Referee Comments are in red text with italics. Author responses are in black text. New additions to the manuscript, where applicable, are noted *in italics*.

Response to GMDD Interactive comment from Referee #1

Questions to answer in overview:

I think the attempt to compare and synthesize different methods and metrics for the identification of ARs is indeed a very valuable and important effort. I wondered if and how systematic model related properties are treated for the evaluation, e.g which role plays horizontal and vertical, temporal resolution for the results and the applicability of the metrics? Is it possible to add a global (or at least northern hemispheric) view for the test period?

Thank you for your review and noting the importance of ARTMIP. During the course of Tier 1, ARTMIP participants will be disentangling these great questions.

We will absolutely include a global view in a Tier 1 overview paper, which we are actively working on now. However, for this GMDD paper (and the results from the 1-month test we conducted last Fall), the purpose is twofold: 1) document the experimental design for the community and outline the goals for the project, and 2) act as a proof-of-concept that such an intercomparison will work and produce interesting results. Given that we are actively working with the full algorithm catalogues (i.e, the 1980-2017 MERRA-2 period), we feel that we can better use our resources and serve the community by presenting the full results in our science overview paper, rather than add an additional plot with a global view to this experimental design paper. The figures we present here are intended to show the flavor of the types of analysis we are doing but are not necessarily meant to be final "results". We expect some of these metrics may change after looking at the full Tier 1 data, especially so given that have a larger pool of catalogues for Tier 1 is available compared to those analyzed for the proof-of-concept. We have added a comment to the paper to make clear that we intend to present global metrics in future research. We have also included a global snapshot of MERRA-2 IVT data with sample ARs labelled in the supplementary material to illustrate example events across the globe.

In terms of temporal resolution, we begin to address the issue in the supplemental text (Figures S1 and S2) where we show the difference between MERRA-2 3 hourly derived data versus 1-hourly data. As for horizontal resolution, all algorithms are applied to the same MERRA-2 data (~50km) for Tier 1, and 25km data for Tier 2-climate change. Note that other Tier 2 plans include intercomparing different reanalysis products at different resolutions.

Additions to manuscript: New text can be found on pages 23 and 27 in the updated manuscript. The new figure is in the supplemental material, Figure S4. We chose February 19th 21Z to tie into Figure 7 on duration where each algorithm detection tag is shown by the black dots.

General points:

Though the comparison of methods is essential and required to judge results of larger model simulations I missed the link to observations. How do the algorithms do compare with observations of atmospheric water vapor columns e.g. from satellite observations? Wouldn't it be useful to define criteria (or refine the definition of) ARs on the basis of satellite observations to allow to estimate the capability of algorithms and methods to identify the structures?

We have actually spent a great deal of time thinking about how we might perform a verification of sorts, but it turns out to be much easier said than done. The AMS Glossary of Meteorology recently adopted a definition for atmospheric rivers, a definition vetted by numerous subject matter experts and commented on during town hall meetings at major conferences. This definition is very broad, and while some generic numbers are provided via schematics, no exact criteria for identifying or tracking an AR are described. The upshot is that while there is broad consensus among the community regarding very general characteristics of ARs, there is no agreement regarding the variables, magnitudes, or precise geometries that should be used to identify them. So, while we could, for example, assess which methods most closely track satellite observed IWV features > 2000 km long and < 500 km wide (one definition, among many), we would already be self-selecting our preferred algorithm. This is true across a host of definitions. In light of this, we have chosen to focus our project on quantifying the uncertainty that arises as a result of these various algorithms and making this data available, at which point other researchers can feel free to pursue additional analyses.

Is it planned to give recommendations of methods to be used to identify ARs?

We will need to wait for the results, but we have some ideas about the form some of these recommendations might take. For example, certain algorithms might produce outliers in terms of AR climatology over certain regions. We may have reason to trust some algorithms more than others over certain regions, or to trust some algorithms more than others in anticipating the effects of a changing climate on ARs. We will need time to work out these details. What we do not want to do is rank algorithms. Recommendations will be based on identifying algorithm(s) that are most appropriate for specific science questions.

How will region specific methods (i.e. only applicable to the Western U.S.) implemented for global analyses?

All ARTMIP participants are running their algorithms for the North American west coast, the European west coast, and globally, or in whichever of those regions their algorithm is capable of running. We also have several algorithms being applied to Polar regions. We expect that algorithms developed specifically for a given region will be in closer agreement with each other than those developed for other regions and then applied there, but we don't yet have those results. To compare *all* algorithms (i.e. global with regional) we need to look at the common

denominator. For example, we have looked at landfalling ARs on the US West Coast because all catalogues include this area. However, this does not preclude sub-setting algorithms for other areas. Note that European landfalling ARs is a subset of the total number of algorithms in ARTMIP. Metrics applicable to global-only algorithms can also be analyzed in isolation. We also note that Tier 1 catalogues will be made available to the community after the completion of this phase and hope that members of the community will take advantage of the catalogues.

Additions to manuscript: We have added text to the manuscript to outline why and how we compare all algorithms (page 23 in the updated manuscript).

Technical Comments:

We have addressed all the technical comments. The point by point comments are below. *Figures 5a/b, 6a/b, and Table 1 have been modified as well as the addition of the Newman et al. reference to Table 1 footnotes.*

Why are captions placed above the Figures?
Agreed, we have moved captions to below the figure.

Fig.5a) and Fig.5b) (also Fig.6): Please label the color bar with units in both cases.
Caption Fig.5b): Clarify text: "Same as Fig.5a) ..."
We have incorporated these suggestions into the revised manuscript.

Please mention in the caption that the number of cases is different in Fig.5a) and Fig.5b) (also 6) due to the regional constraint of the respective definitions.
We have noted this explicitly in the figure captions.

p.23, l.23: Why do not all algorithms participate in the 1-month test case?
The reason all algorithms are not represented in the 1-month test is simply that we had a schedule, and not all participants' data was available at that time (we have also added new participants for Tier 1 and Tier 2 since the posting of this GMDD paper).

*Table 1: instead of using symbols (+,ˆ) for footnotes I suggest to use capital letters, which facilitates reading. Similarly, explanations of *ZN and AR_coeff: Which methods refer to these quantities and what is the meaning of AR_coeff? Why is it 0.3 and is this a general number?*

We have changed the symbols to letters. To avoid confusion with A1, A2, etc. and the notation for the algorithms, we will use lower case lettering “a”, “b”, and “c.”

AR_coeff refers to an empirically-derived coefficient used to define atmospheric rivers in the foundational paper of Zhu and Newell (1998). More work and a deeper exploration of this

value was also done by Newman et al (2012). Methods that use (or are based on) the Zhu and Newell method of AR identification are noted by footnotes and are: the Gorodetskaya et al. and Shields and Kiehl methods. The Zhu and Newell paper is already cited in the references, but we have added the Newman paper to the footnote in Table 1 for those interested in more detail.

Zhu, Y., and R. E. Newell (1998), A proposed algorithm for moisture fluxes from atmospheric rivers, Mon. Weather Rev., 126(3), 725–735, doi:10.1175/1520-0493(1998)126<0725:APAFMF>2.0.CO;2.

Newman, M., G. N. Kiladis, K. M. Weickmann, F. M. Ralph, and P. D. Sardeshmukh (2012), Relative contributions of synoptic and low-frequency eddies to time-mean atmospheric moisture transport, including the role of atmospheric rivers, J. Clim., 25, 7341–7361.

Response to GMDD Interactive comment from Referee #2

We would like to thank the reviewer for these constructive comments. We will address each point below:

Concerning the "Threshold Requirements" in Table 1, it would be very important to know whether the relative thresholds (normally percentiles) have been calculated separately for each month or season, for the winter half-year or for the entire year and I would suggest to include this information at this point.

This is a great suggestion. Although, it is important to note that not all "relative" methods are the same in terms of relying on climatology. For the methods that do, footnotes "c" in Table 1 explain what each percentile-method used for climatology.

Additions to manuscript: We have expanded the footnotes section to be clearer regarding how climatologies are computed by the algorithms that use percentiles to define thresholds.

Section 3.2, page 17, lines: 9-13, "...but future climate research may be better served by relative methodologies, partly ecause of the model biases in the moisture and/or wind fields...": To circumvent the problem of using absolute thresholds in climate model output, you could calculate the percentile corresponding to a given absolute threshold (e.g. an IVT of 250 kg m□1 s□1) in observations/ reanalysis data and find the absolute value corresponding to this percentile in the historical run of the model. This absolute value would then also be used in the RCP run or this model.

Yes, this method of identifying which threshold to use for future runs would work well. Algorithmic choices are left up to the developers themselves, depending on the science

question of interest. Comparing and contrasting the different algorithm choices is a priority for ARTMIP.

Section 4.2.1: You could also consider to use the NOAA-CIRES 20th Century Reanalysis and/or ECMWF's ERA-20C to have AR presence-absence time series for the entire 20th century, but this is of course just a suggestion. Doing so, you could e.g. assess aspects of low frequency variability associated with the PDO or AMO.

We have discussed these datasets our recent workshop, where we will discuss the reanalysis Tier 2 catalogues. The 20th Century Reanalysis is being considered but the timeline for this decision has been pushed back. The high-resolution climate change and the CMIP5 catalogues will precede the reanalysis catalogues in Tier 2.

Additions to manuscript: We have added the 20th Century NOAA-CIRES for consideration, page 23 in the updated manuscript.

Section 5.1, page 24, lines 19-20: Here you state that "a moisture threshold of [...], as in the human control, is potentially too permissive". However, from my point of view, the human control should be always better than any automated method so it is the methods having a problem at this point, not the human eye. Since several persons observing the same IVT field could come to distinct conclusions on whether an AR is present or not, the rough AR definition you cite on page 6 (the AMS one) should be still improved to come to a better consensus on what an AR actually is. Anyway, from my point of view, it is wrong to claim that the human eye is worse than the automated methods.

This is certainly a subject for debate within the community as well as ARTMIP. We do not want to imply that automated methods are better than human controls, or vice versa, and did not intend to give this message. We have amended the language to be more inclusive.

Additions to the manuscript: We have added additional language (in the metrics section) to be more neutral concerning human controls vs automated methods. We chose to add it here, rather than the results section, because we feel this statement should be made when human controls are first discussed. See page 24 in the updated manuscript.

Section 5.4 on page 27 and Figure 8: I would recommend to use an independent "purely observational" precipitation dataset other than MERRA.

This is a good recommendation. We intend to use several different datasets for Tier 1 and Tier 2 analysis. For this paper, and the proof-of-concept analysis, we used the MERRA-2 as a first look, only given this is only one month of data. The intent for this analysis is to show that our design

functions properly and to display the types of metrics we will delve into more deeply for full ARTMIP catalogues.

Figure 4: I would suggest to use a discrete instead of continuous colorbar for this figure.

We have adjusted the color bar to apply more discrete colors as suggested.

Figures 5 to 6: Adding "IVT" below or next-to the colorbar would be helpful in these figures.

Thank you for catching this omission. We have corrected this, as per the request of another referee.

Caption of Figure 6a: A space is missing after "kg" in the parenthesis.

Thank you for catching this syntax problem. This has been corrected.

Atmospheric River Tracking Method Intercomparison Project (ARTMIP): Project Goals and Experimental Design

Christine A. Shields[1], Jonathan J. Rutz[2], Lai-Yung Leung[3], F. Martin Ralph[4], Michael Wehner[5], Brian Kawzenuk[4], Juan M. Lora[6], Elizabeth McClenny[7], Tashiana Osborne[4], Ashley E. Payne[8], Paul Ullrich[7], Alexander Gershunov[4], Naomi Goldenson[9], Bin Guan[10], Yun Qian[3], Alexandre M. Ramos[11], Chandan Sarangi[3], Scott Sellars[4], Irina Gorodetskaya[12], Karthik Kashinath[13], Vitaliy Kurlin[14], Kelly Mahoney[15], Grzegorz Muszynski[13,14],Roger Pierce[16], Aneesh C. Subramanian[4], Ricardo Tome[11], Duane Waliser[17], Daniel Walton[18], Gary Wick[15], Anna Wilson[4], David Lavers[19], Prabhat[5], Allison Collow[20], Harinarayan Krishnan[5], Gudrun Magnusdottir[21], Phu Nguyen[22]

[1] Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder, CO, 80302, USA
[2]Science and Technology Infusion Division, National Weather Service Western Region Headquarters, National Oceanic and Atmospheric Administration, Salt Lake City, Utah, 84138, USA
[3] Earth Systems Analysis and Modeling, Pacific Northwest National Laboratory, Richland, Washington, 99354, USA
[4] Center for Western Weather and Water Extremes, Scripps Institution of Oceanography, La Jolla, California, 92037, USA
[5] Computational Chemistry, Materials, and Climate Group, Lawrence Berkeley National Laboratory, Berkeley, California, 94720, USA
[6] Department of Earth, Planetary, and Space Sciences, University of California, Los Angeles, California, 90095, USA
[7]Department of Land, Air and Water Resources, University of California, Davis, California, 95616, USA
[8]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, Michigan, 48109, USA
[9]Department of Atmospheric Sciences, University of Washington, Seattle, Washington, 98195, USA
[10]Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, California, 90095, USA
[11]Instituto Dom Luiz, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal
[12]Centre for Environmental and Marine Studies, University of Aveiro, 3810-193 Aveiro, Portugal
[13]Data & Analytics Services, National Energy Research Scientific Computing Center (NERSC), Lawrence Berkeley National Laboratory, Berkeley, California, 94720, USA
[14]Department Computer Science Liverpool, Liverpool, L69 3BX, UK
[15]Physical Sciences Division, Earth System Research Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO, 80305, USA
[16]National Weather Service Forecast Office, National Oceanic and Atmospheric Administration, San Diego, CA, 92127, USA
[17]Earth Science and Technology Directorate, Jet Propulsion Laboratory, Pasadena, California, 91109, USA
[18]Institute of the Environment and Sustainability, University of California, Los Angeles, California, 90095, USA
[19]European Centre for Medium-Range Weather Forecasts, Reading, RG2 9AX, UK

[20]Universities Space Research Association, Columbia, MD, 21046, USA
[21]Department of Earth System Science, University of California Irvine, CA 92697, USA
[22]Department of Civil & Environmental Engineering, University of California Irvine, CA 92697, USA

Correspondence to: Christine A. Shields (shields@ucar.edu)

Abstract

The Atmospheric River Tracking Method Intercomparison Project (ARTMIP) is an international collaborative effort to understand and quantify the uncertainties in atmospheric river (AR) science based on detection algorithm alone. Currently, there are many AR identification and tracking algorithms in the literature with a wide range of techniques and conclusions. ARTMIP strives to provide the community with information on different methodologies and provide guidance on the most appropriate algorithm for a given science question or region of interest. All ARTMIP participants will implement their detection algorithms on a specified common dataset for a defined period of time. The project is divided into two phases: Tier 1 will utilize the MERRA-2 reanalysis from January 1980 to June of 2017 and will be used as a baseline for all subsequent comparisons. Participation in Tier 1 is required. Tier 2 will be optional and include sensitivity studies designed around specific science questions, such as reanalysis uncertainty and climate change. High resolution reanalysis and/or model output will be used wherever possible. Proposed metrics include AR frequency, duration, intensity, and precipitation attributable to ARs. Here we present the ARTMIP experimental design, timeline, project requirements, and a brief description of the variety of methodologies in the current literature. We also present results from our 1-month "proof of concept" trial run designed to illustrate the utility and feasibility of the ARTMIP project.

1 Introduction

Atmospheric rivers (ARs) are dynamically driven, filamentary structures that account for ~90% of poleward water vapor transport outside of the tropics, despite occupying only ~10% of the available longitude (Zhu and Newell 1998). ARs are often associated with extreme winter storms and heavy precipitation along the west coasts of mid-latitude continents, including the western US, western Europe, and Chile (e.g., Ralph et al., 2004; Neiman et al., 2008; Viale and Nunez, 2011; Lavers and Villarini, 2015, Waliser and Guan, 2107). Their influence stretches as far as the polar caps as ARs transfer large amounts of heat and moisture poleward influencing the ice sheets surface mass and energy budget (Gorodetskaya et al., 2014; Neff et al., 2014; Bonne et al., 2015). Despite their short-term hazards (e.g., landslides, flooding), ARs provide long-term benefits to regions such as California, where they contribute substantially to mountain snowpack (e.g. Guan et al. 2010), and ultimately refill reservoirs. The sequence of precipitating storms that often accompany ARs may also contribute to relieving droughts (Dettinger 2014). Because ARs play such an important role in the global hydrological cycle (Paltan et al 2017) as well as to water resources in areas such as the western US, understanding how they may vary from subseasonal to interannual time scales and change in a warmer climate is critical to advancing understanding and prediction of regional precipitation (Gershunov et al., 2017).

