In Search of The Optimal Atmospheric River Index for Precipitation: Big Data Analysis of Index Ensembles over the North American West Coast and US Midwest

Chen Zhang[1], **Wen-wen Tung**[1*], **William S. Cleveland**[2,3]

[1]Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, West Lafayette, IN, USA
[2]Department of Statistics, Purdue University, West Lafayette, IN, USA
[3]Department of Computer Science, Purdue University, West Lafayette, IN, USA

Key Points:

- An optimal AR index depends on the defined precipitation impacts, regional physical mechanisms of precipitation, season, and duration.
- IWV with permissive climate thresholds is suitable for representing broad presence and accumulation of precipitation in regions studied.
- IVT necessarily gets extreme West Coast orographic precipitation in cold seasons; IWV fits all seasons for a wide range of Midwest events.

*Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN 47907, USA

Corresponding author: Wen-wen Tung, `wwtung@purdue.edu`

Abstract

Atmospheric rivers (ARs) affect surface hydrometeorology in western North America and the US Midwest. We systematically sought optimal AR indices for expressing surface precipitation impacts within the prevailing Atmospheric River Tracking Method Intercomparison Project (ARTMIP) framework. We adopted a multifactorial ensemble approach. Four factors—moisture fields, climatological thresholds, shape criteria, and temporal thresholds—collectively generated 81 West Coast AR indices and 81 Midwest indices from 2006 to 2015. Two moisture fields were extracted from the MERRA data for ARTMIP: integrated water vapor transport (IVT) and integrated water vapor (IWV). Global Precipitation Climatology Project One-Degree Daily Precipitation data were used. Metrics for precipitation effects included two-way summary statistics relating the concurrence of AR and that of precipitation, per-event averaged precipitation rate, and per-event precipitation accumulation. The analysis was executed via distributed-parallel computing on a Hadoop cluster using R-based DeltaRho software. We found that an optimal AR index depends on types of impact to be addressed, associated physical mechanisms in the affected regions, timing, and duration. In West Coast and Midwest, IWV-based AR indices identified the most abundant AR event time steps, most accurately associated AR to days with precipitation, and represented the gross presence of precipitation the best. With a permissive climatological threshold, they detected the most accumulated precipitation with the longest event duration. IWV-based indices are the overall choice for Midwest ARs. IVT-based indices suitably captured the accumulation of intense orographic precipitation on the West Coast. Combined IVT and IWV only focused on short records of extreme West-Coast AR precipitation.

Plain Language Summary

[Atmospheric rivers (AR), the long narrow filaments of enhanced water vapor transport in the lower troposphere, are known to accompany extreme rain and winds. They are important weather systems for US water resources on the West Coast and in the Midwest. Currently, there are many AR detection algorithms for creating AR indices. We asked which impacts, in which region, and in what time scale and period were of concern. We then used an approach combining climate significant- or extreme-event criteria, image processing, and statistical analysis to create 81 West Coast AR indices and 81 Midwest indices from 2006 to 2015 for answering the questions with detailed visualization. This approach was made possible by high-performance computing with data across multiple computer servers. We found that an optimal AR index depends on the defined precipitation impacts, regional physical mechanisms of precipitation, season, and duration. Integrated water vapor (IWV) can represent the broad-stroke presence and accumulation of precipitation in regions studied. Combined moisture with wind fields, namely integrated water vapor transport (IVT), is necessary to get extreme West Coast AR orographic precipitation in cold seasons. On the other hand, IWV well represents moderate to extreme Midwest AR precipitation events for all seasons. Combination of IVT and IWV is useful to get snapshots of extreme precipitation events.]

1 Introduction

Atmospheric rivers (ARs) are long, narrow filaments of enhanced water vapor transported from the tropics to the higher latitudes in the lower troposphere (Zhu & Newell, 1998). When these moisture-laden ARs make landfall or penetrate inland, water vapor condenses and can release enhanced precipitation (e.g., Guan et al., 2010, 2013; Luo & Tung, 2015). AR precipitation in many parts of the world is paramount for water resources (e.g., Guan et al., 2010; Dettinger et al., 2011; Rutz & Steenburgh, 2012; Dettinger, 2013; Lavers & Villarini, 2015). However, heavy rainfall can lead to floods and ensuing socioeconomic damage. Studies have shown that in North America, ARs have significant surface hydrometeorological effects on the western North America (e.g., Ralph et al., 2006; Neiman et al., 2008; Leung & Qian, 2009; Ralph et al., 2011; Dettinger, 2011; Rutz et al., 2014) and the US Midwest (e.g., Lavers & Villarini, 2013; Nayak & Villarini, 2017).

The first and critical task to study ARs is to develop AR identification methods. There have been many AR detection and tracking methods for different purposes in the literature, as noted in the Atmospheric River Tracking Method Intercomparison Project (ARTMIP, Shields et al., 2018; Rutz et al., 2019; O'Brien et al., 2020). These different detection methods are primarily based on either one or both measurements of Integrated Water Vapor (IWV) and Integrated Water Vapor Transport (IVT).

