERA5. This is because Wille_2.4 was designed for high latitudes such that the threshold and geometry requirements are defined for cold, dry environments while focusing on detecting ARs with relatively high meridional moisture transport.
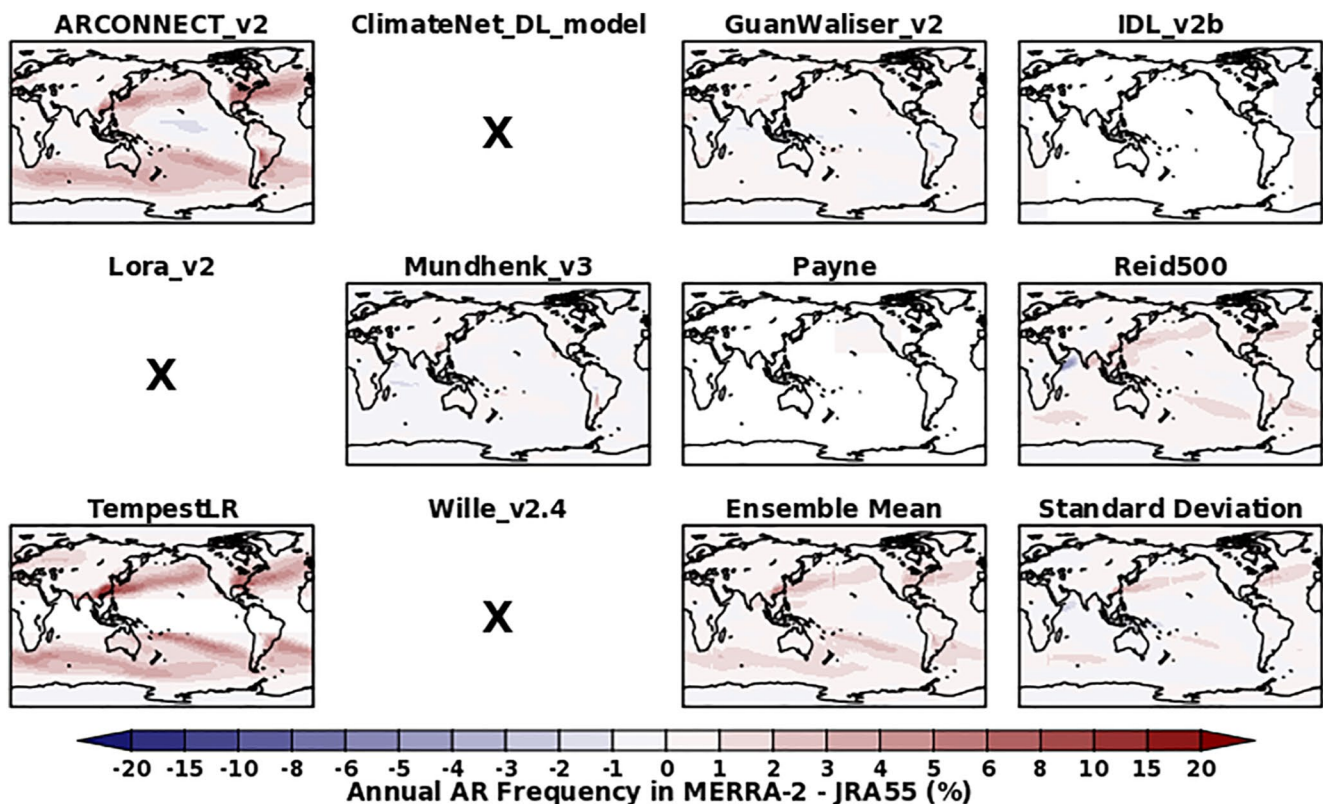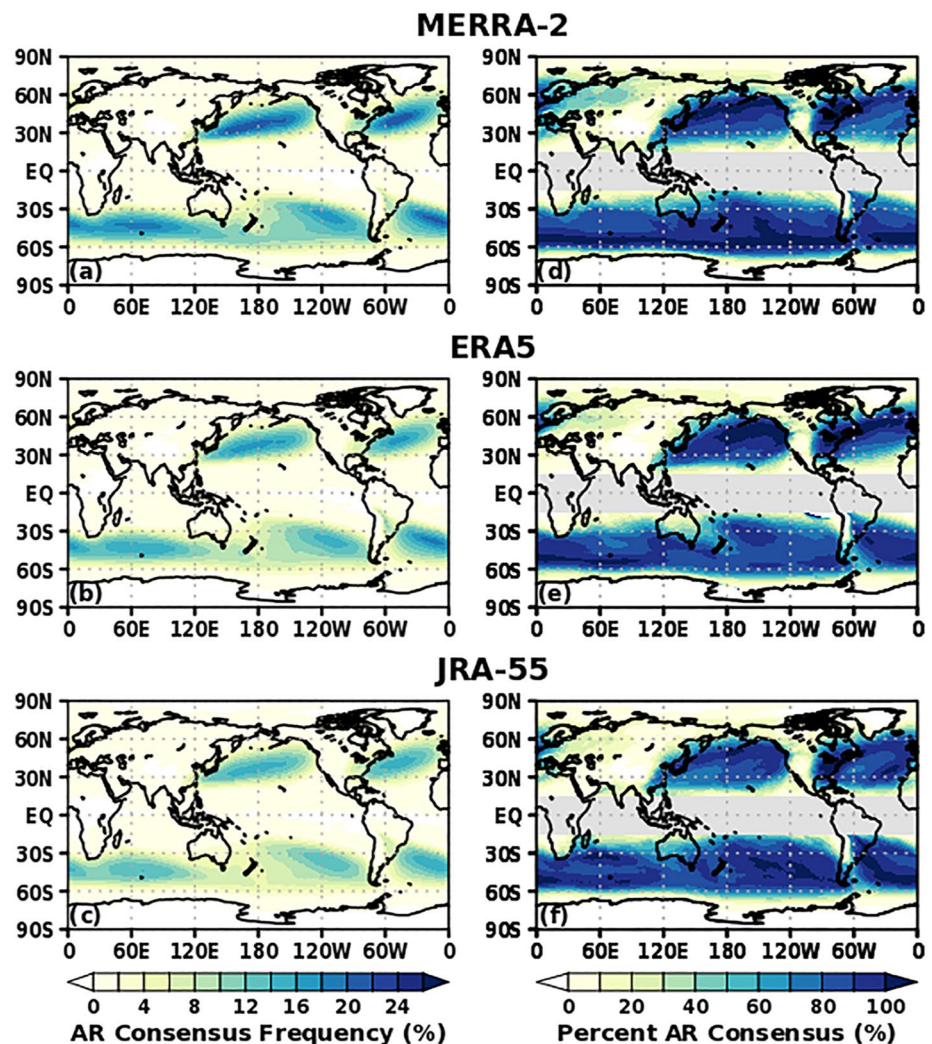
Figure 4. The annual AR frequency of occurrence in JRA-55 relative to MERRA-2 as demonstrated by contributing algorithms as well as the ensemble mean and standard deviation. ARDTs indicated by "X" did not provide data for JRA-55.

When comparing ARs detected in MERRA-2 and ERA5, it is immediately evident that more ARs are detected in MERRA-2 globally and in most of the algorithms. This is because IVT and integrated water vapor (IWV) are larger in magnitude in MERRA-2 nearly everywhere (Figure 1b). The difference between MERRA-2 and ERA5 with respect to AR detection is most notable in ClimateNet_DL_model, which uses IWV for training, though ARs are in general more prominent in MERRA-2 for the algorithms that use an absolute threshold for IVT or use IVT for training: ARCONNECT_v2, Reid500, and TempestLR. This suggests that retuning algorithms with an absolute threshold for a given data product may be necessary, including those that represent gradients in IVT such as in TempestLR.

Although Lora_v2 uses a relative threshold for IVT of 225 kg m^{-1} s^{-1} above a time varying background determined by a 30-day running mean of IWV, perhaps differences in IVT between MERRA-2 and ERA5 are more related to winds and therefore are not captured by this approach. This can be noted by the fact that the difference in AR frequency is larger in Lora_v2 than GuanWaliser_v2 and Mundhenk_v3, which both use a relative threshold of the 85th percentile of IVT in MERRA-2 and ERA5 respectively.