The study of ARs has blossomed from 10 publications in its first 10 years in the 1990s to over 200 papers in 2015 alone (Ralph et al., 2017). This growth in scientific interest is founded on the vital role ARs play in the water budget, precipitation distribution, extreme events, flooding, drought, and many other areas with significant societal relevance, and is evidenced by current (past) campaigns including the multi-agency supported CalWater

(Precipitation, Aerosols, and Pacific Atmospheric Rivers Experiment) and ACAPEX (ARM Cloud Aerosol Precipitation Experiment) field campaigns in 2015 with deployment of a wide range of in-situ and remote sensing instruments from four research aircraft, a research vessel, and multiple ground-based observational networks (Ralph et al., 2016; Neiman et al., 2017). The scientific community involved in AR research has expanded greatly, with 100+ participants from 5 continents attending the First International Atmospheric Rivers Conference in August 2016 (http://cw3e.ucsd.edu/ARconf2016/), many of whom enthusiastically expressed interest in the AR definition and detection comparison project described here.

The increased study of ARs has led to the development of many novel and objective AR identification methods for model and reanalysis data that build on the initial model-based method of Zhu and Newell (1998) and observationally-based methods of Ralph et al. (2004) and Ralph et al. (2013). These different methods have strengths and weaknesses, affecting the resultant AR climatologies and the attribution of high-impact weather and climate events to ARs. Their differences are of particular interest to researchers using reanalysis products to understand the initiation and evolution of ARs and their moisture sources (e.g., Dacre et al., 2015, Ramos et al., 2016a; Ryoo et al., 2015, Payne and Magnusdottir, 2016), to assess weather and subseasonal-to-seasonal (S2S) forecast skill of ARs and AR-induced precipitation (Jankov et al., 2009; Kim et al., 2013; Wick et al., 2013a; Lavers et al., 2014; Nayak et al., 2014 ; DeFlorio et al., 2018; Baggett et al. 2017,), evaluate global weather and climate model simulation fidelity of ARs (Guan and Waliser, 2017), investigate how a warmer or different climate is expected to change AR frequency, duration, and intensity (e.g., Lavers et al., 2013; Gao et al., 2015; Payne and Magnusdottir, 2015; Warner et al., 2015; Shields and Kiehl, 2016 a/b; Ramos et al., 2016b; Lora et al. 2017; Warner and

Deleted: 7a, DeFlorio et al., 2017b

Mass, 2017), and attribute and quantify aspects of freshwater variability to ARs (Ralph et al., 2006; Guan et al., 2010; Neiman et al., 2011; Paltan et al., 2017).

Representing the climatological statistics of ARs is highly dependent on the identification method used (e.g., Huning et al., 2017). For example, different detection algorithms may produce different frequency statistics, not only between observation-based reanalysis products, but also among future climate model projections. The diversity of information on how ARs may change in the future will not be meaningful if we cannot understand how and why, mechanistically, the range of detection algorithms produce significantly different results. The variety of parameter variable types, and different choices that can be made for each variable in AR detection schemes is summarized in Fig. 1 and will be described in more detail in Section 3.

The detection algorithm diversity problem is not unique to ARs. For instance, the CLIVAR (Climate and Ocean – Variability, Predictability, and Change) program's IMILAST (Intercomparison of Mid Latitude Storm Diagnostics) project investigated extratropical cyclones similar to what is proposed here, (Neu et al., 2013). That project found considerable differences across definitions and methodologies and helped define future research directions regarding extratropical cyclones for such storms. Hence, it is imperative to facilitate an objective comparison of AR identification methods, develop guidelines that match science questions with the most appropriate algorithms, and evaluate their performance relative to both observations and climate model data so that the community can direct their future work.

The American Meteorological Society (2017) glossary defines an atmospheric river 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. The water vapor in atmospheric rivers is supplied by tropical and/or extratropical moisture sources. Atmospheric rivers frequently lead to heavy precipitation where they are forced upward—for example, by mountains or by ascent in the warm conveyor belt. Horizontal water vapor transport in the midlatitudes occurs primarily in atmospheric rivers and is focused in the lower troposphere".

ARTMIP strives to evaluate each of the participating algorithms within the context of this AR definition.

2 ARTMIP Goals

Numerous methods are used to detect ARs on gridded model or reanalysis data; therefore, AR characteristics, such as frequency, duration, and intensity, can vary substantially due to the chosen method. The differences between AR identification methods must be quantified and understood to more fully understand present and future AR processes, climatology, and impacts. With this in mind, ARTMIP has the following goals:

Goal #1*: Provide a framework that allows for a systematic comparison of how different AR identification methods affect the climatological, hydrological, and extreme impacts attributed to ARs.*

The co-chairs and committee have established this framework by facilitating meetings, inviting participants, sharing resources for data and information management, and providing a common structure enabling researchers to participate. The experimental design, described in Section 4, is the product of the first ARTMIP workshop, and provides the framework necessary for ARTMIP to succeed. The final design was a collaborative

decision and included participation from researchers from around the world interested in a AR detection comparison project and who are co-authors on this paper.

Goal #2: *Understand and quantify the differences and uncertainties in the climatological characteristics of ARs, as a result of different AR identification methods.*

The second goal is to quantify the extent to which different AR identification criteria (e.g., feature geometry, intensity, temporal, and regional requirements) contribute to the diversity, and resulting uncertainty, in AR statistics, and evaluate the implications to understanding the thermodynamic and dynamical processes associated with ARs, as well as their societal impacts.

The climatological characteristics of ARs, such as AR frequency, duration, intensity, and seasonality (annual cycle), are all strongly dependent on the method used to identify ARs. It is, however, the precipitation attributable to ARs that is perhaps most strongly affected, and this has significant implications for our understanding of how ARs contribute to regional hydroclimate now and in the future. For example, Fig. 2 highlights the results of three separate studies, (Dettinger et al., 2011, Rutz et al., 2014, Guan and Waliser, 2015), which used different AR identification methods to analyze the fraction of total cool-season or annual precipitation attributable to ARs from a variety of reanalysis and precipitation datasets. Differences in AR identification methods as well as the techniques used to attribute precipitation to ARs have important implications for understanding the hydroclimate and managing water resources across the western US. For example, because so much of the western US water supply is accumulated and stored as snowpack during the cool season, scientists and resource managers need to know how much of this water is attributable to ARs, and how changing AR behaviour might affect those numbers in the

future. The purpose of this figure is not to directly compare these analyses, but to motivate ARTMIP and illustrate the different ways of identifying and attributing precipitation that already exist in the literature. These results highlight the importance not only of quantifying the current uncertainty in AR climatology, but also the importance of future projections and reliable estimates of their uncertainty.

Goal #3: *Better understand changes in future ARs and AR-related impacts.*

As a key pathway of moisture transport across the subtropical boundary and from ocean to land, ARs are important elements of the global and regional water cycle. ARs also represent a key aspect of the weather–climate nexus as global warming may influence the synoptic-scale weather systems in which ARs are embedded and affect extreme precipitation in multiple ways. Hence, understanding the processes associated with AR formation, maintenance, and decay, and accurately representing these processes in climate models, is critical for the scientific community to develop a more robust understanding of AR changes in the future climate. A key question that will be addressed is how different AR detection methods may lead to uncertainty in understanding the thermodynamic and dynamical mechanisms of AR changes in a warmer climate. Although the water vapor content in the atmosphere scales with warming following the Clausius-Clapeyron relation, changes in atmospheric circulation such as the jet stream and Rossby wave activity may also have a significant impact on ARs in the future (Barnes et al., 2013, Lavers et al., 2015, Shields and Kiehl, 2016b). Will ARs from different ocean basins respond differently to greenhouse forcing? How do natural modes of climate variability come into play, i.e., El Nino–Southern Oscillation (ENSO), Arctic Oscillation (AO), Madden-Julien Oscillation (MJO), the Pacific Decadal Oscillation (PDO), or the Southern Annular Model (SAM)? How do changes in precipitation efficiency influence regional precipitation associated with ARs in

the future? As the simulation fidelity of ARs is somewhat sensitive to model resolution (Hagos et al., 2015, Guan and Waliser, 2017), another important question is whether certain AR detection and tracking methods may be more sensitive to the resolutions of the simulations than others, and what are the implications to understanding uncertainty in projections of AR changes in the future.

To begin to answer and diagnose these questions, an understanding of how the definition and detection of an AR alters the answers to these questions is needed. A catalogue of ARs and AR-related information will enable researchers to assess which identification methods are most appropriate for the science question being asked, or region of interest. Applying different identification methods to climate simulations of ARs in the present day and future climate will facilitate more robust evaluation of model skill in simulating ARs and identification of mechanisms responsible for changes in ARs and associated extreme precipitation in a warmer climate. Finally, determination of the most appropriate methods of identifying ARs will provide for a set of best practices and community standards that researchers engaged in understanding ARs and climate change can work with and use to develop diagnostic and evaluation metrics for weather and climate models.

3 Method types

Table 1 summarizes the different algorithms adopted by the ARTMIP participants. Details for each parameter type and choice (from Fig. 1) are listed as table columns. The developer of the method is listed by row and refers to individuals or groups who developed the algorithm. The identifier in the first column (A1, A2, etc) will be used for Figs. 3, 5, 7, 8 and denotes those algorithms participating in the initial "proof of concept" phase of the project. Type choices are Condition or Track (see Section 3.1 for definition of these choices). Geometry

requirements refer to the shape and axis requirements, if any, of an AR object. For example, a "Condition" AR algorithm that tests a grid point may also have a requirement that strings grid points together to meet a minimum length, width, or axis. Threshold requirements refer to any absolute or relative threshold, typically for a moisture-related variable, that must be met for an AR object to be defined. Temporal requirements refer to any time conditions to be met. Tracking algorithms typically contain temporal requirements to define an AR object as it is defined in time and space. However, many condition algorithms may also specify a minimum number of time instances (non-varying over a grid point) to be met before an AR object is counted for that grid point. Region refers to whether or not the algorithm is defined to track or count ARs globally or only over specified regions. Reference lists published papers and datasets and their DOI numbers. "Experimental" algorithms have not been published yet.

Table 1. Algorithm methods participating in the early phases of ARTMIP and content of this paper. Developer is listed along with algorithm details, i.e., type; geometry, threshold, and temporal requirements; region of study; DOI reference. Identifiers for the subset of methods participating in the one month "proof of concept" test are in the far-left column and labeled as A1, A2, etc. ARTMIP is an on-going project with the addition of new participants as the project progresses. For the most recent list of developers and participants, please refer to the ARMTIP webpages at http://www.cgd.ucar.edu/projects/artmip/.

Deleted: ¶ ¶ ¶ ¶

Deleted: participants

	Developer	Type	Geometry Requirements	Threshold Requirements	Temporal Requirements	Region	DOI/ Reference
A1	Gershunov et al.[b]	Condition and Track	>= 1500km long	Absolute: 250kgm^{-1}s^{-1} IVT 1.5cm IWV	Time stitching -18 hours (3 time steps for 6 hourly data)	Western U.S.	10.1002/2017GL074175
A2	Goldenson[b]	Condition	>2000km long and < 1000km wide, Object recognition	Absolute: 2cm IWV	Time slice	Western U.S.	Goldenson et. al, submitted
	Gorodetskaya et al.	Condition	IWV > thresh. at the coast (within defined longitudinal sector) and continuously at all latitudes for ≥20° equatorward (length > 2000 km), within ±15° longitude sector (width of 30° ~1000 km at 70°S; requirement of meridional extent)	Relative: [a]ZN using IWV adjusted for reduced tropospheric moisture holding capacity at low temperatures (AR_{coeff} = 0.2)	Time slice	Polar (East Antarctica)	10.1002/2014GL060881
A3	Guan and Waliser[bc]	Condition	Length >2000km and length-width ratio>2; Coherent IVT	Relative: 85th percentile IVT; Absolute min requirement designed for	Time slice	Global	10.1002/2015JD024257; Guan et al., 2017,

Deleted: +

Deleted: +

Deleted: In review

Formatted: Superscript

Deleted: *

Deleted: +

Deleted: ^

			direction within 45^0 of AR shape orientation and with a poleward component	polar locations: $100 kgm^{-1}s^{-1}$ IVT			JHM, submitted
A4	Hagos et al. [b]	Condition	Dependent on threshold requirements to determine footprint; > 2000 km long and < 1000 km wide	Absolute: 2cm IWV $10 ms^{-1}$ wind speed	Time slice	Western U.S.	10.1002/2015GL065435; 10.1175/JCLI-D-16-0088.1
	Lavers et al.	Condition	4.5^o latitude movement allowed	Relative: ~85^{th} percentile determined by evaluation of reanalysis products	Time slice	UK, Western US	10.1029/2012JD018027
A5	Leung and Qian[b]	Track	Moisture flux has an eastward or northward component at landfall; tracks originating north of 25N and east of 140W are rejected	Absolute: mean IVT along track > $500 kgm^{-1}s^{-1}$ and IVT at landfall > $200 kgm^{-1}s^{-1}$; grid points up to 500km to the north and south along the AR tracks are included as part of the AR if their mean IVT > $300 kgm^{-1}s^{-1}$	Time slice	Western U.S.	10.1029/2008GL036445

Deleted: +

Deleted: +

A6, A7	Lora et al.[b]	Condition	Length >= 2000km	Relative: IVT 100kgm^{-1}s^{-1} above climatological area means for N. Pacific	Time slice	Global (A6), North Pacific (A7)	10.1002/2016GL071541
	Mahoney et al.	Condition and Track	Length >= 1500km, Width <=1500km	Absolute: ARDT-IVT 500kgm^{-1}s^{-1} for SEUS.	See Wick	Southeast U.S.	10.1175/MWR-D-15-0279.1 (uses Wick)
	Muszynski et al.	Condition	Topological analysis and machine learned	Threshold-free	N/A	Western U.S., adaptable to other regions	Experimental
A8	Payne and Magnusdottir [bc]	Condition	Length > 1200km, landfalling only	Relative: 85th Percentile of maximum IVT (1000-500mb) Absolute: IWV >2cm, 850mb wind speed > 10m/s	Time stitching (12-hour minimum) -	Western U.S.	10.1002/2015JD023586; 10.1002/2016JD025549
	Ralph et al.	Condition	Length >= 2000km, Width <= 1000km	Absolute: IWV 2cm	Time slice	Western U.S.	10.1175/1520-0493(2004)132<1721:sacaoo>2.0.co;2
A9	Ramos et al.[bc]	Condition	Detected for reference meridians, length >=1500km, latitudinal movement <4.5^{0}N	Relative: IVT 85th percentile (1000-300mb)	Time slice, but 18-hour minimum for persistent ARs	Western Europe, South Africa, adaptable to other regions	10.5194/esd-7-371-2016

A10	Rutz et al.[b]	Condition	Length >= 2000km	Absolute: IVT (surface to 100mb) = 250kgm^{-1}s^{-1}	Time slice	Global, low value on tropics	10.1175/MWR-D-13-00168.1
A11, A12, A13	Sellars et al.[b]	Track	Object identification	Absolute: IVT, thresholds tested = 300 (A11), 500 (A12), 700 (A13) kgm^{-1}s^{-1}	Time stitching, minimum 24-hour period	Global	10.1002/2013EO320001; 10.1175/JHM-D-14-0101.1
A14	Shields and Kiehl[b]	Condition	Ratio 2:1, length to width grid points min 200km length; 850mb wind direction from specified regional quadrants, landfalling only	Relative: [a]ZN moisture threshold using IWV; Wind threshold defined by regional 85th percentile 850mb wind magnitudes	Time slice	Western U.S. Iberian Peninsula, UK, adaptable but regional specific	10.1002/2016GL069476; 10.1002/2016GL070470
A15	TEMPEST[b]	Track	Laplacian IVT thresholds most effective for widths >1000km; cluster size minimum = 120000km^2	IVT >=250kgm^{-1}s^{-1};	Time stitching	Global, but latitude >=15^o	Experimental
	Walton et al.	Condition and Track	Length >= 2000 km	Relative: IVT > 250 kg/m/s + daily IVT climatology	Time stitching, minimum 24-hour period	Western U.S.	Experimental

Deleted: +

Deleted: +

Deleted: +

Formatted: Superscript

Deleted: *

Deleted: +

	Wick et al.	Condition and Track	>=2000km long, <= 1000km wide object identification involving shape and axis	Absolute: ARDT-IWV >2cm	Time slice and stitching	Regional	10.1109/TGRS.2012.2211024

[a]ZN relative threshold formula: $Q >= Q_{zonal\_mean} + AR_{coeff} (Q_{zonalmax} - Q_{zonamean})$ where Q = moisture variable, either IVT (kg $m^{-1}s^{-1}$) or IWV(cm). AR_{coeff} = 0.3 except where noted. (Zhu and Newell, 1998). The Gorodetskaya method uses $Q_{sat, where}$ Q_{sat} represents maximum moisture holding capacity calculated based on temperature (Clausius-Clapeyron), an important distinction for polar ARs. Additional analysis on the ZN method can be found in Newman et al., 2012

[b]Methods used in a 1-month proof-of-concept test (Section 5). These methods are assigned an algorithm id, i.e. A1, A2, etc.

[c] These 1-month proof-of-concept methods apply a percentile approach to determining ARs. A3 and A8 applied the full MERRA2 climatology to compute percentiles. A9, applied the Feb 2017 climatology for this test only. For the full catalogues, A9 will apply extended winter and extended summer climatologies to compute percentiles. Please refer to individual publications (DOI reference column in this table) for climatologies used in earlier published studies by each developer. The climatology used to compute percentile is often dependent on the dataset (re-analysis or model data) being used.

Deleted: *

Formatted: Superscript

Deleted: *

Formatted: Font: Not Bold, Superscript

Deleted: +

Deleted: ^

3.1 Condition vs tracking algorithms

The subtleties in language when describing different algorithmic approaches are best illustrated with the "tracking" versus "condition" parameter type. For ARTMIP purposes, two basic detection "types," defined at the first ARTMIP workshop, represent two fundamentally different ways of detecting ARs. "Condition" refers to counting algorithms that identify a time instance where AR conditions are met. Condition algorithms typically search grid point by grid point for each unique time instance. If AR geometry (involving multiple grid points) and threshold requirements are met, then an AR "condition" is found at that grid point and that point in time. Condition methods may also have an added temporal requirement, but this does not impact the fact that conditions are met at a unique point in space (grid point).

"Tracking" refers to a Lagrangian-style detection method where ARs are objects that can be tracked (followed) in time and space. Objects have specified geometric constraints and can span across grid points. Tracking algorithms must include a temporal requirement that stitches time instances together, i.e., a tracked AR would include several 3-hour time slices stitched together. An example of an object-oriented tracking methods is the Sellars et al., 2015 tracking method.

3.2 Thresholding: absolute versus relative approaches

Another major area where algorithms diverge is in how to determine the intensity of an AR. Some methods follow studies, such as Ralph et al. (2004) and Rutz et al. (2014), that assign an observationally-derived value, such as 2 cm of IWV, or an IVT value of 250 kg $m^{-1}s^{-1}$ to determine the physical threshold required for identification of an AR. Other methods use a statistical approach rather than an absolute value when setting a threshold value, such as the

approach developed by Lavers et al. (2012) where an AR is defined by the 85th percentile values of IVT (kg m^{-1}s^{-1}). Other relative threshold methods, such as Shields and Kiehl (2016a/b), and Gorodetskaya et al. (2014), apply a direct interpretation of the foundational Zhu and Newell (1998) paper that defines ARs by computing anomalies of IWV (cm) or IVT (kg m^{-1}s^{-1}), by latitude band. Further, Gorodetskaya et al. (2014) used the physical approach to define a threshold for IWV depending on the tropospheric moisture holding capacity as a function of temperature at each pressure level (Clausius-Clapeyron relation). The Lora et al. (2017) method is yet another relative thresholding technique wherein ARs are detected for IVT at 100 kg m^{-1}s^{-1} above a climatological-derived mean IVT value and thus changes with the climate state. Although all of these methods "detect" ARs, they do not always detect the same object. Observationally based methods may be best for case studies, forecasts, or current climatologies, but future climate research may be better served by relative methodologies, partly because of model biases in the moisture and/or wind fields. Ultimately, however, the best algorithmic choice will be unique to the science being done, rather than depending on general categories.

4 Experimental Design

ARTMIP will be conducted using a phased experimental approach. All participants must contribute to the first phase to provide a baseline for all subsequent experiments in the second phase. The first phase will be called Tier 1 and will require that participants provide a catalogue of AR occurrences for a set period of time using a common reanalysis product. This phase will focus on defining the uncertainties amongst the various detection method algorithms. The second phase, Tier 2, is optional, and will potentially include creating catalogues for a number of common datasets with different science goals in mind. To some degree, the experiments chosen for Tier 2 will be informed by the outcomes of Tier 1;

however, initially, ARTMIP participants have proposed three separate Tier 2 experiments. The first and second experiments will test AR algorithms under climate change scenarios and different model resolutions, and the third experiment will explore the uncertainties to the various reanalysis products. Table 2 outlines the timeline for ARTMIP.

Table 2. ARTMIP Timeline. Completed targets are in bold.

Target Date	Activity
May 2017	**1st ARTMIP Workshop**
August/September 2017	**1-Month Proof of Concept Test**
January-April 2018	**Full Tier 1 Catalogues Completed**
April 2018	**2nd ARTMIP Workshop**
Spring/Summer/Fall 2018	Tier 1 Analysis and Scientific Papers
Fall 2018, on-going	Tier 2 Climate Change Catalogues Due, Analysis, Papers
Summer 2019, on-going	Tier 2 CMIP5 Catalogues Due, Analysis, Papers
Winter 2019/2020, on-going	Tier 2 Reanalysis Catalogues, Analysis, Papers

Deleted: , and second and third set of experiments will be testing AR algorithms under climate change scenarios and different model resolutions

Formatted Table

Deleted: January

Formatted: Font: Bold

Deleted: Due

Formatted: Font: Bold

Formatted: Superscript

Deleted: Winter 2017

Deleted: /

Deleted: S

Deleted: pring 2018

Deleted: Catalogues Due

Deleted: Spring 2018 ... [1]

4.1 Tier 1 description

ARTMIP participants will run their independent algorithms on a common reanalysis dataset and adhere to a common data format. Tier 1will establish baseline detection statistics for all participants by applying the algorithms to MERRA-2 (Modern Era Retrospective-Analysis for Research and Applications, (Gelaro et al., 2017, Data DOI number: 10.5067/QBZ6MG944HW0) reanalysis data, for the period of January 1980 – June 2017. To eliminate any processing differences between algorithm groups, all moisture and wind

variables have been processed and made available at the University of California, San Diego (UCSD) Center for Western Weather and Water Extremes (CW3E) (B. Kawzenuk, personal communication) at ~50km (.5º x .625º) spatial resolution and 3-hourly instantaneous temporal resolution. Specifically, ARTMIP participants that require IVT (integrated vapor transport, kg $m^{-1}s^{-1}$) information for their algorithms will be using IVT data calculated by UCSD using the MERRA-2 data 3-hourly zonal and meridional winds, and specific humidity variables. IVT is calculated using the following Eq. (1), (from Cordeira et al., 2013),

$$\text{IVT} = -\frac{1}{g}\int_{Pb}^{Pt}(q(p)\,\boldsymbol{V}_h(p))\,dp \qquad (1)$$

where q is the specific humidity (kg/kg), $\mathbf{V}_h$ is the horizontal wind vector (ms^{-1}), Pb is 1000 hPa, Pt is 200 hPa, and g is the acceleration due to gravity. The 1-hourly averaged IVT data available from MERRA-2 directly will not be used. A comparison between 3-hourly UCSD IVT-computed data and 1-hourly MERRA-2 data was completed with details found in supplemental information. Although the 1-hour data provides better temporal resolution, the 3-hourly provides ample temporal information and is sufficient for algorithmic detection comparisons for ARTMIP. Gains using the 1-hourly MERRA-2 IVT data do not outweigh the extra burden in computational resources required for groups to participate in ARTMIP.

Not all algorithms require IVT. Instead, some use IWV, integrated water vapor, or precipitable water (cm). This quantity is derived from MERRA-2 data and is computed as Eq. (2)

$$\text{IWV} = -\frac{1}{g}\int_{Pb}^{Pt} q\,(p)\,dp \qquad (2)$$

where q is the specific humidity (kg/kg) , Pb is 1000 hPa, Pt is 200 hPa, and g is the acceleration due to gravity. Table 3 summarizes all the MERRA-2 data available for AR tracking.

Table 3. ARTMIP variables available for detection algorithms.

Variable	Variable Units	Description	Level
U	ms^{-1}	Zonal wind	All pressure levels
V	ms^{-1}	Meridional wind	All pressure levels
Q	kg/kg	Specific humidity	All pressure levels
T	Kelvin	Air Temperature	All pressure levels
IVT	kg $m^{-1}s^{-1}$	Integrated vapor transport	Integrated from 1000 to 200 hPa
IWV	mm	Integrated water vapor	Integrated from 1000 to 200 hPa
uIVT	kg $m^{-1}s^{-1}$	Zonal wind component of IVT	Available as integrated or pressure level
vIVT	kg $m^{-1}s^{-1}$	Meridional wind component of IVT	Available as integrated or pressure level

Once catalogues are created for each algorithm, data will be made available to all participants. Data format specifications for each catalogue are found in supplementary material.

Many of the ARTMIP participants focus on the North Pacific (Western North America) and North Atlantic (European) regions, however, ARs in other regions, such as the poles and the Southeast U.S. may also be evaluated with ARTMIP data. We are not placing any coverage requirements for participation in ARTMIP, and each group can provide as many global or regional catalogues as desired.

4.2 Tier 2 description

Tier 2 will be similar in structure to Tier 1 in that all participants will create catalogues on a common dataset and follow the same formats, etc. However, instead of algorithms creating catalogues for one reanalysis product, a number of sensitivities studies will be conducted spanning AR detection sensitivity to reanalysis products, and AR detection sensitivity under climate change scenarios.

4.2.1 High-resolution climate change catalogues

For climate model resolution studies, CAM5 (Community Atmosphere Model, Version 5; Neale et al., 2010) 20th century simulations available at 25, 100, and 200 km resolutions from the C20C+ (Climate of the 20th Century Plus Project) Sub-project on Detection and Attribution (portal.nersc.gov/c20c) is available for participants to create AR catalogues for a period of 27 years (1979-2005). For climate change studies, high resolution (25 km) historical (1979-2005) and end of the century RCP8.5 (2080-2099) CAM5 simulation data

Moved down [10]: 4.2.1 Reanalysis catalogues¶

Deleted: ¶

Deleted: ¶
For the reanalysis sensitivity experiment, products chosen may include ERA-I or 5 (European Reanalysis-Interim, or Version 5; Dee et al., 2011), NCEP-NCAR (National Center for Environmental Prediction –National Center for Atmospheric Research; Kalnay et al., 1996), JRA55 (Japanese 55-year Renanalysis; Kobayashi et al., 2015), and CFSR (Climate Forecast System Reanalysis, Saha et al., 2014). Resolution will be coarsened to the lowest resolution and temporal frequency will be chosen by the lowest temporal frequency available amongst all the various products for the necessary variables, (listed in Table 3). ¶

Deleted: 2

Deleted: C

are also provided. This version of CAM5 uses the finite volume dynamical core on a latitude-longitude mesh (Wehner et al., 2014) with data freely available at portal.nersc.gov/C20C.

We use high resolution data for both the Tier 1 (~50km) and Tier 2 (25km) climate change catalogues because it has been shown that high resolution data is important in replicating AR climatology and regional precipitation. Although some climate models have a tendency to overestimate extreme precipitation related to ARs, these biases tend to decrease when high resolution is applied (Hagos et al., 2015, Hagos et al., 2016). In an Earth system modelling framework, regional precipitation is represented more realistically in the higher resolution version compared to the standard lower resolution horizontal grids, (Delworth et al., 2012, Small et al., 2014, Shields et al., 2016). High resolution data will have a better representation of topographical features and be better able to represent regional features at a finer scale.

4.2.2. CMIP5 catalogues

Deleted: 3

Deleted: 3

A number of studies have analyzed CMIP5 model outputs to explore future changes in ARs and the thermodynamic and dynamical mechanisms for the changes (e.g. Lavers et al., 2013; Payne and Magnusdottir, 2015; Warner et al., 2015; Gao et al., 2016; Shields and Kiehl, 2016b, Ramos et al., 2016b). However, there is a lack of systematic comparison of the results and how differences in AR detection and tracking may have influenced the conclusions regarding the changes in AR frequency, AR mean and extreme precipitation, spatial and seasonal distribution of landfalling ARs, and other AR characteristics, impacts, and mechanisms. Characterizing uncertainty in projected AR changes associated with detection algorithms will facilitate more in-depth analysis to understand other aspects of uncertainty

related to model differences, internal variability, and scenario differences, and such uncertainties influence our understanding of AR changes in a warming climate.

4.2.3 Reanalysis catalogues

Moved (insertion) [10]

Deleted: 1

For the reanalysis sensitivity experiment, products chosen may include ERA-I or 5 (European Reanalysis-Interim, or Version 5; Dee et al., 2011), NCEP-NCAR (National Center for Environmental Prediction –National Center for Atmospheric Research; Kalnay et al., 1996), JRA55 (Japanese 55-year Renanalysis; Kobayashi et al., 2015), and CFSR (Climate Forecast System Reanalysis, Saha et al., 2014), and the NOAA-CIRES 20th Century Reanalysis (Compo et al., 2011). Resolution will be coarsened to the lowest resolution and temporal frequency will be chosen by the lowest temporal frequency available amongst all the various products for the necessary variables, (listed in Table 3).

5 Metrics

Once all the catalogues are complete, then analysis will begin. There are many metrics to potentially analyze that are currently found in the literature. The frequency, duration, intensity, climatology of ARs and their relationship to precipitation are common. Other metrics, such as those described in Guan and Waliser (2017) can be adapted for ARTMIP. To test the experimental design, we conducted a 1-month "proof of concept" test to help the basic design and fine tune a few metrics. Here we present a few results from this one-month test that diagnose frequency, intensity and duration for two landfalling AR regions, the North Pacific and North Atlantic. For the full Tier 1 analysis in future publications, global views will be added. Landfalling regions are chosen so that both regional algorithms, focused on impacts to specific continental areas, and global algorithms can be compared directly. For

Deleted: one month

the full catalogues in Tier 1, additional regions will be analyzed, including the East Antarctic, which has proven to have large differences between methodologies that implement a global algorithm compared to a regionally specific polar algorithm (Gorodetskaya et al., 2014). February 2017 was chosen because of the frequent landfalling North Pacific ARs during this time. Algorithms participating in the 1-month test are labelled with a [b] in Table 1 and identified with an algorithm id, i.e, A1, A2), etc. We also conducted a "human" control, where AR conditions and tracks were identified by eye for the month of February for landfalling ARs impacting the western coastlines of North America and Europe. Full details on the human control dataset are explained in supplemental material. We emphasize here that the human control is not considered "truth", nor is it better or worse than automated methods, but merely another (subjective) method to add to the spectrum of detection algorithms participating in ARTMIP.

Deleted: +

Deleted: ¶
¶

5.1 Frequency

Fig. 3 shows frequency (in 3-hour instances) by latitude band for landfalling ARs. The human control as well as each of the methods are plotted for February 2017. Each color represents a unique detection algorithm, and the black lines represent the human controls where both IVT and IWV were utilized to identify ARs "by eye". The IVT threshold (solid black line) is 250 kg $m^{-1}s^{-1}$ and the IWV thresholds (two different dashed lines) are 2 cm and 1.5 cm, respectively. For western North America, all of the algorithms and the human controls agree on the shape of the latitudinal distribution with most AR 3-hour-period detections accumulating along the coast of California. ARs over the North Atlantic are latitudinally more diverse, but the majority of algorithms and controls peak around 53°N. Regarding the actual number of 3-hour periods, there is a large spread in the frequency values across all the automated algorithms with the human control "detections" far exceeding most algorithms.

This preliminary result suggests that setting a moisture threshold of 250 $kgm^{-1}s^{1}$ or an IWV value of 2cm for North Atlantic ARs, as in the human control, is potentially too permissive.

To help identify case study events, a methodology count of how many (and which) methods detect an AR along the coast can be conducted. Fig. 4 plots the number of methods that detect an AR at the North American coastline for a sample of days in February, 2017. The number of methods detections for each 3-hour time instance per day was computed, but only the maximum time instance per day is plotted for simplicity. The polygons represent the number of methods. For example, if only 1 method detects an AR at a specific grid point along the coast, then a beige circle is plotted at that grid point along the coast; if 14 methods detect an AR at a specific grid point along the coast, then a dark blue circle is plotted at that grid point along the coast, and so forth. Even with this basic representation, the diversity in numbers of method detections for each day is large. There are days where there is good method agreement in identifying AR conditions along the coastline. For example, February 7th, most methods identify AR conditions in Southern California, and on the 9th and 15th many methods detect ARs in the Pacific Northwest. However, there are many days where only a handful of methods detect ARs (i.e. February 22nd and 28th). The ability of individual algorithms to detect the duration of events listed here is examined in further detail in Section 5.3.