ARDTs that demonstrate good agreement between MERRA-2 and ERA5 do so for JRA-55 as well. There is little difference between MERRA-2 and JRA-55 in GuanWaliser_v2, IDL_v2b, Mundhenk_v3, and Payne. An interesting comparison is how the differences between MERRA-2 and the other two reanalyses contrast between ARCONNECT_v2 and Reid500. While more ARs are detected in MERRA-2 for both algorithms, the difference with JRA-55 is muted compared to ERA5 in Reid500 but amplified in ARCONNECT_v2. Given the absolute threshold, the response in Reid500 makes sense given that global IVT is slightly higher in JRA-55. A similar driver is also likely responsible for lower AR detection rate from TempestLR, in JRA-55 compared to ERA5, as the relatively low resolution of JRA-55 produces weaker gradients in the IVT field. TempestLR relies on the Laplacian of IVT, which makes the gradient of IVT an important factor for AR detection. The larger difference between MERRA-2 and JRA55 in ARCONNECT_v2 is more nuanced. Due the coarser spatial resolution, JRA-55 does not capture the more extreme values of IVT that are on the tails of the distribution in MERRA-2 and

Figure 5. The frequency of occurrence of consensus ARs, with at least half of the global ARDTs detecting an AR, in (a) MERRA-2, (b) ERA5, and (c) JRA55 and the frequency of occurrence of consensus ARs as a percentage of the ensemble mean AR frequency in (d) MERRA-2, (e) ERA5, and (f) JRA55.

ERA5. ARCONNECT_v2 uses a threshold of 700 kg m^{-1} s^{-1} for seeding the object identification. It is likely that this threshold is too strict to detect weak and moderate ARs in JRA-55, though adequate for the other reanalyses. An additional factor might be the 6-hour time increment, allowing ARs to move too far between timesteps for the object identification to capture the continuity of the AR feature.

One exception to the greater frequency of ARs in MERRA-2 is interior South America, and this has to do with the South American Low Level Jet (SALLJ), which transports moisture from the Amazon to the La Plata basin in subtropical South America (Montini et al., 2019; Ramos et al., 2019; Vera et al., 2006). ARDTs pick up the signature of the SALLJ due to its intense moisture transport with greater frequency in ERA5 with respect to MERRA-2. The opposite is true with respect to JRA-55, as MERRA-2 still has more ARs detected in the subtropics of South America. Whether the SALLJ feature is considered an AR is up for debate (Gimeno et al., 2021). A reanalysis comparison of the SALLJ was performed by Montini et al. (2019). Although ERA-Interim was used instead of ERA5, the study concluded that overall, ERA-Interim had a slightly better representation of the SALLJ compared to MERRA-2 and JRA-55, though MERRA-2 is still suitable for studies on the SALLJ (Montini et al., 2019). JRA-55 was unable to capture the peak intensity of the jet likely due to its coarser resolution.

Even if algorithms agree on the frequency of ARs, it is important to know if the same ARs are being detected by the ARDTs in the three reanalysis products. This is accomplished through investigating an AR consensus within

the ensemble (Lora et al., 2020). ARs that are detected by at least half of the ARDTs are considered a consensus AR for the given reanalysis. In order to maximize the number of ARDTs included in the consensus without spatial and temporal inhomogeneities, the consensus AR frequency was determined using global ARDTs only for the period of 2000 through 2019. Since MERRA-2 has more ARs detected in general, the consensus frequency of occurrence is also analyzed as a percentage of the ensemble AR frequency for each reanalysis. The ensemble mean used here was recomputed from what is displayed in Figures 2–4 to match the ARDTs and time period included in the consensus calculation. Note that the tropics have been masked out in the percentage consensus frequency to allow for the inclusion of TempestLR.

The spatial pattern for the consensus AR frequency is similar to the ensemble mean for the reanalyses, with a higher frequency of occurrence in the center of the AR tracks extending poleward in the northern and southern Atlantic and Pacific Oceans and zonally in the southern Indian Ocean (Figure 5, left). As demonstrated by the case studies presented by Lora et al. (2020), ARDTs tend to agree on the core of an AR object though a discrepancy arises on the boundary of the AR periphery particularly for relatively weak ARs. Unsurprisingly, due to the overall higher occurrence of ARs in MERRA-2, the frequency of occurrence for consensus ARs is highest in MERRA-2 in all five AR track regions, followed by ERA5. When the occurrence frequency of the consensus AR is considered as a percentage of the ensemble mean frequency, the largest consensus occurs on the periphery of the AR tracks (Figure 5, right). From this perspective, there is a strong agreement in the ARDTs in ERA5 in the northern hemisphere but relatively weaker agreement in the southern hemisphere. Also noteworthy is the greater percentage of occurrence of the consensus AR in MERRA-2 over land, particularly over Europe, central Australia, and northwestern North America.

Next, we assess whether a given AR is detected in each reanalysis. A consensus AR is defined for the individual reanalyses with the approach above in which at least half of the ARDTs must detect an AR. This "consensus AR" is then compared between MERRA-2, ERA5, and JRA55. Due to the differing temporal and spatial resolutions, consensus ARs in ERA5 and MERRA-2 were regridded to match the 1.25° spatial resolution of JRA-55 at the matching timesteps corresponding to available data in JRA-55. For a comparison between ERA5 and MERRA-2, the hourly temporal resolution was maintained. Due to the higher frequency of ARs in MERRA-2, the consensus AR frequency between MERRA-2 and either ERA5 or JRA-55 is higher than the consensus between ERA5 and JRA-55.

It is also instructive to determine when a consensus AR is detected in ERA5 and JRA-55, but not in MERRA-2, which is termed a "disagreement". Across the entire region, this occurs on average about 1% of the time, however the spatial patterns are different for ERA5 and JRA-55 (Figures 6e and 6f). ERA5 has a larger "disagreement" frequency along the northern periphery of the AR track, in addition to hints of further inland penetration (Figure 6e). On the other hand, JRA-55 tends to have an AR consensus within the AR track when there is not an AR detected in MERRA-2. Maximum frequency of "disagreement" in JRA-55 aligns with the gradient of the consensus AR frequency. This could be related to the 6- hour temporal resolution of the AR tags in JRA-55 relative to the one-hour resolution of MERRA-2. Although the time stamps were matched properly to compute the "disagreement" frequency, there is a possibility that one reanalysis propagates the AR at a faster rate. Resolution could also be playing a role in the ERA5 "disagreement" frequency, but in this case spatial rather than temporal resolution. With a finer spatial resolution and additional vertical levels, there are likely occasions when ERA5 is able to transport additional water vapor around the complex terrain along the U.S. west coast.