Deleted: . ¶

Deleted: o

Deleted: 3

5.2 Intensity

Intensity can be defined in many ways but often refers to the amount of moisture present in an AR and/or the strength of the winds. IVT is an obvious quantity to use when evaluating the strength of an AR because it incorporates both wind and moisture. There is value, however, at looking at these quantities separately when trying to decompose dynamic and

thermodynamic influences. For the 1-month test, we looked at IVT for time instances where ARs exist.

In Figs. 5 and 6 we show two different ways of looking at mean AR-IVT across applicable methods to highlight how the definition of intensity can also vary. Fig. 5a/b show composites (for the North Pacific and European sectors, respectively) only at grid points where detection algorithms are implemented and include all time instances. This provides a look at the mean IVT for all ARs at all locations for all times. Not all algorithms search for AR conditions at all points. For example, A14 (Shields and Kiehl) only detects ARs that make landfall along coastal grid points, and A9 (Ramos et al.) detects ARs along reference meridians (for masks for regional algorithms, see Fig. S3 in supplementary material). Fig. 6 comparatively, shows IVT composites for each grid point, focusing only on specific time periods where landfalling ARs exist. While Fig. 5 shows mean IVT for all ARs at detection points, Fig. 6 is the composite for landfalling ARs only. Each of these methods show intensity but are looking at different quantities. The landfalling ARs have a different signature and a less intense distribution, compared to the all-location AR composites. As one would expect, for both Figs. 5 and 6, methods with higher thresholds on IVT produce much higher AR average intensities, thus, AR intensity metrics could be thought of as self-selecting for some cases.

5.3 Duration

Duration of ARs also must be defined. Typically, this is expressed as the length of time an AR affects a point location, for example, a coastal location for a landfalling storm. However, for tracking algorithms, duration may be defined as the life-cycle of an AR. For the 1-month proof-of-concept test, we use the first definition and look at the duration at coastal locations along the North American west coast and specific European locations. The top panel of Fig. 7

shows a time series of daily IVT anomalies along the western coastlines of the (orange line) Iberian Peninsula, (teal line) United States and (blue line) Ireland and the United Kingdom. Four "human" observed AR tracks for events in each region are shaded and the composite magnitude of IVT for each are shown in panels a – d. These four events are compared over a variety of algorithms, indicated by algorithm id in the top panel, where each black dot indicates detection of an AR along the coastline. While all algorithms are listed, it is important to note that they are a mix of regional and global in scope. An example snapshot of IVT from a global view is shown in supplemental material (Figure S4). February 19th, 21Z, 2017 was chosen to illustrate individual ARs in the MERRA-2 dataset during the month examined here.

Deleted: ¶

The four selected events in Figure 7 demonstrate the large diversity of AR geometry, landfall location and intensity that must be identified by each algorithm. The agreement between the different algorithms, hinted at in Fig. 4, is apparent in a comparison of the two West Coast examples mentioned in section 5.1 (Figs. 7b and d). The three versions of the Sellars et al. (2015) algorithm can be used as a benchmark of AR intensity, in which the IVT threshold increases from 300 kg m^{-1} s^{-1} (in A11) to 700 kg m^{-1} s^{-1} (in A13). Relatively strong events are well captured by most algorithms (Fig. 7a - c), with few exceptions that are likely related to domain size. Agreement between algorithms on the duration or presence of an AR during weaker events is much more variable, such as that seen in Fig. 7d.

5.4 Comparison with precipitation observational datasets

The importance of understanding and tracking ARs ultimately boils down to impacts. AR-related precipitation can be the cause of major flooding, can fill local reservoirs, and can relieve droughts. How much precipitation falls, the rate at which it falls, and when and where

it falls, specifically during AR events, is a metric we must consider for this project. The variation among the different algorithms can be seen in a comparison of precipitation characteristics for the event shown in Fig. 7b using MERRA-2 precipitation data (Fig. 8) . The inset shows the landfalling mask from Shields and Kiehl (2016), which is used as a common base of comparison for landfall between the different algorithms. Precipitation related to the landfalling AR is isolated by focusing only on gridboxes that are tagged by each algorithm. Comparison shows a positive relationship between the average spatial coverage of the detected landfalling plume (y-axis) and the average maximum precipitation rate at each time slice (x-axis). Generally, the durations of AR conditions along the coastline are higher for algorithms with broader coverage. The wide range of characteristics for this single well-defined event motivate further investigation.

As a part of Tier 1, methods will be evaluated using a variety of precipitation products in addition to MERRA-2, most relevant to the areas of interest. These include the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) 3B42 product, version 7 (Huffman et al., 2007), the Global Precipitation Climatology Project (GPCP) dataset (Huffman et al., 2001), the Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks (PERSIANN; Sorooshian et al., 2000), Livneh (Livneh et al., 2013), and E-OBS (Haylock et al., 2008). Tier 2 climate studies will use precipitation output, both convective and large-scale, from the CAM5 simulations. Finally, it is important to consider not only the uncertainties in attributing precipitation due to detection method, but also the manner or technique used when assigning precipitation values to individual ARs.

6 Summary

ARTMIP is a community effort designed to diagnose the uncertainties surrounding atmospheric river science based on detection methodology alone. Understanding the uncertainties and, importantly, the implications of those uncertainties, is the primary motivation for ARTMIP, whose goals are to provide the community with a deeper understanding of AR tracking, mechanisms, and impacts for both the weather forecasting and climate community. There are many detection algorithms currently in the literature that are often fundamentally different. Some algorithms detect ARs based on a condition at a certain point in time and space, while others follow, or track, ARs as a whole object through space and time. Some algorithms use absolute thresholds to determine moisture intensity, while others use relative measures, such as statistical or anomaly-based approaches. The many degrees of freedom, in both detection parameter and choice of thresholds or geometry, add to the uncertainty on defining an AR, in particular for gridded datasets such as reanalysis products, or model output. This project aims to disentangle some of these problems by providing a framework to compare detection schemes. The project is divided into two tiers. The first tier is mandatory for all participants and will provide a baseline by applying all algorithms to a common dataset, the MERRA-2 reanalysis. The second tier is optional and will focus on sensitivity studies such as comparison amongst a variety of reanalysis products, and a comparison using climate model data, utilizing both historical and future climate simulations. Metrics diagnosed by ARTMIP will, at minimum, include AR frequency, intensity, duration, climatology, and relationship to precipitation. Participation is open to any group with an AR detection algorithm or an interest in evaluating ARTMIP data. Participants will have full access to all ARTMIP data.

7 Data Availability

Data for ARTMIP is described in section 4. All 1-month proof of concept catalogues used for the figures and preliminary results in this paper are included as supplemental material. Source data for the full MERRA-2 Tier 1 catalogues are available from the Climate Data Gateway (CDG), doi:10.5065/D62R3QFS. Full ARTMIP catalogues will be available to ARTMIP participants after the respective tier phases have been completed. Participation in ARTMIP is open to any person or group with an AR detection scheme and/or interest in analyzing data produced by ARTMIP. To do so, contact C. Shields (shields@ucar.edu) or J. Rutz (jonathan.rutz@noaa.gov).

8 Acknowledgments

The contributions from NCAR (Cooperative Agreement DE-FC02-97ER62402), PNNL, and LBNL to ARTMIP are supported by the U.S. Department of Energy Office of Science Biological and Environmental Research (BER) as part of the Regional and Global Climate Modeling program. PNNL is operated for DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830. LBNL is operated for DOE by the University of California under contract number DE-AC02-05CH11231. Computing resources (ark:/85065/d7wd3xhc) were partially provided by the Climate Simulation Laboratory at NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation and other agencies, as well as Scripps Institute for Oceanography at the University of California, San Diego. Alexandre M. Ramos was supported through a postdoctoral grant (SFRH/BPD/84328/2012) from the Portuguese Science Foundation (Fundação para a Ciência e a Tecnologia, FCT). We also thank CW3E (Center for Western Weather and Extremes) for providing support for the 1st ARTMIP Workshop, and the many people and their sponsoring institutions involved with the ARTMIP project.

Formatted: Pattern: Clear (White)

Deleted:

Deleted: applied to the 1-month proof of concept test presented in this paper is

Deleted: at the University of California, San Diego (UCSD) Center for Western Weather and Water Extremes (CW3E) from B. Kawzenuk

Formatted: Font: (Default) +Body (Times New Roman), 12 pt, Not Bold

Deleted: .

Formatted: Font: (Default) +Body (Times New Roman), Font color: Custom Color(RGB(34,34,34))

Formatted: Pattern: Clear

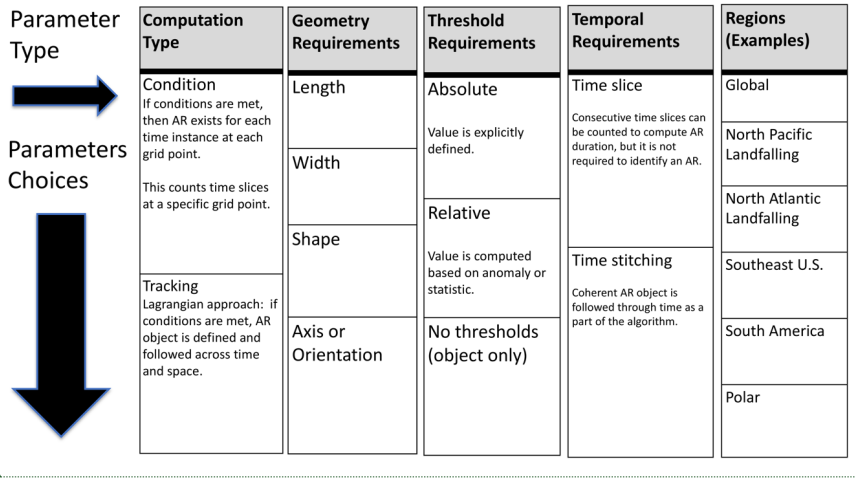
Formatted: Font: (Default) +Body (Times New Roman), 12 pt

Formatted: Font: (Default) +Body (Times New Roman), 12 pt

Formatted: Font color: Auto

9 Competing Interests

Paul Ullrich is a topical editor of GMD, otherwise, the authors declare that they have no conflicts of interest.



Moved down [1]: Figure 1: Schematic diagram illustrating the diversity on AR detection algorithms found in current literature by categorizing the variety of parameters used for identification and tracking, and then listing different types of choices available per category. ¶

Moved (insertion) [1]

Figure 1: Schematic diagram illustrating the diversity on AR detection algorithms found in current literature by categorizing the variety of parameters used for identification and tracking, and then listing different types of choices available per category.

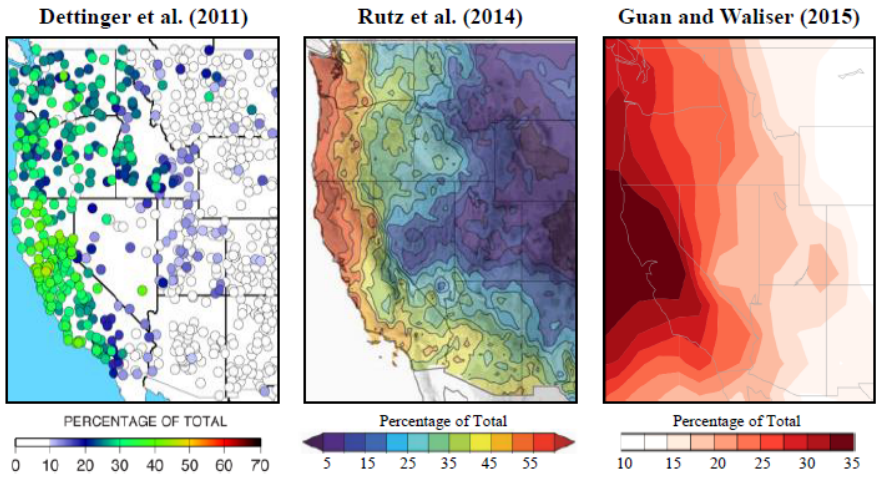


Figure 2: Examples of different algorithm results. (Left and center) The fraction of total cool-season precipitation attributable to ARs from Dettinger et al. (2011) and Rutz et al. (2014). (right) As in (left and center), but for annual precipitation from Guan and Waliser (2015). These studies use different AR identification methods, as well as different atmospheric reanalyses and observed precipitation data sets.

Moved down [2]: **Figure 2**: Examples of different algorithm results. (Left and center) The fraction of total cool-season precipitation attributable to ARs from Dettinger et al. (2011) and Rutz et al. (2014). (right) As in (left and center), but for annual precipitation from Guan and Waliser (2015). These studies use different AR identification methods, as well as different atmospheric reanalyses and observed precipitation data sets. ¶

Moved (insertion) [2]

Deleted: ¶
¶
¶
¶

Moved down [3]: **Figure 3**: Human control vs method counts (3-hour instances) at the coastline for landfalling ARS by latitude for the month of February using MERRA-2 3-hourly data. West refers to North Pacific ARs making landfall along Western North America, and East refers to North Atlantic ARs impacting European latitudes. Color lines represent detection algorithms and black lines represent the "human" control. The black solid line represents a static IVT 250 $kgm^{-1}s^{-1}$ threshold, and the black dashed (and dotted) lines represent static 2 and 1.5 cm IWV thresholds, respectively. Algorithm identifiers (A1, A2, etc.) specified in Table 1.¶

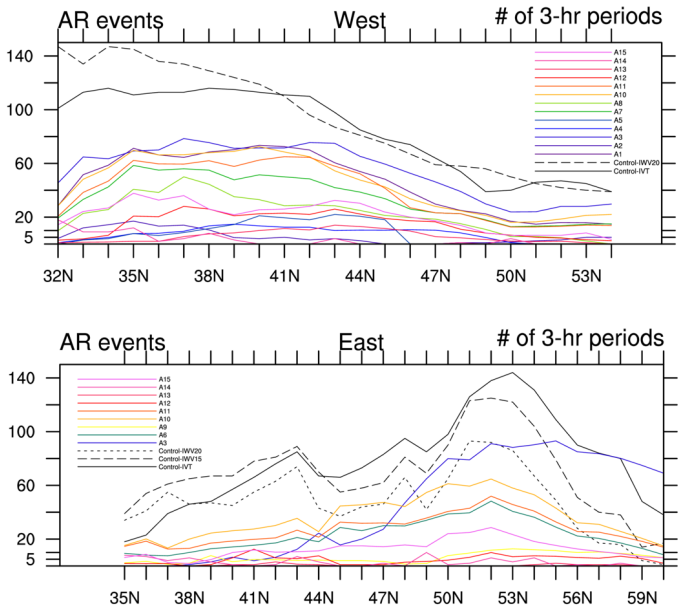


Figure 3: Human control vs method counts (3-hour instances) at the coastline for landfalling ARS by latitude for the month of February using MERRA-2 3-hourly data. West refers to North Pacific ARs making landfall along Western North America, and East refers to North Atlantic ARs impacting European latitudes. Color lines represent detection algorithms and black lines represent the "human" control. The black solid line represents a static IVT 250 $kgm^{-1}s^{-1}$ threshold, and the black dashed (and dotted) lines represent static 2 and 1.5 cm IWV thresholds, respectively. Algorithm identifiers (A1, A2, etc.) specified in Table 1.