While not the most intense or influential AR in the forty-year record, a case study was selected for 18z on 6 November 2014. This case was selected because it is in the required period for ERA5, it contains features that are representative of AR detection in the reanalyses and is a rare situation where the consensus for ERA5 detected an AR in the Pacific Northwest of the United States that was not agreed upon by the MERRA-2 consensus (Figure 7). The fraction of ARDTs registering a detection is used, as opposed to the count, since there are seven global ARDTs available for MERRA-2 and ERA5 but five global ARDTs for JRA-55. There are two features for the event on 6 November 2014, each with its own area of low pressure: one over the Pacific Ocean and the other over the Pacific Northwest of the U.S. The finer resolution of ERA5 results in the detection of narrower ARs for both AR features. This is typical behavior and is likely a contributing factor to the fact that the AR frequency is higher in MERRA-2 (Figure 3). Also evident is the finer scale features in ERA5 such as the bulge around 38 N and 145 W that is not as well resolved in JRA-55. With respect to JRA-55, a smaller fraction of ARDTs detect the core
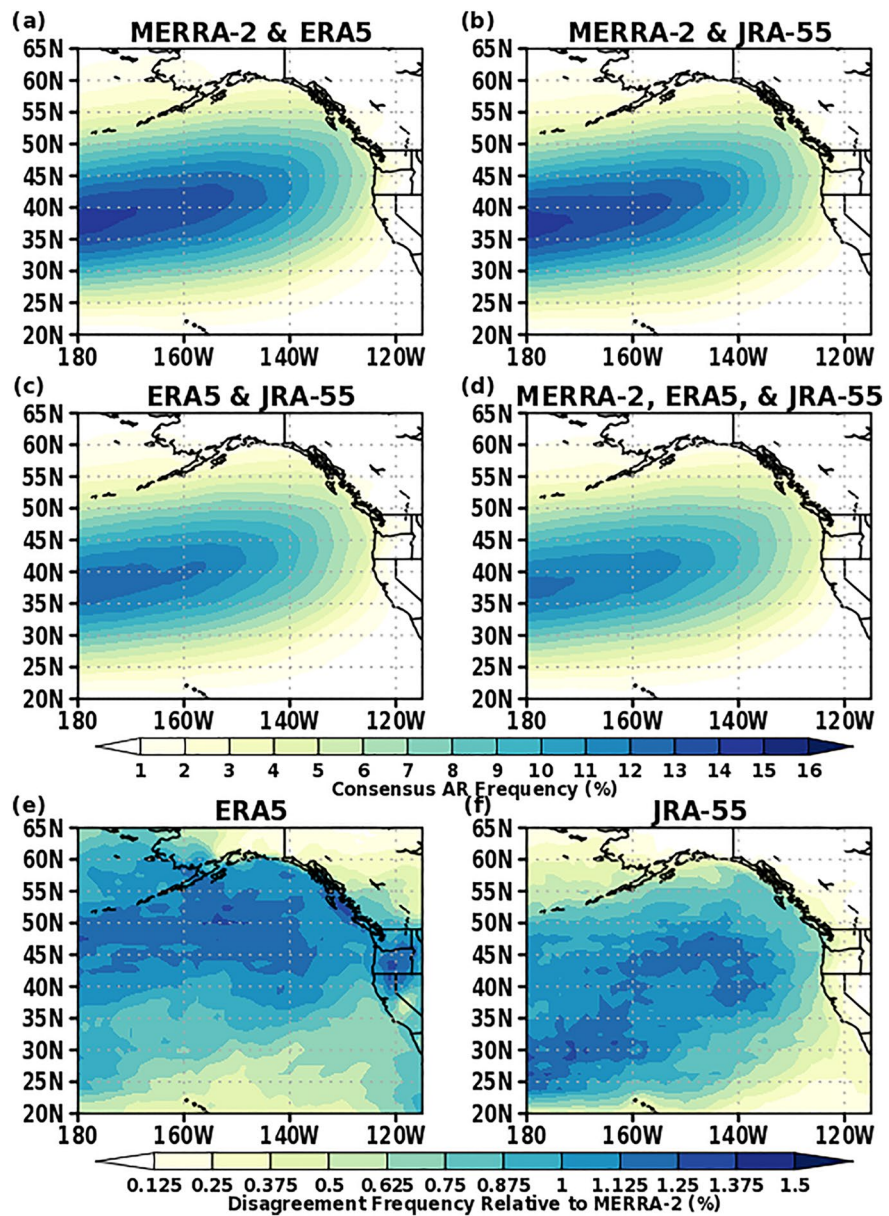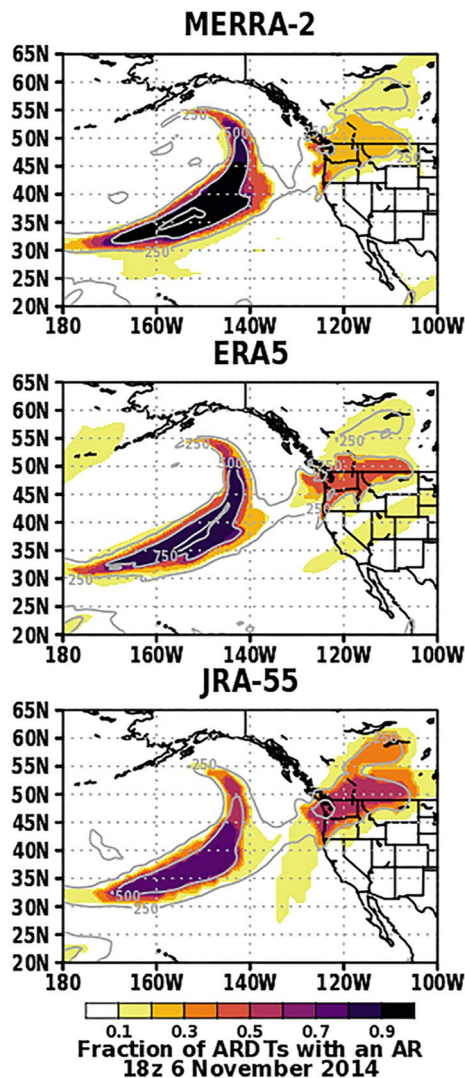
Figure 6. Annual mean frequency of ARs detected in (a) MERRA-2 and ERA5, (b) MERRA-2 and JRA-55, (c) ERA5 and JRA-55, and (d) MERRA-2, ERA5, and JRA-55, as well as the frequency that an AR is detected in (e) ERA5 and (f) JRA-55 but not MERRA-2.

of the AR over the Pacific Ocean, and even less extending northward to the contour for an IVT of 500 kg m^{-1} s^{-1}. While MERRA-2 and ERA5 have an IVT contour exceeding 750 kg m^{-1} s^{-1}, JRA-55 does not.

Conversely, ARDTs tend not to detect the AR feature over the Pacific Northwest in MERRA-2. It is evident that the enhanced IVT has a greater latitudinal extent in MERRA-2 relative to ERA5. The one algorithm that detected an AR in ERA5 but not MERRA-2 in this region was Lora_v2. In the Pacific Northwest region, JRA-55 has the highest values for IVT of the three reanalyses, above 150 kg m^{-1} s^{-1} greater than MERRA-2 just off the Washington coast and over central Washington. As a result, at least half of the ARDTs detect an AR in this region in JRA-55.

For a different perspective of global AR detection, a time series of the mean hemispheric percent area that had an AR for the 20-year period beginning in 2000 is shown in Figure 8. This temporal subset is presented because the

Figure 7. Case study for 18z on 6 November 2014 showing the fraction of global ARDTs that detected an AR. Gray contours indicate lines of constant IVT at the values of 250, 500, and 750 kg m^{-1} s^{-1}.

statistics for ERA5 are more robust when all available ARDTs are included. Shading represents the time series of standard deviation as a measure of variance among the ARDTs. Analogous figures depicting the individual ARDTs can be found in the supplemental document. Like the time series in global mean TPW, there is a subtle increase in the percent area with an AR with time in the three reanalyses for both the northern and southern hemispheres (Figures 8a and 8b). It is unsurprising that the reanalyses are well correlated, although MERRA-2 has a larger area covered by ARs. The average percent area with an AR is lower in JRA-55 relative to MERRA-2 and ERA5. Only one ARDT, Reid500, indicates a percent area with an AR in MERRA-2 that is on par with the ensemble mean for JRA-55 in both hemispheres. The standard deviation for MERRA-2, ERA5, and JRA-55 remains relatively consistent throughout the time series indicating that variability among the ARDTs does not change in time. Overall, the consistency in the annual mean time series and the standard deviation among the reanalyses is important to note considering the relative changes in IVT and TPW as demonstrated by Figure 1.

The seasonal cycle of percent area with an AR is more interesting. The northern hemisphere has a bimodal mean seasonal cycle with peaks in the boreal winter and summer months (Figure 8c). As demonstrated by Mundhenk et al. (2016), ARs peak during the boreal summer months in southwest Alaska, yet the boreal winter months on the U.S. west coast. Gradual transitions in the dynamics associated with atmospheric rivers, for example within the northeast Pacific (Mundhenk et al., 2016), as well as disagreements among the ARDTs result in the bimodal seasonal cycle. Conversely, the southern hemisphere has a seasonal cycle that only peaks during their summer season (Figure 8d).

Unlike the annual mean time series, there is seasonal variability in the standard deviation for both hemispheres in MERRA-2 and ERA5. A seasonal cycle is not present in the standard deviation for JRA-55 due to the ARDTs that submitted the optional AR tags for JRA-55 (see the supplemental document for results in MERRA-2 and ERA5 using the available ARDTs for JRA-55). There are seven global ARDTs included in the calculation for MERRA-2 and ERA5 but five for JRA-55, with the two missing as Lora_v2 and ClimateNet_DL. This is particularly important for the seasonal cycle of standard deviation. Most ARDTs have a pronounced seasonal maximum in the northern hemisphere during the summer months in all three reanalyses. Although a summer peak is still present, Mundenk_v3 has a muted seasonal cycle with respect to the other ARDTs. On the other hand, Lora_v2 and ClimateNet_DL have a pronounced minima during the months of July and August, driving the increase in standard deviation during the boreal summer. Intuitively, it would make sense for differences in the seasonal cycle to be a function of fixed, relative, and absolute thresholds; however, this is not necessarily the case. Each pairing of ARDTs associated with the different seasonal cycles have different types of criteria for IVT. Two of the ARDTs with a maximum in the boreal summer, GuanWaliser_v2 and Reid500, use the same geometry threshold requiring a length of at least 2000 km and a length to width ratio greater than 2. It is difficult to determine why Lora_v2 has an opposing seasonal cycle considering it employs the same length criteria as GuanWaliser_v2 and Reid500, but perhaps this indicates the importance of a length to width ratio in the detection of ARs for regional AR detection.
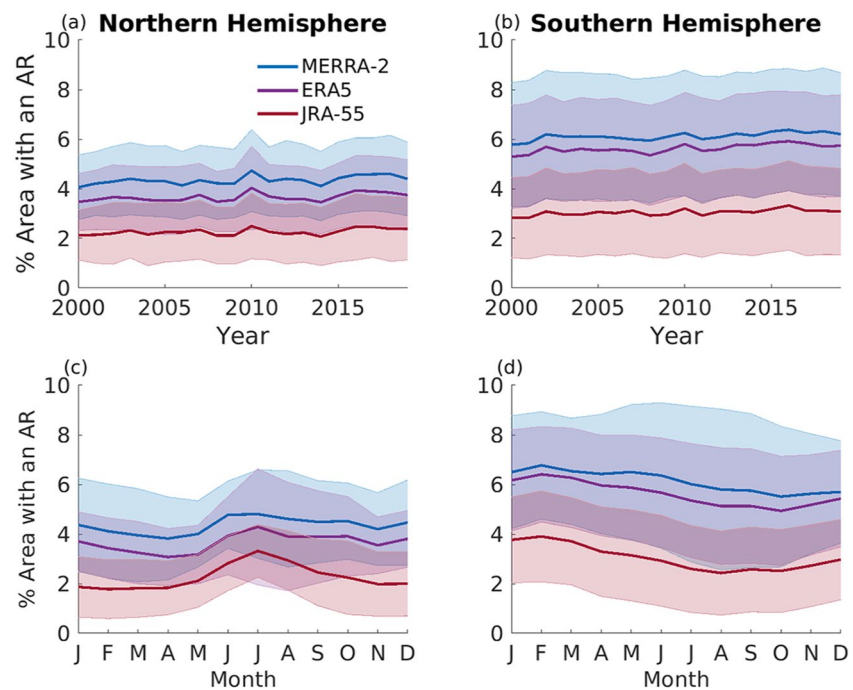
4.2. Regional Results

4.2.1. Coastal AR Frequency

A more nuanced approach for AR detection is evaluated through AR frequency along the coast of three regions. The west coast of the United States and Europe is examined in the same manner as Figure 6 of Rutz et al. (2019),

Figure 8. (a),(b) Annual and (c),(d) monthly mean time series of the percent area covered by ARs in the (a),(c) northern hemisphere and (b),(d) southern hemisphere. Shading represents plus and minus one standard deviation among the ARDTs.

though New Zealand is also considered. Like along the west coast of the United States, the most intense precipitation and flooding events in New Zealand can be attributed to ARs (Reid et al., 2021). Grid points with a fractional coverage of land and ocean between 23 and 55 N and 130 to 115 W for the United States (left panel of Figure 4 in Rutz et al., 2019), 35–65 N and 15 W to 10 E for Europe (right panel of Figure 4 in Rutz et al., 2019), and the western coast of New Zealand between 35 and 46 S were considered. Zonal means of AR frequency are presented along the meridional coastal transects. Unlike Rutz et al. (2019), interior transects are not investigated.

Coastlines are used as opposed to transects to highlight differences between the reanalysis products related to resolution. The ability of each of the reanalyses to resolve the coast lines is immediately evident (Figure 9). Individual features of the coast lines are apparent in ERA5 and not resolved in JRA-55. A good example of this is at 42° N along the west coast of the United States. An indented curve in the coastline of northern California is resolved in ERA5, but not the coarser MERRA-2 or JRA-55, which places the grid box signaling the coast slightly to the east in ERA5. As a result, over half of the ARDTs indicate a dip in the frequency of AR occurrence at that latitude. In the column for New Zealand in Figure 9, it is evident that JRA-55 responds in an opposite manner to the Cook Strait, with less instead of more ARs around 40°S. This furthers the conclusion that JRA-55 may not be suitable for regional AR studies.

There are a lot of common features when comparing the results across regions and reanalyses due to the routinely applied criteria. Regardless of region or reanalysis, more lenient ARDTs such as ClimateNet_DL_model and GuanWaliser_v2 tend to detect more ARs than stricter ARDTs like Reid500 and TempestLR. In addition, ARDTs tend to be more clustered with fewer ARs detected in ERA5 than MERRA-2. While there are other regions that could be investigated such as South America and the poles, it is unlikely that the overall conclusions would change.

### 4.3. Impact of AR Detection on Precipitation

The consensus AR for each reanalysis was used to determine the mean AR related precipitation over the extratropical oceans using the daily, half degree observational data from GPCP version 3.1 (Huffman et al., 2021). Since this is a daily precipitation product, the daily consensus AR was considered as an AR if the grid box had at least one time step over the course of the day with a consensus AR. Results are presented in Figure 10 with
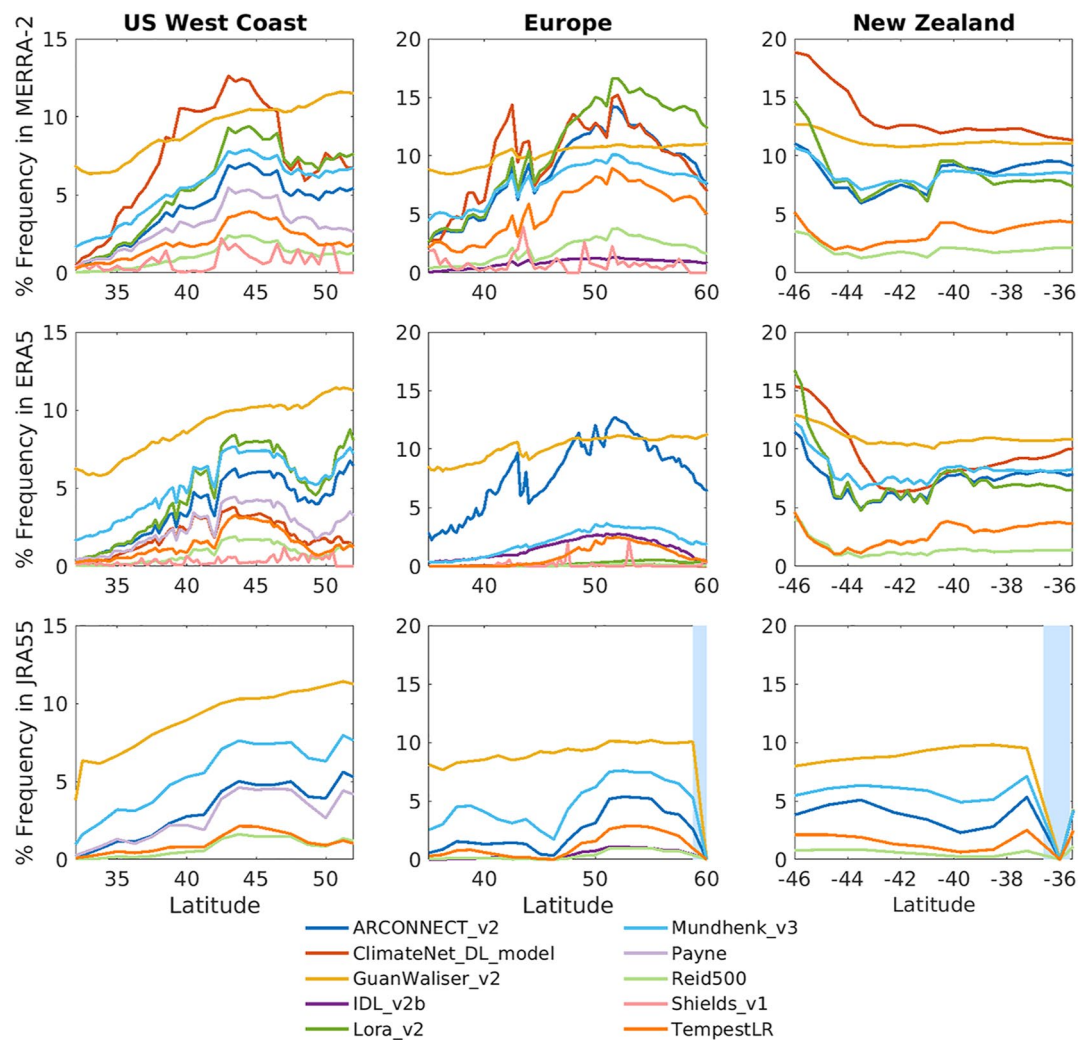
Figure 9. The frequency of occurrence of consensus ARs along the western coast lines of the U.S. (left column), Europe (middle column), and New Zealand for participating ARDTs in MERRA-2 (top row), ERA5 (middle row), and JRA-55 (bottom row). Blue shading indicates latitudes where the coastline is not resolved in JRA-55.

ERA5 and JRA-55 shown as a difference from MERRA-2. Most AR related precipitation falls in the northern hemisphere, along the northeastern periphery of the AR tracks in the Atlantic and Pacific Oceans (Figure 10a). In the southern hemisphere, there is enhanced precipitation along the AR tracks in the Pacific and Atlantic Oceans, but not the southern Indian Ocean (Figure 10d). Overall, more precipitation is associated with ARs in MERRA-2, particularly with respect to JRA-55 (Figure 10c). Where there are differences between the reanalyses, it is in the core of the AR tracks and not along the periphery where the maximum precipitation occurs. Despite an enhanced frequency in MERRA-2 relative to ERA5 in the southern hemisphere, the amount of AR related precipitation is comparable between the two reanalyses.