Moved (insertion) [3]

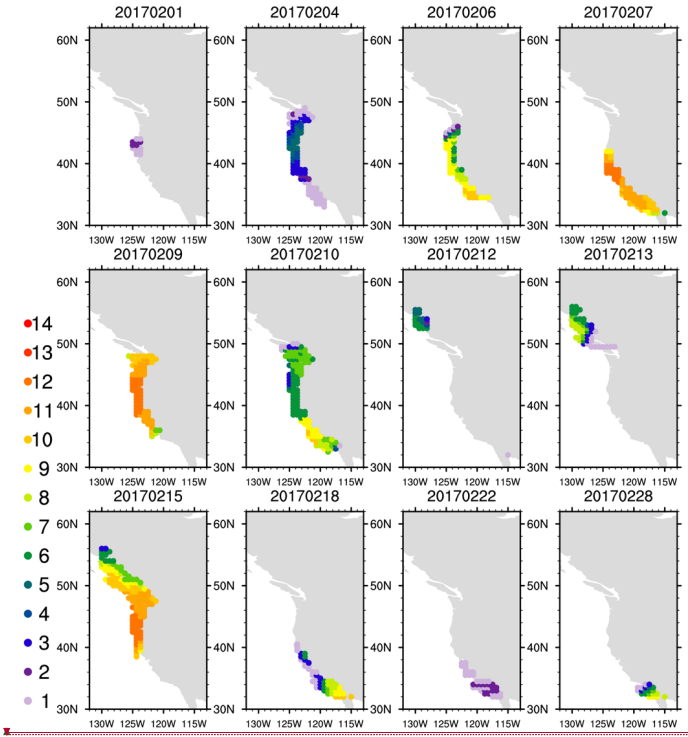
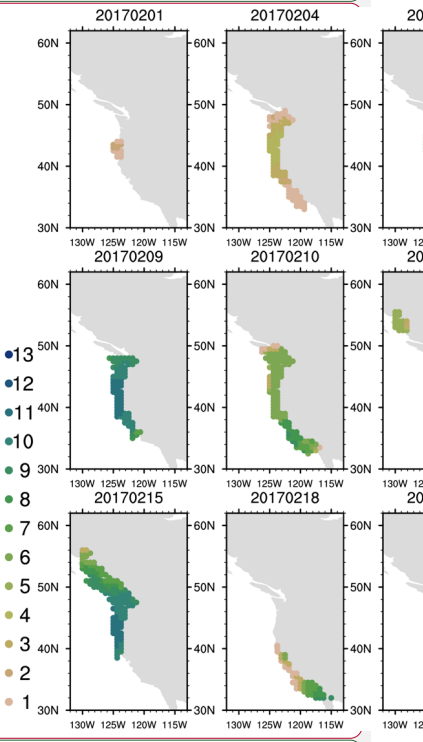


Figure 4: The number of methods that detect an AR at the coastline for sample days in February is plotted, plots are labeled with date in YYYYMMDD, i.e. 20170201 is February 1st, 2017. Because each day had 8 associated time steps, the maximum number of methods for each day is plotted. The polygons represent the number of methods, i.e. if only 1 method detected an AR at a specific grid point along the coast, then a light beige circle is plotted at that gridpoint along the coast; if 14 methods detected an AR at a specific grid point along the coast, then the darkest blue star is plotted at that grid point along the coast. Individual methods are not identified.

Moved down [4]: **Figure 4**: The number of methods that detect an AR at the coastline for sample days in February is plotted, plots are labeled with date in YYYYMMDD, i.e. 20170201 is February 1st, 2017. Because each day had 8 associated time steps, the maximum number of methods for each day is plotted. The polygons represent the number of methods, i.e. if only 1 method detected an AR at a specific grid point along the coast, then a light beige circle is plotted at that gridpoint along the coast; if 12 methods detected an AR at a specific grid point along the coast, then the darkest blue star is plotted at that grid point along the coast. Individual methods are not identified. ¶



Deleted:

Moved (insertion) [4]

Deleted: 2

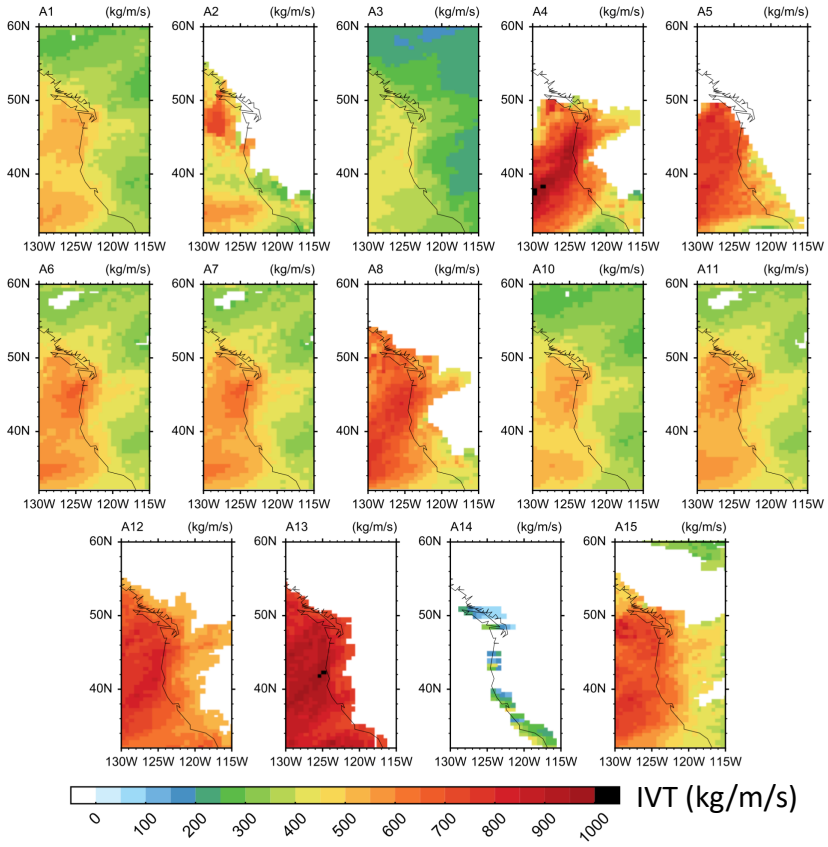
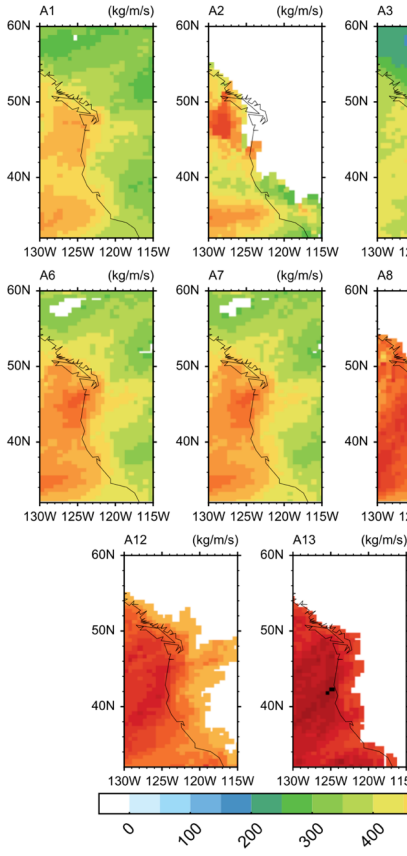


Figure 5a Composite MERRA-2 IVT (kg $m^{-1}s^{-1}$) for Western North America for all AR occurrences for all grid points where ARs are detected. Algorithm ids are found in Table 1. Algorithm A14 computes AR detection only for landfalling ARs at coastline grid points. The absence of color indicates no AR detection.

Moved down [5]: **Figure 5a** Composite MERRA-2 IVT (kg $m^{-1}s^{-1}$) for Western North America for all AR occurrences for all grid points where ARs are detected. Algorithm ids are found in Table 1. Algorithm A14 computes AR detection only for landfalling ARs at coastline grid points. The absence of color indicates no AR detection. ¶



Deleted:

Moved (insertion) [5]

Deleted: ¶

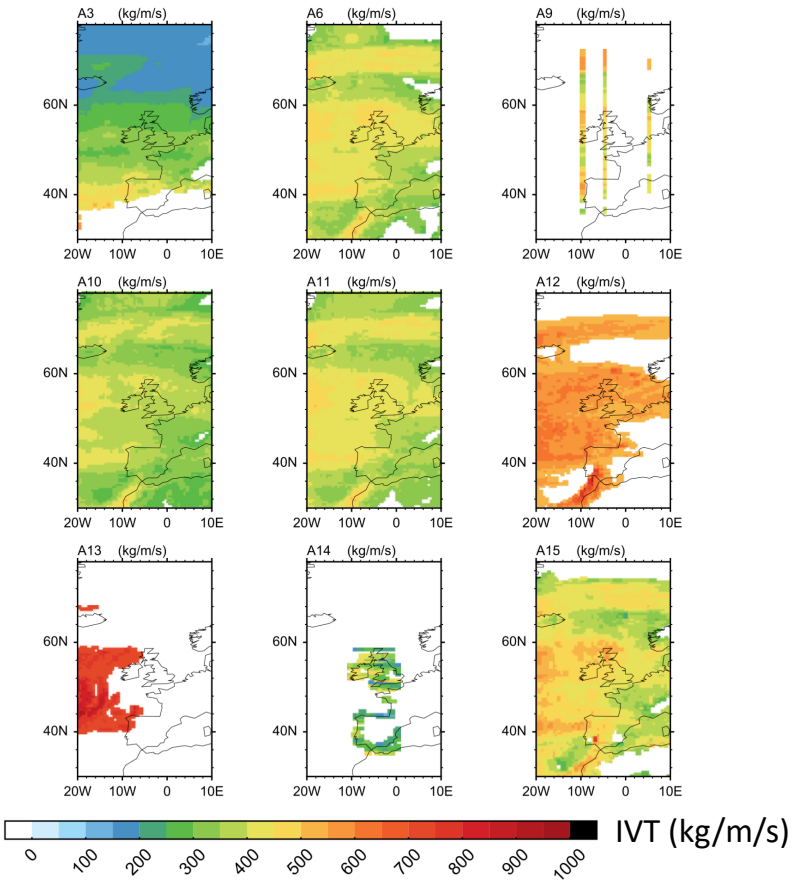
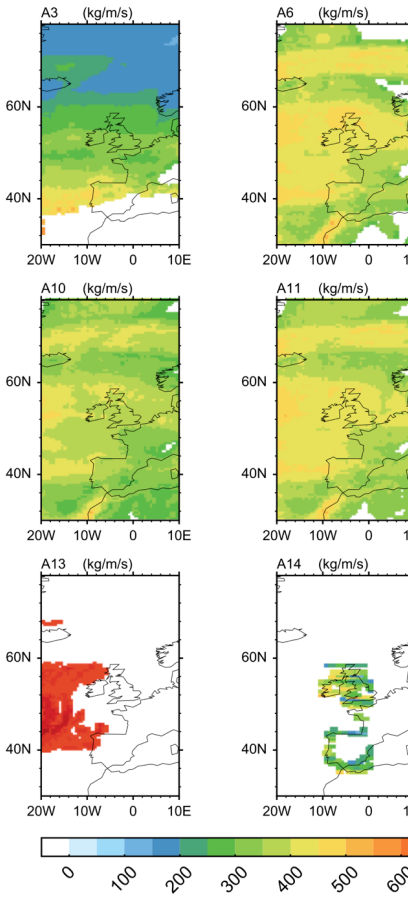


Figure 5b: Same as Figure 5a except of North Atlantic ARs. Algorithm A9 detects ARs at reference meridians. Note that the number of algorithms in this figure differs from 5a due to the regional constraint of the respective definitions.

Moved down [6]: **Figure 5b**: Same as a except of North Atlantic ARs. Algorithm A9 detects ARs at reference meridians. ¶



Deleted:

Moved (insertion) [6]

Deleted: a

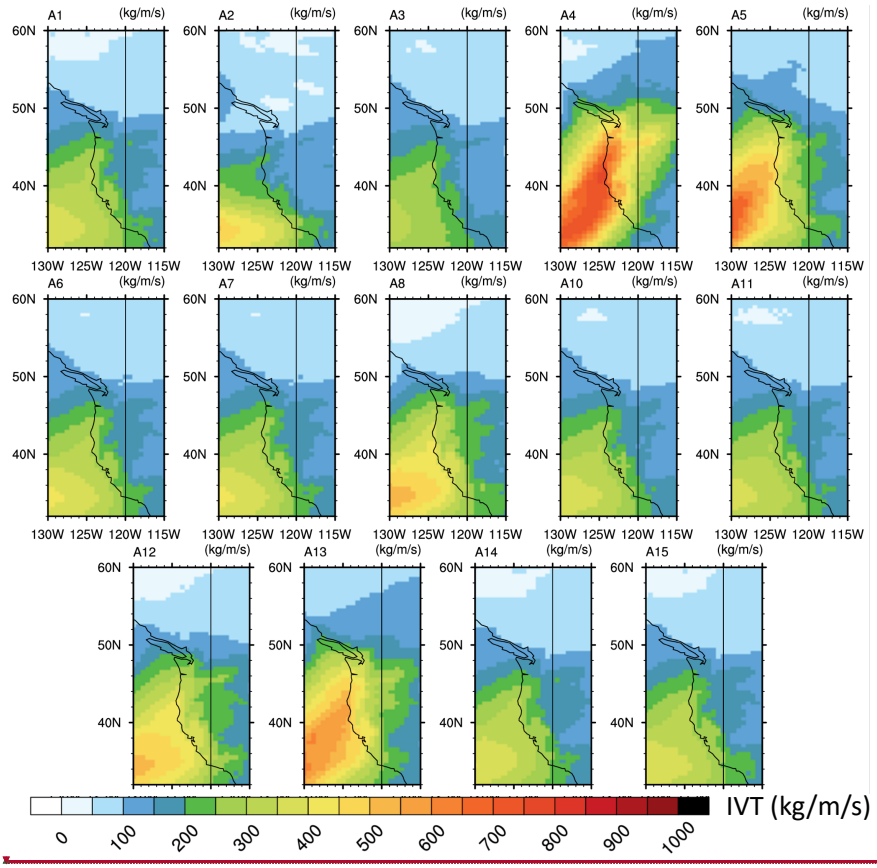
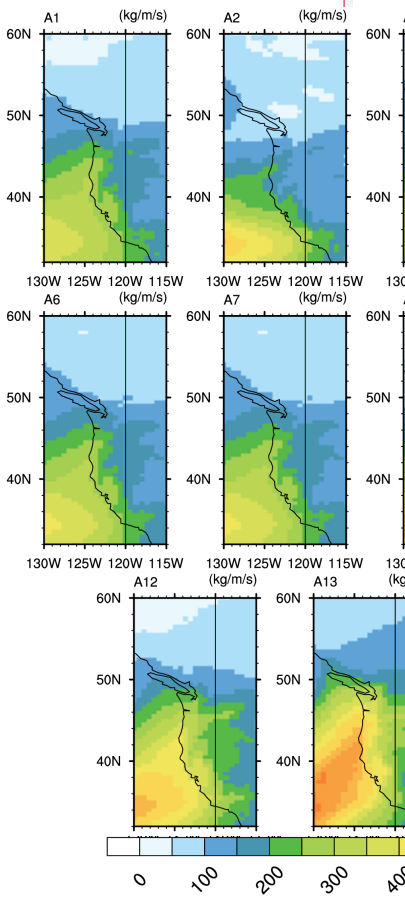


Figure 6a. Composite MERRA-2 IVT (kg $m^{-1}s^{-1}$) but for landfalling ARs only along North American west coast. Time instances where an AR was detected along the coastline where composited for the entire region. Algorithm masks are not necessary.

Deleted: ¶

Moved down [7]: **Figure 6a** Composite MERRA-2 IVT (kgm^{-1}s^{-1}) but for landfalling ARs only along North American west coast. Time instances where an AR was detected along the coastline where composited for the entire region. Algorithm masks are not necessary.¶



Deleted:

Moved (insertion) [7]

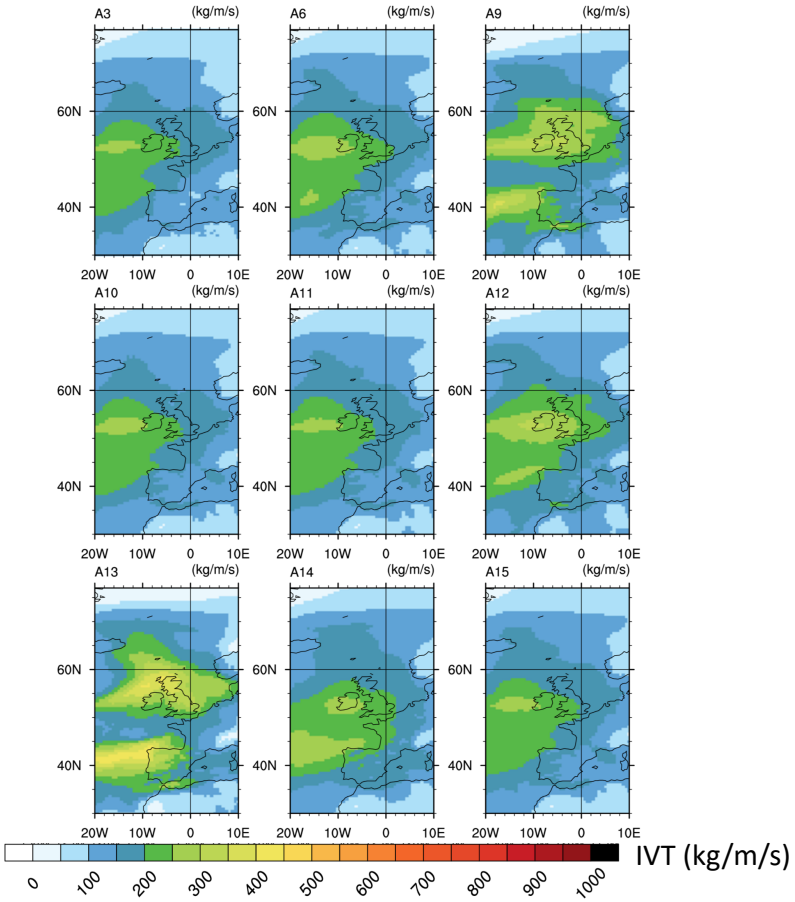


Figure 6b: Same as Figure 6a except for European coastlines. Note that the number of algorithms in this figure differs from 6a due to the regional constraint of the respective definitions.