A well-documented region for ARs is the northeast Pacific, with ARs making landfall on the west coast of the United States and British Columbia. The literature contains numerous studies that have computed the percentage or amount of precipitation that is associated with ARs along the coast of the United States, though there is little agreement due to varying techniques for AR detection, datasets, and seasons (Arabzadeh et al., 2020; Collow et al., 2020; Dettinger et al., 2011; Guan et al., 2010; Guan & Waliser, 2015; Rutz et al., 2014, 2019). Precipitation associated with consensus ARs as well as the ensemble of ARDTs are investigated in the MERRA-2, ERA5, and JRA-55 using an independent observational data set for precipitation, the Climate Prediction Center Unified (CPCU) gauge-based analysis of daily precipitation (Chen & Xie, 2008). Prior to computing any statistics, the precipitation observations were regridded to match the resolution of the individual reanalyses.
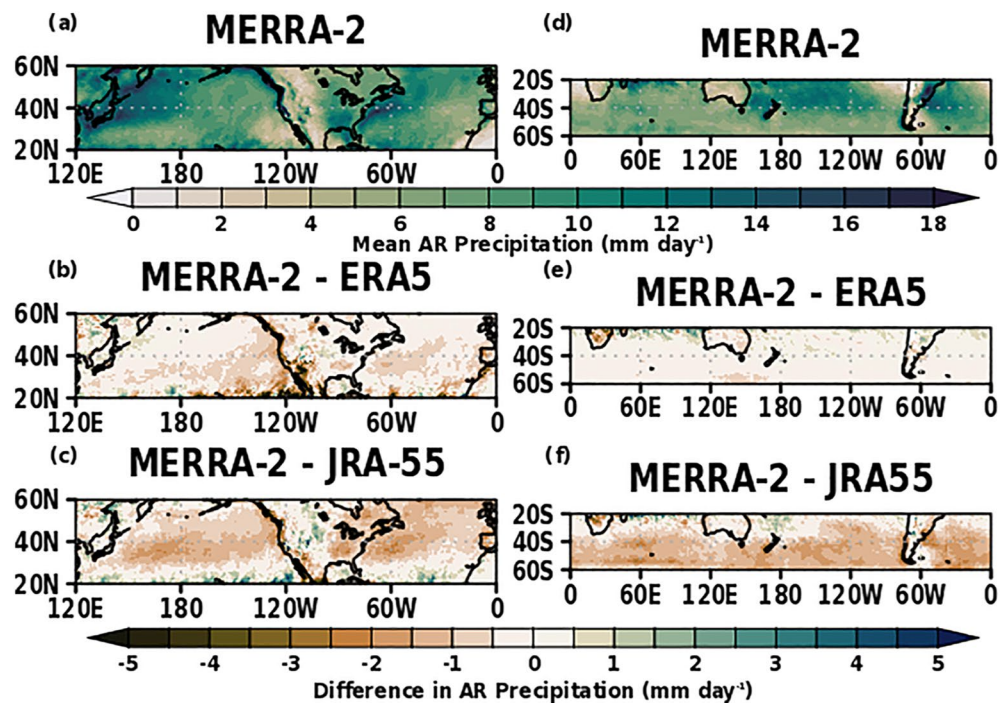
Figure 10. Mean AR precipitation for the (a, b) and (c) northern and (d, e) and (f) southern hemisphere extratropics in MERRA-2 and the difference between (b),(g) MERRA-2 and ERA5 and (c),(f) MERRA-2 and JRA-55.

The overall spatial patterns and magnitudes for the annual mean precipitation associated with consensus ARs in MERRA-2, ERA-5, and JRA-55 are similar due to the influence of topography on precipitation (Figures 11a–11c). Differences in the spatial resolution are evident as the field is smoothed in JRA-55 especially relative to ERA-5. Along the Sierra Nevada Mountains, there is more precipitation associated with ARs in MERRA-2 than ERA5 and JRA-55. ARDTs with strict thresholds tend to result in high values for the mean precipitation associated with an AR as more intense events are included in the average. Given that more ARs are generally detected in MERRA-2, this is somewhat counter-intuitive. This region is, however, located at a gradient where ARDTs begin to no longer detect the inland penetration of ARs. Though difficult to say for sure, it can be hypothesized that the intensity distribution of consensus ARs detected in MERRA-2 in the Sierra Nevada Mountains includes fewer weaker ARs and more intense ARs than ERA5 and JRA55.

Another possibility for the higher mean precipitation in MERRA-2 is that a few ARDTs only submitted AR tags in ERA5 for the 20-year period beginning in 2000. Embedded within the results for precipitation is the temporal trend in precipitation for the region, particularly in southern California and the desert southwest. This is manifested through a lower mean precipitation from ARs (Figure 11b), higher percentage of precipitation resulting from ARs Figure 11e), and more precipitation associated with consensus ARs than the ensemble means (Figure 12b). As a reminder, the ensemble mean incorporates all ARs detected by all ARDTs in the calculation while precipitation associated with the consensus AR requires at least half of the ARDTs to detect an AR, but does not consider which or how many ARDTs detect the AR.

Overall, consensus ARs result in more intense precipitation than the ensemble means along the west coast of the U.S. extending from Washington to central California, and less intense precipitation further inland. ARDTs tend to agree with whether an AR makes landfall, but there is less agreement regarding the inland penetration of ARs. The results for ERA5 presented in Figure 12b are, however, dependent on methodology and the temporal period as mentioned earlier. With respect to the standard deviation of precipitation from the ensemble of ARDTs, the spatial pattern is similar among the three reanalyses and dictated by topography. The magnitude of the standard deviation is noticeably smaller for ERA5.
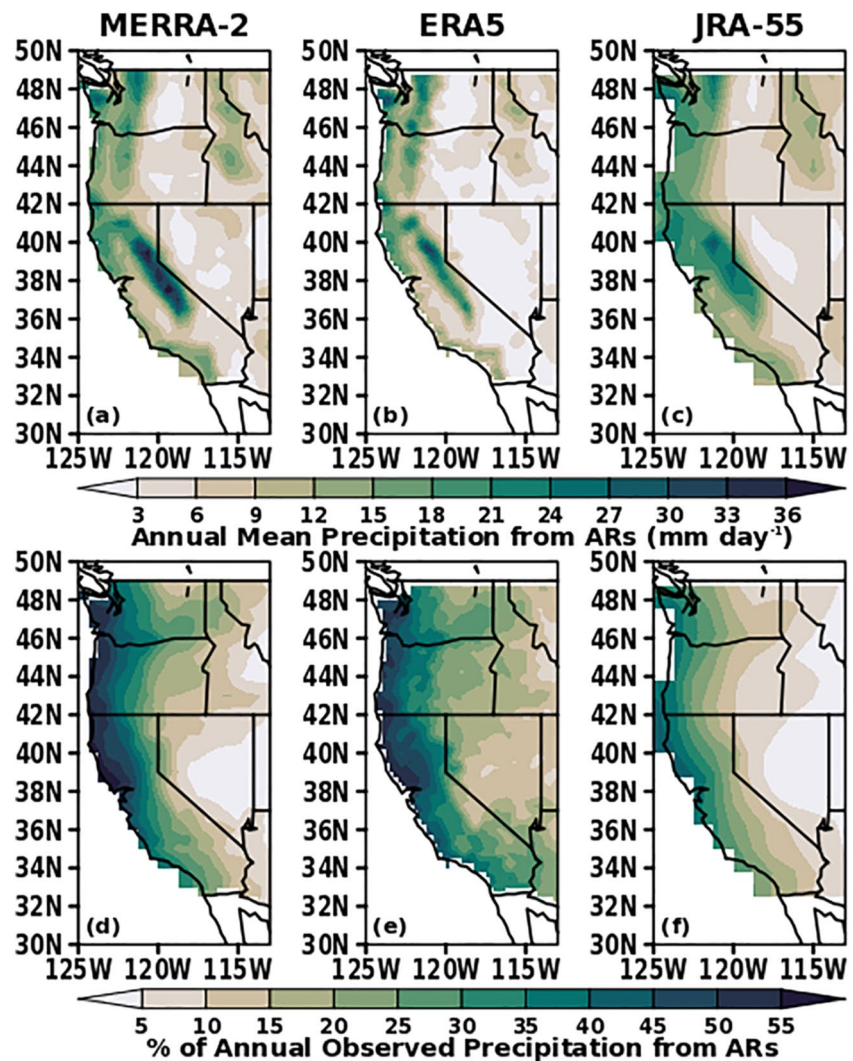
Figure 11. Annual mean intensity of precipitation associated with consensus ARs in (a) MERRA-2, (b) ERA5, and JRA-55 and (d–f) AR cumulative precipitation as a percentage of the total annual precipitation.

5. Summary and Conclusions

There were eleven participating ARDTs in the reanalysis intercomparison that employed a variety of techniques including absolute and relative thresholding, geometry requirements, and machine learning to detect ARs in MERRA-2, ERA5, and JRA55. This data set of hourly AR binary tags in the three reanalyses is now available for the public to use for future studies of AR science. More ARs tended to be detected in MERRA-2 due to elevated values of TPW and IVT relative to ERA5 and JRA55. This indicates that if MERRA-2 were used instead of ERA5 or JRA-55 as a baseline for evaluating a climate model, it would be expected that the climate model would also have an enhanced frequency of ARs, or it would be considered deficient in the frequency of ARs. An additional difference between the reanalyses is that due to the finer resolution, ERA5 has finer ARs with stronger gradients.

ARDTs that used seasonally and spatially varying thresholds, such as GuanWaliser_v2 and Mundhenk_v3, had better agreement between the reanalysis products compared to ARDTs that use a fixed threshold, such as Reid500. Large differences in AR detection between the reanalysis products were also present in ClimateNet_DL_model, which does not use any explicit thresholds, but rather deep learning based on expertly labeled tags. ARTMIP recommendations regarding absolute and relative threshold based ARDTs have been discussed as part of the Tier 1 effort (Rutz et al., 2019), but the variance in the seasonal cycle of global percent area with an AR indicated the importance of geometry thresholds. Differences due to ARDTs are larger than differences due to reanalysis.
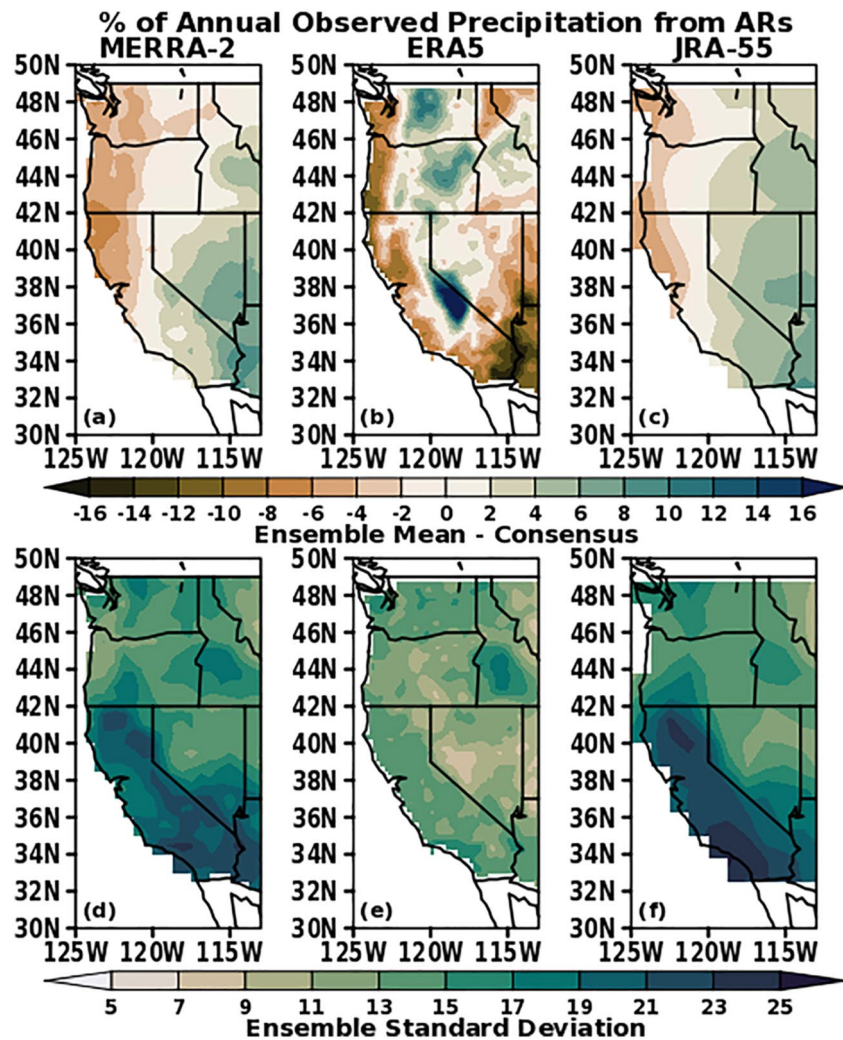
Figure 12. Ensemble mean percent of annual observed precipitation coinciding with detected ARs minus the precipitation from the consensus AR in (a) MERRA-2, (b) ERA5, and JRA-55 and (d–f) the standard deviation of the ensemble.

Despite discrepancies in the frequency of ARs detected in MERRA-2, ERA5, and JRA-55, there is more uncertainty in the spatial pattern of AR frequency due to individual requirements of ARDTs than the reanalysis that is used. This is also the case for global percent area with an AR if comparing the results for MERRA-2 and ERA5. However, differences in the IVT field among the reanalyses along coastlines result in considerable uncertainty in the detection of coastal ARs. Noticeably fewer ARs were detected along the European coastlines by multiple ARDTs in ERA5.

The volume of data from forty years of global, hourly data at one-quarter degree spatial resolution for ERA5 proved challenging. On a global scale, this resulted in time consuming downloads, storage concerns, and memory errors when processing. There were nevertheless rewards of the finer resolution data at a regional scale and for case study purposes. Finer scale features of the anatomy of an AR became apparent as well as a more nuanced view of the impact of coast lines on landfalling ARs. JRA-55 was much easier to work with and provided reasonable results for the global spatial pattern of AR frequency. However, the mean hemispheric percent area with a detected AR was smaller than the other two reanalyses and the coarse resolution was unable to resolve details along coastlines.

The results presented here are not intended to determine which ARDT or reanalysis is best for scientific studies on ARs, but rather to assist with the interpretation of results for future studies. Future work as part of the Tier 2 Reanalysis Intercomparison will focus on topics not addressed here. Specifically, the suitability of century scale

reanalyses for determining observed changes in the characteristics of ARs, global impacts of ARs on precipitation and extreme precipitation events, furthering our understanding of how to handle regional differences in AR detection, and determining whether a specific reanalysis is best used for the evaluation of climate models.

Data Availability Statement

All ARTMIP data are available from the Climate Data Gateway (doi: 10.26024/rawv-yx53). MERRA-2 is available from the Goddard Earth Sciences Data and Information Services Center (GES DISC) at https://disc.gsfc.nasa.gov/, ERA5 data is available from the Copernicus Climate Change Service (C3S) Climate Data Store at https://cds.climate.copernicus.eu/ and JRA-55 and JRA-55C data is available from the NCAR Research Data Archive at https://rda.ucar.edu/. CPC US Unified Precipitation data was provided by the NOAA/OAR/ESRL PSL, Boulder, Colorado, USA, from https://psl.noaa.gov/data/gridded/data.unified.daily.conus.html. Monthly GPCP Precipitation data was provided by the NOAA/OAR/ESRL PSL, Boulder, Colorado, USA, from https://psl.noaa.gov/data/gridded/data.gpcp.html. Daily GPCP precipitation data was provided by GESDISC at https://disc.gsfc.nasa.gov/datasets/GPCPDAY_3.1/summary.

References

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., et al. (2003). The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979-Present). *Journal of Hydrometeorology*, *4*, 11472–1167. https://doi.org/10.1175/1525-7541(2003)004<1147:tvgpcp>2.0.co;2

Arabzadeh, A., Ehsani, M. R., Guan, B., Heflin, S., & Behrangi, A. (2020). Global intercomparison of atmospheric rivers precipitation in remote sensing and reanalysis products. *Journal of Geophysical Research: Atmospheres*, *125*, e2020JD033021. https://doi.org/10.1029/2020JD033021

Benedict, J. J., Clement, A. C., & Medeiros, B. (2019). Atmospheric blocking and other large-scale precursor patterns of landfalling atmospheric rivers in the North Pacific: A CESM2 study. *Journal of Geophysical Research: Atmospheres*, *124*, 11330–11353. https://doi.org/10.1029/2019JD030790

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

Bosilovich, M. G., Robertson, F. R., Takacs, L., Molod, A., & Mocko, D. (2017). Atmospheric water balance and variability in the MERRA-2 reanalysis. *Journal of Climate*, *30*(4), 1177–1196. https://doi.org/10.1175/JCLI-D-16-0338.1

Chen, M., & Xie, P. (2008). *CPC unified gauge-based analysis of global daily precipitation. Western Pacific Geophysics Meeting*. American Geophysical Union. Abstract A24A-05.