Deleted: **Figure 6b:** As in a but for European coastlines.¶ ... [2]

Deleted: ¶

Moved down [8]: Figure 7: (top panel) Time series of daily IVT anomalies for (orange) Iberia, (teal) the U.S. West Coast and (blue) Ireland and the UK. Four events of varying geometry and intensity are shaded in the top panel and composites for each event are shown in the bottom four panels a - d. The black dots above the time series in the top panel indicate time slices in which each event is detected by an algorithm. ¶

Formatted: Font: Not Bold

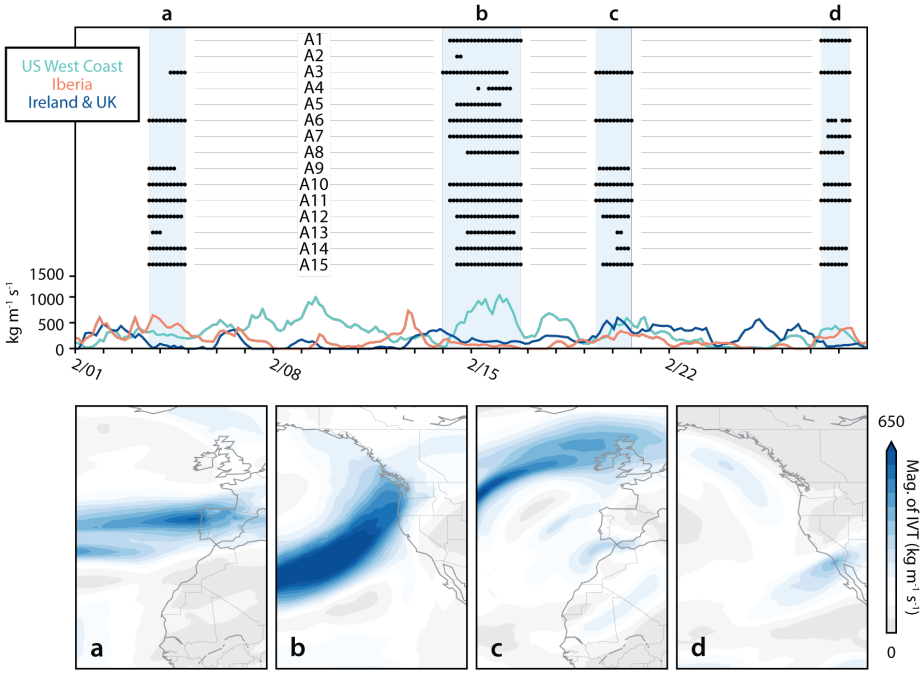


Figure 7: (top panel) Time series of daily IVT anomalies for (orange) Iberia, (teal) the U.S. West Coast and (blue) Ireland and the UK. Four events of varying geometry and intensity are shaded in the top panel and composites for each event are shown in the bottom four panels a - d. The black dots above the time series in the top panel indicate time slices in which each event is detected by an algorithm.

Moved (insertion) [8]

Moved down [9]: **Figure 8**: Focusing on the landfalling event in Fig. 7b, the average areal extent of the landfalling plume (y-axis) and average of the maximum precipitation rate at each detected time slice (x-axis) are compared for each algorithm. The size of the markers corresponds to the duration of the event as described in Fig. 7.¶

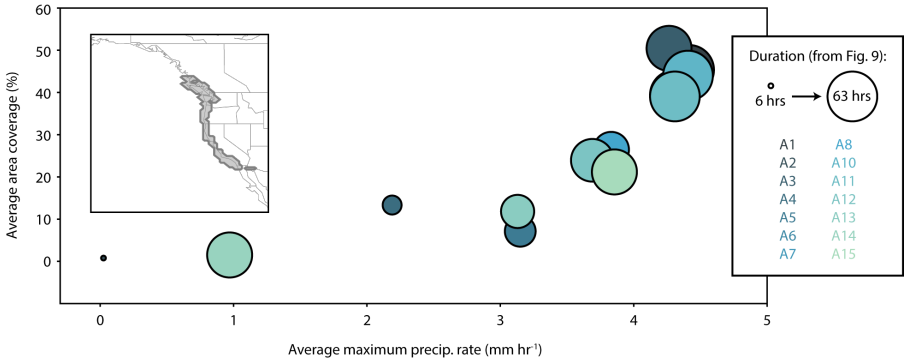


Figure 8: Focusing on the landfalling event in Fig. 7b, the average areal extent of the landfalling plume (y-axis) and average of the maximum precipitation rate at each detected time slice (x-axis) are compared for each algorithm. The size of the markers corresponds to the duration of the event as described in Fig. 7.

Moved (insertion) [9]

10 References

American Meteorological Society, cited 2017: "Atmospheric River". Glossary of Meteorology. [Available online at http://glossary.ametsoc.org/wiki/atmospheric river]

Baggett, C. F., Barnes, E. A., Maloney, E. D. and Mundhenk, B. D.: Advancing atmospheric river forecasts into subseasonal-to-seasonal time scales, Geophysical Research Letters, 44(14), 7528–7536, doi:10.1002/2017gl074434, 2017.

Barnes, E. A. and Polvani, L.: Response of the midlatitude jets, and of their variability, to increased greenhouse gases in the CMIP5 models, Journal of Climate, 26(18), 7117–7135, doi:10.1175/jcli-d-12-00536.1, 2013.

Bonne, J., Steen-Larsen, H.C., Risi, C., Werner, M., Sodemann, H., Lacour, J. Fettweis, X., Cesana, G., Delmotte, M., Cattani, O., Vallelonga, P., Kjær H.A., Clerbaux, C., Sveinbjörnsdóttir, A.E., and Masson-Delmotte,V.: The summer 2012 Greenland heat wave: In situ and remote sensing observations of water vapor isotopic composition during an atmospheric river event. J. Geophys. Res. Atmos., 120, 2970–2989. doi: 10.1002/2014JD022602, 2015.

Compo, G.P.,Whitaker, J.S., Sardeshmukh, P.D., Matsui, N.,Allan, R,J., Yin, X., Gleason,B.E., Vose, R.S., Rutledge, G., Bessemoulin, P., Brönnimann, S. Brunet, M., Crouthamel, R.I., Grant, A.N., Groisman, P.Y., Jones, P.D., Kruk, M., Kruger, A.C., Marshall, G.J., Maugeri, M., Mok, H.Y., Nordli, Ø. , Ross, T.F. , Trigo, R.M. , Wang, X.L., Woodruff, S.D. ,and Worley, S.J. : The Twentieth Century Reanalysis Project. Quarterly J. Roy. Meteorol. Soc., 137, 1-28. http://dx.doi.org/10.1002/qj.776, 2011.

Formatted: Line spacing: 1.5 lines

Cordeira, J. M., Ralph, F. M. and Moore, B. J.: The development and evolution of two atmospheric rivers in proximity to Western North Pacific tropical cyclones in October 2010, Monthly Weather Review, 141(12), 4234–4255, doi:10.1175/mwr-d-13-00019.1, 2013.

Dacre, H. F., Clark, P. A., Martinez-Alvarado, O., Stringer, M. A. and Lavers, D. A.: How do atmospheric rivers form?, Bulletin of the American Meteorological Society, 96(8), 1243–1255, doi:10.1175/bams-d-14-00031.1, 2015.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137(656), 553–597, doi:10.1002/qj.828, 2011.

DeFlorio, M., Waliser, D.E., Guan, B., Lavers, D., Ralph, F.M., and Vitart, F.: Global Assessment of Atmospheric River Prediction Skill. *J. Hydrometeor.*, **19**, 409–426, https://doi.org/10.1175/JHM-D-17-0135.1, 2018.

Delworth, T. L., Rosati, A., Anderson, W., Adcroft, A. J., Balaji, V., Benson, R., Dixon, K., Griffies, S. M., Lee, H.-C., Pacanowski, R. C., Vecchi, G. A., Wittenberg, A. T., Zeng, F. and

Formatted: Line spacing: 1.5 lines

Deleted: Global assessment of atmospheric river prediction skill, Journal of Hydrometeorology, *Submitted with Revisions*., 2017a.

Formatted: Font: (Default) +Body (Times New Roman)

Deleted: ¶
DeFlorio, M., Waliser, D.E., Guan, B., Lavers, D., Ralph F.M., and Vitart, F.: Global evaluation of atmospheric river subseasonal prediction skill, Journal of Climate, *In Prep*, 2017b.¶

Zhang, R.: Simulated climate and climate change in the GFDL CM2.5 high-resolution coupled climate model, Journal of Climate, 25(8), 2755–2781, doi:10.1175/jcli-d-11-00316.1, 2012.

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

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

Gao, Y., Lu, J., Leung, L. R., Yang, Q., Hagos, S. and Qian, Y.: Dynamical and thermodynamical modulations on future changes of landfalling atmospheric rivers over western North America, Geophysical Research Letters, 42(17), 7179–7186, doi:10.1002/2015gl065435, 2015.

Gao, Y., Lu, J. and Leung, L. R.: Uncertainties in projecting future changes in atmospheric rivers and their impacts on heavy precipitation over Europe, Journal of Climate, 29(18), 6711–6726, doi:10.1175/jcli-d-16-0088.1, 2016.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M. and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), Journal of Climate, 30(14), 5419–5454, doi:10.1175/jcli-d-16-0758.1, 2017.

Gershunov, A., Shulgina, T., Ralph, F. M., Lavers, D. A. and Rutz, J. J.: Assessing the climate-scale variability of atmospheric rivers affecting western North America, Geophysical Research Letters, 44(15), 7900–7908, doi:10.1002/2017gl074175, 2017.

Global Modeling and Assimilation Office (GMAO)(2015), MERRA-2 inst3_3d_asm_Np: 3d,3-Hourly,Instantaneous,Pressure-Level,Assimilation,Assimilated Meteorological Fields V5.12.4, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed **[*Data Access Date*]** 10.5067/QBZ6MG944HW0.

Goldenson, N., Leung, L.R., Bitz, C.M., Blanchard-Wrigglesworth, E.: Influence of Atmospheric River Events on Mountain Snowpack of the Western U.S., (submitted, JCLI-D-18-0268).

Gorodetskaya, I. V., Tsukernik, M., Claes, K., Ralph, M. F., Neff, W. D. and Van Lipzig, N. P. M.: The role of atmospheric rivers in anomalous snow accumulation in East Antarctica, Geophysical Research Letters, 41(17), 6199–6206, doi:10.1002/2014gl060881, 2014.

Guan, B., Molotch, N. P., Waliser, D. E., Fetzer, E. J. and Neiman, P. J.: Extreme snowfall events linked to atmospheric rivers and surface air temperature via satellite measurements, Geophysical Research Letters, 37(20), n/a-n/a, doi:10.1029/2010gl044696, 2010.

Deleted: in review

Guan, B. and Waliser, D. E.: Detection of atmospheric rivers: Evaluation and application of an algorithm for global studies, Journal of Geophysical Research: Atmospheres, 120(24), 12514–12535, doi:10.1002/2015jd024257, 2015.

Guan, B. and Waliser, D. E.: Atmospheric rivers in 20 year weather and climate simulations: A multimodel, global evaluation, Journal of Geophysical Research: Atmospheres, 122(11), 5556–5581, doi:10.1002/2016jd026174, 2017.

Guan, B., Waliser D., Ralph, F. M. : An inter-comparison between reanalysis and dropsonde observations of the total water vapor transport in individual atmospheric river, JHM, revision *in review*, 2017.

Hagos, S., Leung, L. R., Yang, Q., Zhao, C. and Lu, J.: 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, 2015.

Hagos, S. M., Leung, L. R., Yoon, J.-H., Lu, J. and Gao, Y.: A projection of changes in landfalling atmospheric river frequency and extreme precipitation over western North America from the Large Ensemble CESM simulations, Geophysical Research Letters, 43(3), 1357–1363, doi:10.1002/2015gl067392, 2016.

Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D. and New, M.: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, Journal of Geophysical Research, 113(D20), doi:10.1029/2008jd010201, 2008.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P. and Stocker, E. F.: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-Global, Multiyear, Combined-Sensor Precipitation Estimates at Fine Scales, Journal of Hydrometeorology, 8(1), 38–55, doi:10.1175/jhm560.1, 2007.

Huffman, G. J., Adler, R. F., Morrissey, M. M., Bolvin, D. T., Curtis, S., Joyce, R., McGavock, B. and Susskind, J.: Global precipitation at one-degree daily resolution from multisatellite observations, Journal of Hydrometeorology, 2(1), 36–50, doi:10.1175/1525-7541(2001)002<0036:gpaodd>2.0.co;2, 2001.

Huning, L. S., Margulis, S. A., Guan, B., Waliser, D. E. and Neiman, P. J.: Implications of detection methods on characterizing atmospheric river contribution to seasonal snowfall across Sierra Nevada, USA, Geophysical Research Letters, doi:10.1002/2017gl075201, 2017.

Jankov, I., Bao, J.-W., Neiman, P. J., Schultz, P. J., Yuan, H. and White, A. B.: Evaluation and comparison of microphysical algorithms in ARW-WRF model simulations of atmospheric river events affecting the California coast, Journal of Hydrometeorology, 10(4), 847–870, doi:10.1175/2009jhm1059.1, 2009.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R. and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, Bulletin of the American Meteorological Society, 77(3), 437–471, doi:10.1175/1520 0477(1996)077<0437:tnyrp>2.0.co;2, 1996.

Kim, J., Waliser, D. E., Neiman, P. J., Guan, B., Ryoo, J.-M. and Wick, G. A.: Effects of atmospheric river landfalls on the cold season precipitation in California, Climate Dynamics, 40(1–2), 465–474, doi:10.1007/s00382-012-1322-3, 2012.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., MiyaokaI, K. and Takahashi, K.: The JRA-55 Reanalysis: General Specifications and Basic Characteristics, Journal of the Meteorological Society of Japan. Ser. II, 93(1), 5–48, doi:10.2151/jmsj.2015-001, 2015.

Lavers, D. A., Villarini, G., Allan, R. P., Wood, E. F. and Wade, A. J.: 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(D20), doi:10.1029/2012jd018027, 2012.

Lavers, D. A., Allan, R. P., Villarini, G., Lloyd-Hughes, B., Brayshaw, D. J. and Wade, A. J.: Future changes in atmospheric rivers and their implications for winter flooding in Britain, Environmental Research Letters, 8(3), 34010, doi:10.1088/1748-9326/8/3/034010, 2013.

Lavers, D. A., Pappenberger, F. and Zsoter, E.: Extending medium-range predictability of extreme hydrological events in Europe, Nature Communications, 5, 5382, doi:10.1038/ncomms6382, 2014.

Lavers, D. A. and Villarini, G.: The contribution of atmospheric rivers to precipitation in Europe and the United States, Journal of Hydrology, 522, 382–390, doi:10.1016/j.jhydrol.2014.12.010, 2015.

Lavers, D. A., Ralph, F.M., Waliser, D. E., Gershunov, A. and Dettinger, M. D., Climate change intensification of horizontal water vapor transport in CMIP5, Geophys. Res. Lett.,42, doi:10.1002/2015GL064672, 2015.

Leung, L. R. and Qian, Y.: Atmospheric rivers induced heavy precipitation and flooding in the western U.S. simulated by the WRF regional climate model, Geophysical Research Letters, 36(3), n/a-n/a, doi:10.1029/2008gl036445, 2009.

Livneh, B., Rosenberg, E. A., Lin, C., Nijssen, B., Mishra, V., Andreadis, K. M., Maurer, E. P. and Lettenmaier, D. P.: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: Update and Extensions*, Journal of Climate, 26(23), 9384–9392, doi:10.1175/jcli-d-12-00508.1, 2013.