Chen, X., Leung, L. R., Gao, Y., Liu, Y., Wigmosta, M., & Richmond, M. (2018). Predictability of extreme precipitation in western U.S. watersheds based on atmospheric river occurrence, intensity, and duration. *Geophysical Research Letters*, *45*(11), 693701–693711. https://doi.org/10.1029/2018GL079831

Chen, X., Leung, L. R., Wigmosta, M., & Richmond, M. (2019). Impact of atmospheric rivers on surface hydrological processes in western U.S. watersheds. *Journal of Geophysical Research: Atmospheres*, *124*, 8896–8916. https://doi.org/10.1029/2019JD030468

Collow, A. B. M., Mersiovsky, H., & Bosilovich, M. G. (2020). Large-scale influences on atmospheric river–induced extreme precipitation events along the coast of Washington State. *Journal of Hydrometeorology*, *21*, 2139–2156. https://doi.org/10.1175/JHM-D-19-0272.1

Dettinger, M. D. (2013). Atmospheric rivers as drought busters on the U.S. West Coast. *Journal of Hydrometeorology*, *14*(6), 1721–1732. https://doi.org/10.1175/JHM-D-13-02.1

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*, 445–478. https://doi.org/10.3390/w3020445

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The modern-era retrospective analysis for research and applications, Version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Gimeno, L., Algarra, I., Eiras-Barca, J., Ramos, A. M., & Nieto, R. (2021). Atmospheric River, a term encompassing different meteorological patterns. *Wiley Interdisciplinary Reviews: Water*, *8*(6), e1558. https://doi.org/10.1002/wat2.1558

Global Modeling & Assimilation Office (GMAO) (2015b). *MERRA-2 inst3_3d_asm_Np: 3d,3-Hourly,Instantaneous,Pressure-Level,Assimilation,Assimilated Meteorological Fields V5.12.4*. Goddard Earth Sciences Data and Information Services Center (GES DISC). Accessed: 26 April 2021. https://doi.org/10.5067/QBZ6MG944HW0

Global Modeling & Assimilation Office (GMAO) (2015a). *MERRA-2 tavg1_2d_int_Nx: 2d,1-Hourly,Time-Averaged,Single-Level,Assimilation,Vertically Integrated Diagnostics V5.12.4*. Goddard Earth Sciences Data and Information Services Center (GES DISC). Accessed: 1 April 2020. https://doi.org/10.5067/Q5GVUVUIVGO7

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

Guan, B., Molotch, N. P., Waliser, D. E., Fetzer, E. J., & Neiman, P. J. (2010). Extreme snowfall events linked to atmospheric rivers and surface air temperature via satellite measurements. *Geophysical Research Letters*, *37*, L20401. https://doi.org/10.1029/2010GL044696

Guan, B., & Waliser, D. E. (2015). Detection of atmospheric rivers: Evaluation and application of an algorithm for global studies. *Journal of Geophysical Research: Atmospheres*, *120*(24), 12514–12535. https://doi.org/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: Atmospheres*, *122*, 5556–5581. https://doi.org/10.1002/2016JD026174

Guan, B., & Waliser, D. E. (2019). Tracking atmospheric rivers globally: Spatial distributions and temporal evolution of life cycle characteristics. *Journal of Geophysical Research: Atmospheres*, *124*, 12523–12552. https://doi.org/10.1029/2019JD031205

Guan, B., Waliser, D. E., & Ralph, F. M. (2018). An intercomparison between reanalysis and dropsonde observations of the total water vapor transport in individual atmospheric rivers. *Journal of Hydrometeorology*, *19*(2), 321–337. https://doi.org/10.1175/jhm-d-17-0114.1

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*, 1999–2049. https://doi.org/10.1002/qj.3803

Hu, J. M., & Nolin, A. W. (2019). Snowpack contributions and temperature characterization of landfalling atmospheric rivers in the Western Cordillera of the United States. *Geophysical Research Letters*, *46*, 6663–6672. https://doi.org/10.1029/2019GL083564

Huang, H., Patricola, C. M., Bercos-Hickey, E., Zhou, Y., Rhoades, A., Risser, M. D., & Collins, W. D. (2021). Sources of subseasonal-to-seasonal predictability of atmospheric rivers and precipitation in the western United States. *Journal of Geophysical Research: Atmospheres*, *126*, e2020JD034053. https://doi.org/10.1029/2020JD034053

Huffman, G. J., Behrangi, A., Bolvin, D. T., & Nelkin, E. J. (2021). In G. J. Huffman, A. Behrangi, D. T. Bolvin, E. J. Nelkin, & M. Greenbelt (Eds.), *GPCP version 3.1 Daily Precipitation Data Set*. USA, Goddard Earth Sciences Data and Information Services Center (GES DISC). Accessed: 5 January 2022. https://doi.org/10.5067/MEASURES/GPCP/DATA303

Huning, L. S., Guan, B., Waliser, D. E., & Lettenmaier, D. P. (2019). Sensitivity of seasonal snowfall attribution to atmospheric rivers and their reanalysis-based detection. *Geophysical Research Letters*, *46*, 794–803. https://doi.org/10.1029/2018GL080783

Ionita, M., Nagavciuc, V., & Guan, B. (2020). Rivers in the sky, flooding on the ground: The role of atmospheric rivers in inland flooding in central Europe. *Hydrology and Earth System Sciences*, *24*(11), 5125–5147. https://doi.org/10.5194/hess-24-5125-2020

Kim, S., & Chiang, J. C. H. (2021). Atmospheric river lifecycle characteristics shaped by synoptic conditions at Genesis. *International Journal of Climatology*, *1*, 18–538. https://doi.org/10.1002/joc.7258

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., et al. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan*, *93*, 5–48. https://doi.org/10.2151/jmsj.2015-001

Lavers, D. A., Allan, R. P., Wood, E. F., Villarini, G., Brayshaw, D. J., & Wade, A. J. (2011). Winter floods in Britain are connected to atmospheric rivers. *Geophysical Research Letters*, *38*, L23803. https://doi.org/10.1029/2011GL049783

Lavers, D. A., Villarini, G., Allan, R. P., Wood, E. F., & Wade, A. J. (2012). 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*, *117*, D20106. https://doi.org/10.1029/2012JD018027

Lora, J. M., Mitchell, J. L., Risi, C., & Tripati, A. E. (2017). North Pacific atmospheric rivers and their influence on western North America at the Last Glacial Maximum. *Geophysical Research Letters*, *44*(2), 1051–1059. https://doi.org/10.1002/2016GL071541

Lora, J. M., Shields, C. A., & Rutz, J. J. (2020). Consensus and disagreement in Atmospheric River detection: ARTMIP global catalogues. *Geophysical Research Letters*, *47*, e2020GL089302. https://doi.org/10.1029/2020GL089302

Massoud, E. C., Espinoza, V., Guan, B., & Waliser, D. E. (2019). Global climate model ensemble approaches for future projections of atmospheric rivers. *Earth's Future*, *7*, 1136–1151. https://doi.org/10.1029/2019EF001249

McCarty, W., Coy, L., Gelaro, R., Huang, A., Merkova, D., Smith, E. B., et al. (2016). *MERRA-2 Input Observations: Summary and Assessment. NASA Technical Report Series on Global Modeling and Data Assimilation* (Vol. 46, p. 61). NASA/TM-2016-104606. Retrieved from https://gmao.gsfc.nasa.gov/pubs/docs/McCarty885.pdf

McClenny, E. E., Ullrich, P. A., & Grotjahn, R. (2020). Sensitivity of atmospheric river vapor transport and precipitation to uniform sea surface temperature increases. *Journal of Geophysical Research: Atmospheres*, *125*, e2020JD033421. https://doi.org/10.1029/2020JD033421

Montini, T. L., Jones, C., & Carvalho, L. M. V. (2019). The South American low-level jet: A new climatology, variability, and changes. *Journal of Geophysical Research: Atmospheres*, *124*, 1200–1218. https://doi.org/10.1029/2018JD029634

Mundhenk, B. D., Barnes, E. A., & Maloney, E. D. (2016). All-season climatology and variability of atmospheric river frequencies over the North Pacific. *Journal of Climate*, *29*(13), 4885–4903. https://doi.org/10.1175/JCLI-D-15-0655.1

Neiman, P. J., Schick, L. J., Ralph, F. M., Hughes, M., & Wick, G. A. (2011). Flooding in western Washington: The connection to atmospheric rivers. *Journal of Hydrometeorology*, *12*, 1337–1358. https://doi.org/10.1175/2011JHM1358.1

O'Brien, T. A., Payne, A. E., Shields, C. A., Rutz, J., Brands, S., Castellano, C., et al. (2020). Detection uncertainty matters for understanding atmospheric rivers. *Bulletin of the American Meteorological Society*, *101*, E790–E796. https://doi.org/10.1175/BAMS-D-19-0348.1

O'Brien, T. A., Wehner, M. F., Payne, A. E., Shields, C. A., Rutz, J. J., Leung, L. R., et al. (2021). Increases in future AR count and size: Overview of the ARTMIP Tier 2 CMIP5/6 Experiment. *Earth and Space Science Open Archive*, *31*. https://doi.org/10.1002/essoar.10504170.1

Park, C., Son, S.-W., & Kim, H. (2021). Distinct features of atmospheric rivers in the early versus late East Asian summer monsoon and their impacts on monsoon rainfall. *Journal of Geophysical Research: Atmospheres*, *126*, e2020JD033537. https://doi.org/10.1029/2020JD033537