Lora, J. M., Mitchell, J. L., Risi, C. and Tripati, A. E.: North Pacific atmospheric rivers and their influence on western North America at the Last Glacial Maximum, Geophysical Research Letters, 44(2), 1051–1059, doi:10.1002/2016gl071541, 2017.

Mahoney, K., Jackson, D. L., Neiman, P., Hughes, M., Darby, L., Wick, G., White, A., Sukovich, E. and Cifelli, R.: Understanding the role of atmospheric rivers in heavy precipitation in the Southeast United States, Monthly Weather Review, 144(4), 1617–1632, doi:10.1175/mwr-d-15-0279.1, 2016.

Deleted: Lavers, D. A., Ralph, F.M., Waliser, D.E., Gershunov, A. and, Dettinger, M.D. : Intensification of water vapor transport along the atmospheric branch of the global water cycle in CMIP5 models, Geophysical Research Letters, *Submitted*, 2015.

Formatted: Font: (Default) +Body (Times New Roman), 12 pt

Formatted: Line spacing: 1.5 lines

Formatted: Font: (Default) +Body (Times New Roman), 12 pt

Formatted: Font: (Default) +Body (Times New Roman), 12 pt

Formatted: Font: (Default) +Body (Times New Roman), 12 pt

Formatted: Font: (Default) +Body (Times New Roman), 12 pt

Formatted: Font: (Default) +Body (Times New Roman), 12 pt

Formatted: Font: (Default) +Body (Times New Roman), 12 pt

Formatted: Font: (Default) +Body (Times New Roman), 12 pt

Formatted: Font: (Default) +Body (Times New Roman), 12 pt

Deleted: ¶

Nayak, M. A., Villarini, G. and Lavers, D. A.: On the skill of numerical weather prediction models to forecast atmospheric rivers over the central United States, Geophysical Research Letters, 41(12), 4354–4362, doi:10.1002/2014gl060299, 2014.

Neale, R.B., Chen, C.C., Gettelman, A., Lauritzen, P.H., Park, S., Williamson, D.L., Conley, A.J., Garcia, R., Kinnison, D., Lamarque, J.F. and Marsh, D.: Description of the NCAR community atmosphere model (CAM 5.0). *NCAR Tech. Note NCAR/TN-486+ STR*, 2010.

Neff, W., Compo, G.P., Ralph F.M., and Shupe, M.D.: Continental heat anomalies and the extreme melting of the Greenland ice surface in 2012 and 1889, J. Geophys. Res. Atmos., 119, 6520–6536, doi:10.1002/2014JD021470, 2014.

Neiman, P. J., Ralph, F. M., Wick, G. A., Kuo, Y.-H., Wee, T.-K., Ma, Z., Taylor, G. H. and Dettinger, M. D.: 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, 2008.

Neiman, P. J., Schick, L. J., Ralph, F. M., Hughes, M. and Wick, G. A.: Flooding in western Washington: The connection to atmospheric rivers, Journal of Hydrometeorology, 12(6), 1337–1358, doi:10.1175/2011jhm1358.1, 2011.

Newman, M., Kiladis, G.N., Weickmann, K.M., Ralph, F.M., and Sardeshmukh, P.D.: Relative contributions of synoptic and low-frequency eddies to time-mean atmospheric moisture transport, including the role of atmospheric rivers, J. Clim., 25, 7341–7361, 2012.

Formatted: Line spacing: 1.5 lines

Deleted: ¶

Neu, U., Akperov, M. G., Bellenbaum, N., Benestad, R., Blender, R., Caballero, R., Cocozza, A., Dacre, H. F., Feng, Y., Fraedrich, K., Grieger, J., Gulev, S., Hanley, J., Hewson, T., Inatsu, M., Keay, K., Kew, S. F., Kindem, I., Leckebusch, G. C., Liberato, M. L. R., Lionello, P., Mokhov, I. I., Pinto, J. G., Raible, C. C., Reale, M., Rudeva, I., Schuster, M., Simmonds, I., Sinclair, M., Sprenger, M., Tilinina, N. D., Trigo, I. F., Ulbrich, S., Ulbrich, U., Wang, X. L. and Wernli, H.: IMILAST: A community effort to intercompare extratropical cyclone detection and tracking algorithms, Bulletin of the American Meteorological Society, 94(4), 529–547, doi:10.1175/bams-d-11-00154.1, 2013.

Paltan, H., Waliser, D., Lim, W. H., Guan, B., Yamazaki, D., Pant, R. and Dadson, S.: Global floods and water availability driven by atmospheric rivers, Geophysical Research Letters, doi:10.1002/2017gl074882, 2017.

Payne, A. E. and Magnusdottir, G.: 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, doi:10.1002/2015jd023586, 2015.

Payne, A. E. and Magnusdottir, G.: Persistent landfalling atmospheric rivers over the west coast of North America, Journal of Geophysical Research: Atmospheres, 121(22), 13,287-13,300, doi:10.1002/2016jd025549, 2016.

Ralph, F. M., Neiman, P. J. and Wick, G. A.: 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, 2004.

Ralph, F. M., Neiman, P. J., Wick, G. A., Gutman, S. I., Dettinger, M. D., Cayan, D. R. and White, A. B.: Flooding on California's Russian River: Role of atmospheric rivers, Geophysical Research Letters, 33(13), doi:10.1029/2006gl026689, 2006.

Ralph, F. M., Coleman, T., Neiman, P. J., Zamora, R. J. and Dettinger, M. D.: 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, doi:10.1175/jhm-d-12-076.1, 2013.

Ralph, F. M., Prather, K. A., Cayan, D., Spackman, J. R., DeMott, P., Dettinger, M., Fairall, C., Leung, R., Rosenfeld, D., Rutledge, S., Waliser, D., White, A. B., Cordeira, J., Martin, A., Helly, J. and Intrieri, J.: CalWater Field studies designed to quantify the roles of atmospheric rivers and aerosols in modulating U.S. West Coast precipitation in a changing climate, Bulletin of the American Meteorological Society, 97(7), 1209–1228, doi:10.1175/bams-d-14-00043.1, 2016.

Ralph, F. M., Dettinger, M., Lavers, D., Gorodetskaya, I. V., Martin, A., Viale, M., White, A. B., Oakley, N., Rutz, J., Spackman, J. R., Wernli, H. and Cordeira, J.: 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, 2017.

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

Ramos, A. M., Tomé, R., Trigo, R. M., Liberato, M. L. R. and Pinto, J. G.: Projected changes in atmospheric rivers affecting Europe in CMIP5 models, Geophysical Research Letters, 43(17), 9315–9323, doi:10.1002/2016gl070634, 2016.

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

Ryoo, J.-M., Waliser, D. E. and Fetzer, E. J.: Trajectory analysis on the origin of air mass and moisture associated with Atmospheric Rivers over the west coast of the United States, Atmospheric Chemistry and Physics Discussions, 11(4), 11109–11142, doi:10.5194/acpd-11-11109-2011, 2011.

Ryoo, J.-M., Waliser, D. E., Waugh, D. W., Wong., S., Fetzer, E. J., and Fung, I.: Classification of atmospheric river events on the U.S. West Coast using a trajectory model, Journal of Geophysical Research Atmospheres, 120, 3007–3028. doi: 10.1002/2014JD022023.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M. P., van den Dool, H., Zhang, Q., Wang, W., Chen, M. and Becker, E.: The NCEP Climate Forecast System Version 2, Journal of Climate, 27(6), 2185–2208, doi:10.1175/jcli-d-12-00823.1, 2014.

Sellars, S., Nguyen, P., Chu, W., Gao, X., Hsu, K. and Sorooshian, S.: Computational Earth Science: Big data transformed into insight, Eos, Transactions American Geophysical Union, 94(32), 277–278, doi:10.1002/2013eo320001, 2013.

Sellars, S. L., Gao, X. and Sorooshian, S.: An Object-oriented approach to investigate impacts of climate oscillations on precipitation: A western United States case study, Journal of Hydrometeorology, 16(2), 830–842, doi:10.1175/jhm-d-14-0101.1, 2015.

Shields, C. A. and Kiehl, J. T.: Atmospheric river landfall-latitude changes in future climate simulations, Geophysical Research Letters, 43(16), 8775–8782, doi:10.1002/2016gl070470, 2016.

Shields, C. A. and Kiehl, J. T.: Simulating the Pineapple Express in the half degree Community Climate System Model, CCSM4, Geophysical Research Letters, 43(14), 7767–7773, doi:10.1002/2016gl069476, 2016.

Shields, C. A., Kiehl, J. T. and Meehl, G. A.: Future changes in regional precipitation simulated by a half-degree coupled climate model: Sensitivity to horizontal resolution, Journal of Advances in Modeling Earth Systems, 8(2), 863–884, doi:10.1002/2015ms000584, 2016.

Small, R. J., Bacmeister, J., Bailey, D., Baker, A., Bishop, S., Bryan, F., Caron, J., Dennis, J., Gent, P., Hsu, H., Jochum, M., Lawrence, D., Muñoz, E., diNezio, P., Scheitlin, T., Tomas, R., Tribbia, J., Tseng, Y. and Vertenstein, M.: A new synoptic scale resolving global climate simulation using the Community Earth System Model, Journal of Advances in Modeling Earth Systems, 6(4), 1065–1094, doi:10.1002/2014ms000363, 2014.

Sorooshian, S., Hsu, K.-L., Gao, X., Gupta, H. V., Imam, B. and Braithwaite, D.: Evaluation of PERSIANN system satellite–based estimates of tropical rainfall, Bulletin of the American Meteorological Society, 81(9), 2035–2046, doi:10.1175/1520-0477(2000)081<2035:eopsse>2.3.co;2, 2000.

Viale, M. and Nuñez, M. N.: 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, 2011.

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

Warner, M. D., Mass, C. F. and Salathé, E. P., Jr.: Changes in winter atmospheric rivers along the North American West Coast in CMIP5 Climate Models, Journal of Hydrometeorology, 16(1), 118–128, doi:10.1175/jhm-d-14-0080.1, 2015.

Warner, M. D., and Mass, C. F.: Changes in the climatology, structure, and seasonality of Northeast Pacific atmospheric rivers in CMIP5 climate simulations, Journal of Hydrometeorology, 18, 2131 – 2141, doi: 10.1175/JHM-D-16-0200.1.

Wehner, M.F., Reed, K.,Li, F., Prabhat, Bacmeister, J., Chen, C., Paciorek, C., Gleckler,P. , Sperber, K., Collins, W.D., Gettelman, A., Jablonowski, C.: The effect of horizontal resolution on simulation quality in the Community Atmospheric Model, CAM5.1. Journal of Modeling the Earth System 06, 980-997. doi:10.1002/2013MS000276, 2014.

Wick, G. A., Neiman, P. J., Ralph, F. M. and Hamill, T. M.: Evaluation of forecasts of the water vapor signature of atmospheric rivers in operational numerical weather prediction models, Weather and Forecasting, 28(6), 1337–1352, doi:10.1175/waf-d-13-00025.1, 2013.

Wick, G. A., Neiman, P. J. and Ralph, F. M.: Description and validation of an automated objective technique for identification and characterization of the integrated water vapor signature of atmospheric rivers, IEEE Transactions on Geoscience and Remote Sensing, 51(4), 2166–2176, doi:10.1109/tgrs.2012.2211024, 2013.

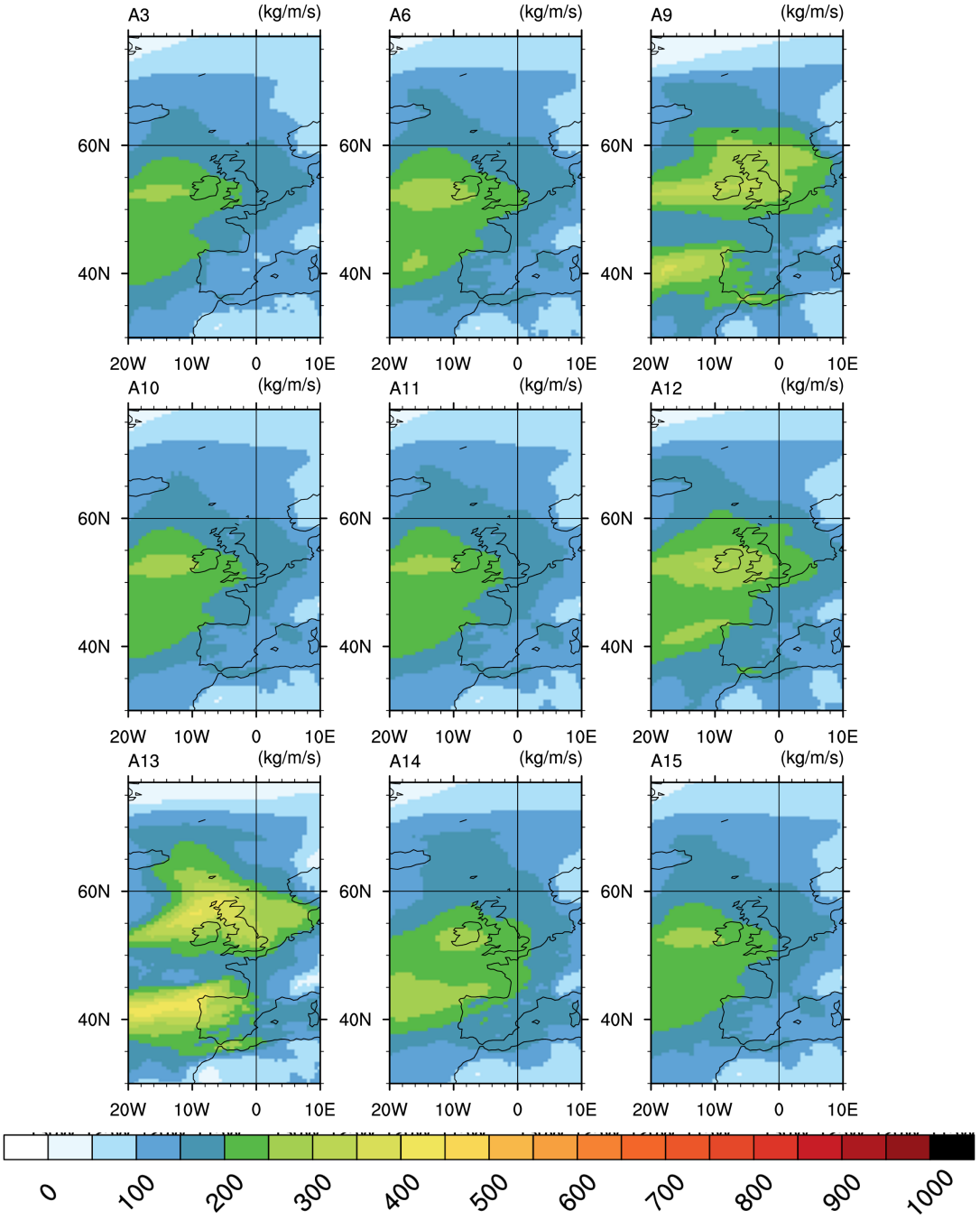
Zhu, Y. and Newell, R. E.: A proposed algorithm for moisture fluxes from atmospheric rivers, Monthly Weather Review, 126(3), 725–735, doi:10.1175/1520-0493(1998)126<0725:apafmf>2.0.co;2, 1998.

Page 18: [1] Deleted	Microsoft Office User	5/2/18 12:40:00 PM

Spring 2018	2nd ARTMIP Workshop
Summer/Fall 2018	Tier 2 Analysis and Scientific Papers

Page 39: [2] Deleted	Microsoft Office User	5/1/18 3:24:00 PM

Figure 6b: As in a but for European coastlines.



Atmospheric River Tracking Method Intercomparison Project (ARTMIP): Project Goals and Experimental Design

Christine A. Shields[1], Jonathan J. Rutz[2], Lai-Yung Leung[3], F. Martin Ralph[4], Michael Wehner[5], Brian Kawzenuk[4], Juan M. Lora[6], Elizabeth McClenny[7], Tashiana Osborne[4], Ashley E. Payne[8], Paul Ullrich[7], Alexander Gershunov[4], Naomi Goldenson[9], Bin Guan[10], Yun Qian[3], Alexandre M. Ramos[11], Chandan Sarangi[3], Scott Sellars[4], Irina Gorodetskaya[12], Karthik Kashinath[13], Vitaliy Kurlin[14], Kelly Mahoney[15], Grzegorz Muszynski[13,14],Roger Pierce[16], Aneesh C. Subramanian[4], Ricardo Tome[11], Duane Waliser[17], Daniel Walton[18], Gary Wick[15], Anna Wilson[4], David Lavers[19], Prabhat[5], Allison Collow[20], Harinarayan Krishnan[5], Gudrun Magnusdottir[21], Phu Nguyen[22]

[1] Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder, CO, 80302, USA
[2]Science and Technology Infusion Division, National Weather Service Western Region Headquarters, National Oceanic and Atmospheric Administration, Salt Lake City, Utah, 84138, USA
[3] Earth Systems Analysis and Modeling, Pacific Northwest National Laboratory, Richland, Washington, 99354, USA
[4] Center for Western Weather and Water Extremes, Scripps Institution of Oceanography, La Jolla, California, 92037, USA
[5] Computational Chemistry, Materials, and Climate Group, Lawrence Berkeley National Laboratory, Berkeley, California, 94720, USA
[6] Department of Earth, Planetary, and Space Sciences, University of California, Los Angeles, California, 90095, USA
[7]Department of Land, Air and Water Resources, University of California, Davis, California, 95616, USA
[8]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, Michigan, 48109, USA
[9]Department of Atmospheric Sciences, University of Washington, Seattle, Washington, 98195, USA
[10]Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, California, 90095, USA
[11]Instituto Dom Luiz, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal
[12]Centre for Environmental and Marine Studies, University of Aveiro, 3810-193 Aveiro, Portugal
[13]Data & Analytics Services, National Energy Research Scientific Computing Center (NERSC), Lawrence Berkeley National Laboratory, Berkeley, California, 94720, USA
[14]Department Computer Science Liverpool, Liverpool, L69 3BX, UK
[15]Physical Sciences Division, Earth System Research Laboratory, National Oceanic and Atmospheric Administration, Boulder, CO, 80305, USA
[16]National Weather Service Forecast Office, National Oceanic and Atmospheric Administration, San Diego, CA, 92127, USA
[17]Earth Science and Technology Directorate, Jet Propulsion Laboratory, Pasadena, California, 91109, USA
[18]Institute of the Environment and Sustainability, University of California, Los Angeles, California, 90095, USA
[19]European Centre for Medium-Range Weather Forecasts, Reading, RG2 9AX, UK

[20]Universities Space Research Association, Columbia, MD, 21046, USA
[21]Department of Earth System Science, University of California Irvine, CA 92697, USA
[22]Department of Civil & Environmental Engineering, University of California Irvine, CA 92697, USA
Correspondence to: Christine A. Shields (shields@ucar.edu)

Supplemental Text

1 IVT-UCSD derived 3-hourly versus IVT-MERRA-provided 1hourly

Comparing the 1-h to 3-h MERRA-2 provided time-averaged data suggests that there are only very small differences between the two, meaning that there is little value added to using 1-h data. Figure S1 shows the difference between the average IVT magnitude for February 2017 computed with those two frequencies. Note the small magnitudes of the colorbar. However, comparing the time-averaged 3-h MERRA-2-provided IVT (kg $m^{-1}s^{-1}$) to the 3-h IVT computed by the UCSD group, there are more substantial discrepancies over regions of substantial topographical relief, as well as regions where the surface pressure is considerably different from 1000hPa. Supplemental Figure2 shows the same difference as Figure S1, but for these two datasets. In particular, the UCSD-provided data overestimates IVT over topographic highs and underestimates it by various amounts over ocean regions, most dramatically in tropical and subtropical regions. This is consistent with the vertical integrals being done on the model grid in the one case, and with interpolated values over a limited pressure range in the other. Because some of these differences could be the difference between a detection and a non-detection, it is imperative that all participants use the same dataset to create their 1980-2017 AR catalogues. Due to resource constraints, as a group, we are using the lower temporal resolution dataset provided by UCSD.

2 Data formats for ARTMIP Submissions

2.1 All Submissions

The output should take the form of a NetCDF4 file with filename formatted as follows. For the purposes of the one-month analysis of MERRA2 data, Reanalysis datasetshould be MERRA2and time frequency should be 3hourly. The dateis either a single day in the form YYYYMMDD or a range of dates in the form YYYYMMDD-YYYYMMDD.

```
<Reanalysis dataset>.ar_tag.<Algorithm>.<hourly/3hourly/6hourly/daily>.<Date(s)>.nc4
```

The structure of the NetCDF4 file should follow the following format, for an example algorithm entitled Namev1run on 3-hourly data. The time, lat, and lonvariables should be inherited directly from the MERRA2 dataset. The ar_binary_tag variable should be of type byte(an 8-bit integer) so as to limit the size of each submission. This variable should be 1 if an atmospheric river is detected at this grid point / time and 0 if no detection occurred. Any number of files can be provided, although we suggest one file per processed day.

```
ncdump -h MERRA2.ar_tag.Name_v1.3hourly.20170201.nc4
netcdf

MERRA2.ar_tag.Name_v1.3hourly.20170201 { dimensions:

    time = UNLIMITED ; // (8
    currently) lat = 361 ;

    lon = 576 ;
variables:
```

```
    double time(time) ;
        time:standard_name = "time"
        ; time:long_name = "time" ;

        time:units = "minutes since 2017-02-01
        00:00:00" ; time:calendar = "standard" ;

    double lat(lat) ;

        lat:standard_name =
        "latitude" ; lat:long_name =
        "latitude" ; lat:units =
        "degrees_north" ; lat:axis =
        "Y" ;

    double lon(lon) ;

        lon:standard_name =
        "longitude" ; lon:long_name
        = "longitude" ; lon:units =
        "degrees_east" ; lon:axis =
        "X" ;

    byte ar_binary_tag(time, lat, lon) ;

        ar_binary_tag:description = "binary indicator of atmospheric river" ;
        ar_binary_tag:scheme = "Jiang" ;

        ar_binary_tag:version = "1.0" ;

}
```

2.2 Methods with Regional Coverage

We acknowledge that not all algorithms will provide global data. In the case that only regional data is available, or in the case of some algorithms, data only at several grid points along the coast, we would request that an additional file is submitted with filename given by:

```
<Reanalysis dataset>.ar_mask.<Algorithm>.nc4
```

This NetCDF4 file should be based on the following template:

```
ncdump -h MERRA2.ar_mask.nc4

netcdf
MERRA2.ar_mask {
dimensions:
    lat =
    361 ;
    lon =
    576 ;

variables:
    double lat(lat) ;

        lat:standard_name =
        "latitude" ; lat:long_name =
        "latitude" ; lat:units =
        "degrees_north" ; lat:axis =
        "Y" ;
```

```
    double lon(lon) ;

        lon:standard_name =
        "longitude" ; lon:long_name
        = "longitude" ; lon:units =
        "degrees_east" ; lon:axis =
        "X" ;

    byte ar_binary_mask(lat, lon) ;

        ar_binary_mask:description = "atmospheric river regional coverage mask" ;
        ar_binary_mask:scheme = "Jiang" ;

        ar_binary_mask:version = "1.0" ;

}
```

The value of ar_binary_mask should be binary (either 1 or 0). Grid cells that are covered by the detection scheme should be given a value of 1, and those that are not covered should be given a value of 0.

Example masks from 1-month test participants are seen in Supplementary Figure S3.

3 Human Control Methodology

The same criteria for identification was used for both counting and tracking human controls and included the following:

1) IVT > 250 $kgm^{-1}s^{-1}$
2) IWV > 20mm
3) Length ~2 x width in IVT and/or IWV field;
4) Or at least a general stretched structure

5) Generally westerly flow
6) IVTy > 0 (poleward transport)

Data was compiled onto spreadsheets and available to ARTMIP participants.

3.1 Human control "Tracking"

The brightest IVT pixel is what is chosen as the recorded longitude and latitude. This single pixel is not followed temporally, so the given longitude and latitude might deviate from expectations (e.g., from the given longitude and latitude, it might look as if the AR is tracking south or west, but this is not the case). We expect any analysis performed on these locations to include some room for error. The locations we have pinpointed are good starting points, but of course do not capture the true scale of these events. Sometimes our best but still subjective judgement was used. These are highly dynamic systems, and they merge with or split from one another. We made notes where ARs track together, merge, or split off from another, but it is entirely possible we have missed some distinctions between events.

3.2 Human control "Counting" for (landfalling) events

An AR event is counted here if its IVT or IWV crosses a coastline. The vertical axis displays time step. Counting is binned along coasts at 1-degree resolution along the horizontal axis of each entered into a counting spreadsheet. Note that IVT and IWV were assessed separately since there is a spatiotemporal lag between IVT and IWV. The western coast of North America from a latitude of 32°N to 55°N, and the western and southwestern coast of Europe from 35°N to 61°N were considered. Each column of the last row of each counting spreadsheet sums all the values in each column. Each sum, therefore, represents the total number of instances where the IVT or IWV associated with an AR event crosses the coastline

at that 1-degree latitude range during the entire examined time period and is the data used for Figure 3.

Deleted: 5

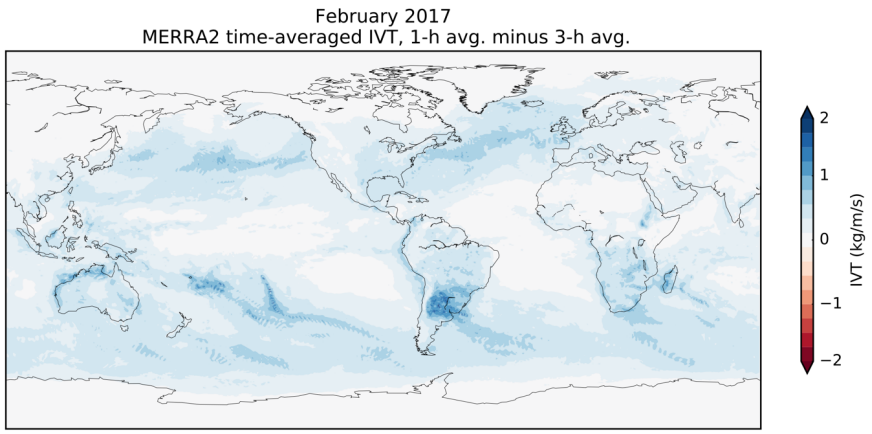


Figure S1. February 2017 MERRA-2 time-average IVT 1-hour versus 3-hour differences. Units are $kgm^{-1}s^{-1}$.

Moved down [3]: Figure S1. February 2017 MERRA-2 time-average IVT 1-hour versus 3-hour differences. Units are $kgm^{-1}s^{-1}$. ¶

Formatted: Normal

Deleted: ¶

Moved (insertion) [3]

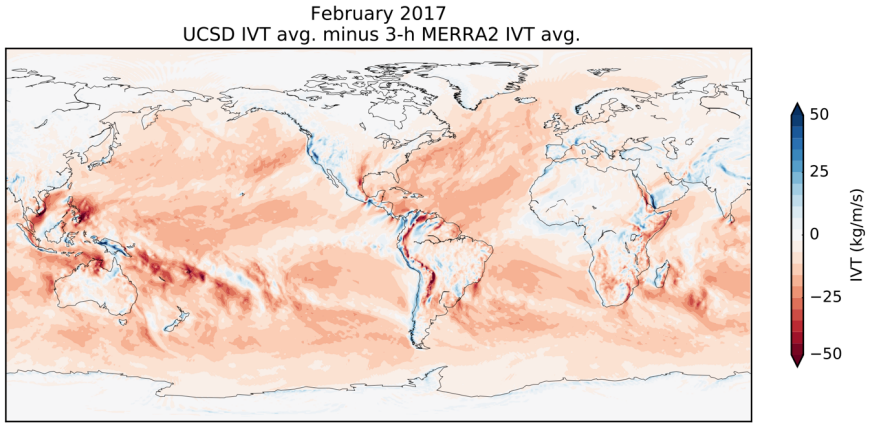


Figure S2. February 2017 MERRA-2 UCSD-computed 3-hourly IVT data versus 3-hour time-averaged data computed form averaging 1-hour MERRA-2 provided IVT differences (as in Figure S1). Units are $kgm^{-1}s^{-1}$.

Moved down [2]: Figure S2. February 2017 MERRA-2 UCSD-computed 3-hourly IVT data versus 3-hour time-averaged data computed form averaging 1-hour MERRA-2 provided IVT differences (as in Figure S1). Units are $kgm^{-1}s^{-1}$.¶

Moved (insertion) [2]

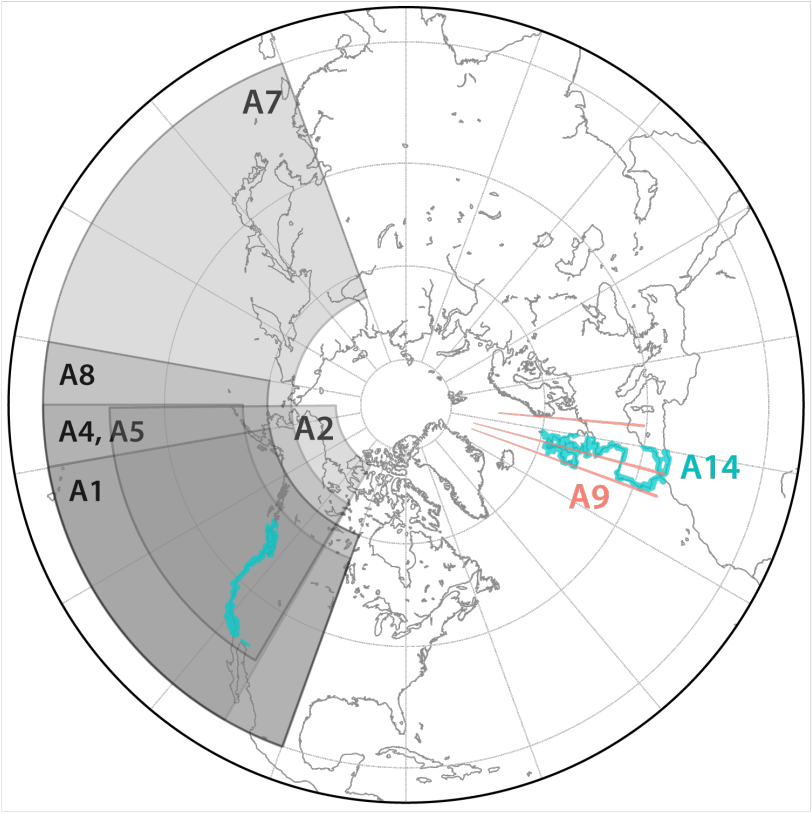


Figure S3. Regional masks provided by ARTMIP 1-month test participants.

Moved down [1]: Figure S3. Regional masks provided by ARTMIP 1-month test participants.¶

Moved (insertion) [1]

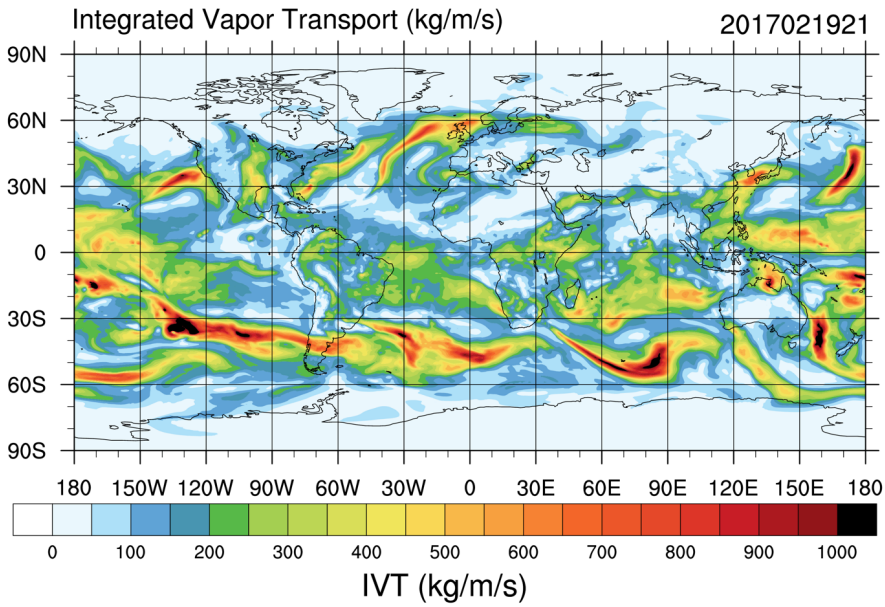


Figure S4. Integrated vapor transport (IVT) (kg/m/s) for a sample time slice in the MERRA-2 dataset. Date stamp in the upper right corner in the format of YYYYDDMMHH, i.e. 2017, February 19th, 21Z. Note the various individual ARs in the Pacific and Atlantic Ocean Basins with landfalling ARs impacting the west coast of the United States, the UK, and mid-latitude coastal regions on the west coast of South America.

Formatted: Superscript