Payne, A. E., Demory, M.-E., Leung, L. R., Ramos, A. M., Shields, C. A., Rutz, J. J., et al. (2020). Responses and impacts of atmospheric rivers to climate change. *Nature Reviews Earth & Environment*, *1*(3), 143–157. https://doi.org/10.1038/s43017-020-0030-5

Payne, A. E., & Magnusdottir, G. (2014). Dynamics of landfalling atmospheric rivers over the North Pacific in 30 years of MERRA Reanalysis. *Journal of Climate*, *27*(18), 7133–7150. https://doi.org/10.1175/JCLI-D-14-00034.1

Payne, A. E., & Magnusdottir, G. (2015). 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*, 11–173. https://doi.org/10.1002/2015JD023586

Prabhat, K., Mudigonda, M., Kim, S., Kapp-Schwoerer, L., Graubner, A., Karaismailoglu, E., et al. (2021). ClimateNet: An expert-labeled open dataset and deep learning architecture for enabling high-precision analyses of extreme weather. *Geoscientific Model Development*, *14*, 107–124. https://doi.org/10.5194/gmd-14-107-2021

Ralph, F. M., Dettinger, M. D., Cairns, M. M., Galarneau, T. J., & Eylander, J. (2018). Defining "Atmospheric River": How the glossary of meteorology helped resolve a debate. *Bulletin of the American Meteorological Society*, *99*(4), 837–839. https://doi.org/10.1175/BAMS-D-17-0157.1

Ralph, F. M., Neiman, P. J., Wick, G. A., GutmanDettinger, S. I. M. D., Cayan, D. R., White, A. B., & White, A. B. (2006). Flooding on California's Russian River: Role of atmospheric rivers. *Geophysical Research Letters*, *33*, L13801. https://doi.org/10.1029/2006GL026689

Ralph, F. M., Wilson, A. M., Shulgina, T., Kawzenuk, B., Sellars, S., Rutz, J. J., et al. (2019). ARTMIP-early start comparison of Atmospheric River detection tools: How many atmospheric rivers hit northern California's Russian River watershed? *Climate Dynamics*, *52*, 4973–4994. https://doi.org/10.1007/s00382-018-4427-5

Ramos, A. M., Blamey, R. C., Algarra, I., Nieto, R., Gimeno, L., Tomé, R., et al. (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*, 217–230. https://doi.org/10.1111/nyas.139602019

Ramos, A. M., Nieto, R., Tomé, R., Gimeno, L., Trigo, R. M., Liberato, M. L. R., & Lavers, D. A. (2016). Atmospheric rivers moisture sources from a Lagrangian perspective. *Earth System Dynamics*, *7*(2), 371–384. https://doi.org/10.5194/esd-7-371-2016

Ramos, A. M., Trigo, R. M., Liberato, M. L. R., & Tomé, R. (2015). Daily precipitation extreme events in the Iberian Peninsula and its association with atmospheric rivers. *Journal of Hydrometeorology*, *16*(2), 579–597. https://doi.org/10.1175/JHM-D-14-0103.1

Reid, K. J., King, A. D., Lane, T. P., & Short, E. (2020). The sensitivity of atmospheric river identification to integrated water vapor transport threshold, resolution, and regridding method. *Journal of Geophysical Research: Atmospheres*, *125*, e2020JD032897. https://doi.org/10.1029/2020JD032897

Reid, K. J., Rosier, S. M., Harrington, L. J., King, A. D., & Lane, T. P. (2021). Extreme rainfall in New Zealand and its association with Atmospheric Rivers. *Environmental Research Letters*, *16*, 044012. https://doi.org/10.1088/1748-9326/abeae0

Rhoades, A. M., Jones, A. D., Srivastava, A., Huang, H., O'Brien, T. A., Patricola, C. M., et al. (2020). The shifting scales of western U.S. landfalling atmospheric rivers under climate change. *Geophysical Research Letters*, *47*, e2020GL089096. https://doi.org/10.1029/2020GL089096

Rutz, J. J., Shields, C. A., Lora, J. M., Payne, A. E., Guan, B., Ullrich, P., et al. (2019). The Atmospheric River tracking method intercomparison project (ARTMIP): Quantifying uncertainties in Atmospheric River climatology. *Journal of Geophysical Research: Atmospheres*, *124*, 13777–13802. https://doi.org/10.1029/2019JD030936

Rutz, J. J., Steenburgh, W. J., & Ralph, F. M. (2014). Climatological characteristics of atmospheric rivers and their inland penetration over the Western United States. *Monthly Weather Review*, *142*(2), 905–921. https://doi.org/10.1175/MWR-D-13-00168.1

Ryoo, J., Kaspi, Y., Waugh, D. W., Kiladis, G. N., Waliser, D. E., Fetzer, E. J., & Kim, J. (2013). Impact of Rossby Wave Breaking on U.S. West Coast Winter Precipitation during ENSO Events. *Journal of Climate*, *26*(17), 6360–6382. https://doi.org/10.1175/JCLI-D-12-00297.1

Shearer, E. J., Nguyen, P., Sellars, S. L., Analui, B., Kawzenuk, B., Hsu, K., & Sorooshian, S. (2020). Examination of global midlatitude atmospheric river lifecycles using an object-oriented methodology. *Journal of Geophysical Research: Atmospheres*, *125*, e2020JD033425. https://doi.org/10.1029/2020JD033425

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

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

Shields, C. A., Rosenbloom, N., Bates, S., Hannay, C., Hu, A., Payne, A. E., et al. (2019). Meridional heat transport during atmospheric rivers in high-resolution CESM climate projections. *Geophysical Research Letters*, *46*, 14702–14712. https://doi.org/10.1029/2019GL085565

Shields, C. A., Rutz, J. J., Leung, L.-Y., Ralph, F. M., Wehner, M., Kawzenuk, B., et al. (2018). Atmospheric River Tracking Method Intercomparison Project (ARTMIP): Project goals and experimental design. *Geoscientific Model Development*, *11*(6), 2455–2474. https://doi.org/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 mid-Holocene. *Earth and Planetary Science Letters*, *541*, 116293. https://doi.org/10.1016/j.epsl.2020.116293

Vera, C., Baez, J., Douglas, M., Emmanuel, C. B., Marengo, J., Meitin, J., et al. (2006). The South American Low-Level Jet Experiment. *Bulletin of the American Meteorological Society*, *87*(1), 63–78. https://doi.org/10.1175/BAMS-87-1-63

Waliser, D., & Guan, B. (2017). Extreme winds and precipitation during landfall of atmospheric rivers. *Nature Geoscience*, *10*(3), 179–183. https://doi.org/10.1038/ngeo2894

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*, 911–916. https://doi.org/10.1038/s41561-019-0460-1

Wille, J. D., Favier, V., Gorodetskaya, I. V., Agosta, C., Kittel, C., Beeman, J. C., et al. (2021). Antarctic atmospheric river climatology and precipitation impacts. *Journal of Geophysical Research: Atmospheres*, *126*, e2020JD033788. https://doi.org/10.1029/2020JD033788

Xue, Y., Li, J., Menzel, W. P., Borbas, E., Ho, S.-P., Li, Z., & Li, J. (2019). Characteristics of satellite sampling errors in total precipitable water from SSMIS, HIRS, and COSMIC observations. *Journal of Geophysical Research: Atmospheres*, *124*, 6966–6981. https://doi.org/10.1029/2018JD030045

Zavadoff, B. L., & Kirtman, B. P. (2020). Dynamic and thermodynamic modulators of European Atmospheric Rivers. *Journal of Climate*, *33*(10), 4167–4185. https://doi.org/10.1175/JCLI-D-19-0601.1

Zhou, Y., & Kim, H. (2019). Impact of distinct origin locations on the life cycles of landfalling atmospheric rivers over the U.S. West Coast. *Journal of Geophysical Research: Atmospheres*, *124*, 11897–11909. https://doi.org/10.1029/2019jd031218

Zhou, Y., O'Brien, T. A., Ullrich, P. A., Collins, W. D., Patricola, C. M., & Rhoades, A. M. (2021). Uncertainties in atmospheric river lifecycles by detection algorithms: Climatology and variability. *Journal of Geophysical Research: Atmospheres*, *126*, e2020JD033711. https://doi.org/10.1029/2020JD033711

Zhu, Y., & Newell, R. E. (1998). A proposed algorithm for moisture fluxes from atmospheric rivers. *Monthly Weather Review*, *126*. https://doi.org/10.1175/1520-0493(1998)126<0725:apafmf>2.0.co;2

Reference From the Supporting Information

Eiras-Barca, J., Ramos, A. M., Algarra, I., Vázquez, M., Dominguez, F., Miguez-Macho, G., et al. (2021). European West Coast atmospheric rivers: A scale to characterize strength and impacts. *Weather and Climate Extremes*, *31*, 100305. https://doi.org/10.1016/j.wace.2021.100305