Ralph et al. (2004, 2005, 2006) created an objective AR identification method using satellite-based IWV for case studies in the North American West Coast. They defined ARs with IWV content > 20 mm, length > 2000 km, and width < 1000 km. Similar approaches have since been widely applied (e.g., Neiman et al., 2008; Wick et al., 2013). Furthermore, IVT derived from reanalysis or models incorporates the effects of advection. Zhu and Newell (1998) first defined ARs through IVT. Lavers et al. (2012) and Lavers and Villarini (2013), respectively, established percentile-based IVT thresholds to study ARs affecting Britain and Central US. Guan and Waliser (2015) applied 85th percentile seasonal climatological thresholds to IVT for global AR detection. Meanwhile, Rutz et al. (2014) used absolute thresholds, preferring $IVT \geq 250$ kg m^{-1} s^{-1} to $IWV \geq 20$ mm as a threshold to emphasize inland-penetrating ARs in the Western US.

IVT-based detection method is increasingly chosen over IWV-based ones in research and operation as horizontal moisture transport is qualitatively related with orographic precipitation (e.g., Neiman et al., 2009; Rutz et al., 2014; Guan & Waliser, 2015). The combination of IVT and IWV (IVT+IWV thereafter) was recently adopted (e.g., Eiras-Barca et al., 2016; Gershunov et al., 2017). The $IVT + IWV$ method was proposed to reduce tracking errors from considering only one of the measurements. It requires both IVT and IWV values to meet their corresponding thresholds simultaneously.

Furthermore, the duration of an AR is important for its hydrometeorological effects. Longer-lived ARs are more likely to bring higher rainfall (in total and on average) and streamflow than shorter-duration ones (Ralph et al., 2013; Nayak & Villarini, 2018). However, there has not been a consensus in duration criteria. Duration thresholds were not required in some early case studies (e.g., Ralph et al., 2004). Subsequently, a minimum of at least 8 (Ralph et al., 2013), 12 (Payne & Magnusdottir, 2016), 18 (Lavers et al., 2012; Lavers & Villarini, 2013; Nayak & Villarini, 2017; Gershunov et al., 2017), or 24 consecutive hours (Sellars et al., 2015) were included as a part of detection algorithms.

Although systematic comparisons among different AR identification methods are underway (Shields et al., 2018; Rutz et al., 2019; Ralph et al., 2019), the relationships between the identified ARs and precipitation remain to be quantified. Important questions to ask include: between the two common detection measurements of IVT and IWV, which one, or both, should be used when surface precipitation is concerned? How do more restrictive duration criteria perform if long-lived ARs produce larger amounts of precipitation than short-lived ones (Ralph et al., 2013)? In probing these questions, we attempted to establish an optimal AR detection algorithm suited for expressing surface precipitation impacts. We used a multi-factorial ensemble analysis, well suited for uncertainty quantification, within the ARTMIP framework of prevailing detection methods and reanalysis data from January 2006 to December 2015. The paper is organized as follows: data and methods are in section 2. Surface precipitation effects associated with different AR detection indices are analyzed and discussed in section 3. Sections 4 and 5 provide discussions and conclusions, respectively.

2 Data and Methods

2.1 Data

2.1.1 MERRA data for ARTMIP

The two conventional moisture measurements for AR detection, IVT and IWV, were extracted from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) source data for ARTMIP through Climate Data Gateway (NCAR CDG, 2019). This dataset was calculated by the Center for Western Weather and Water Extremes at the University of California, San Diego, according to the following formula (Shields et al., 2018):

$$IVT = -\frac{1}{g}\int_{1000}^{200} q(p)|\mathbf{V}_{\mathrm{h}}(p)|\, dp, \tag{1}$$

$$IWV = -\frac{1}{g}\int_{1000}^{200} q(p)\, dp \tag{2}$$

The three variables, horizontal wind ($\mathbf{V}_{\mathrm{h}} = (u, v)$ where u is the zonal and v the meridional winds in m s^{-1}), specific humidity (q in kg kg^{-1}), and pressure (p in hPa), used in the formula were from NASA MERRA-2 (Gelaro et al., 2017). The horizontal spatial resolution and temporal resolution of the vertically integrated fields are 0.5° longitude by 0.625° latitude and 3 hours. We used all of the MERRA-2 Tier 1 data available at the time of download, from January 1980 to June 2017, to create climatological thresholds. Then, we applied the AR detection algorithm to the subset of the recent decade, January 2006 to December 2015, to generate AR indices.

2.1.2 GPCP Version 1.3 One-Degree Daily Precipitation Data Set

Global Precipitation Climatology Project (GPCP) One-Degree Daily Precipitation Version 1.3 (Huffman et al., 2001) provides daily precipitation estimates at a global 1.0°

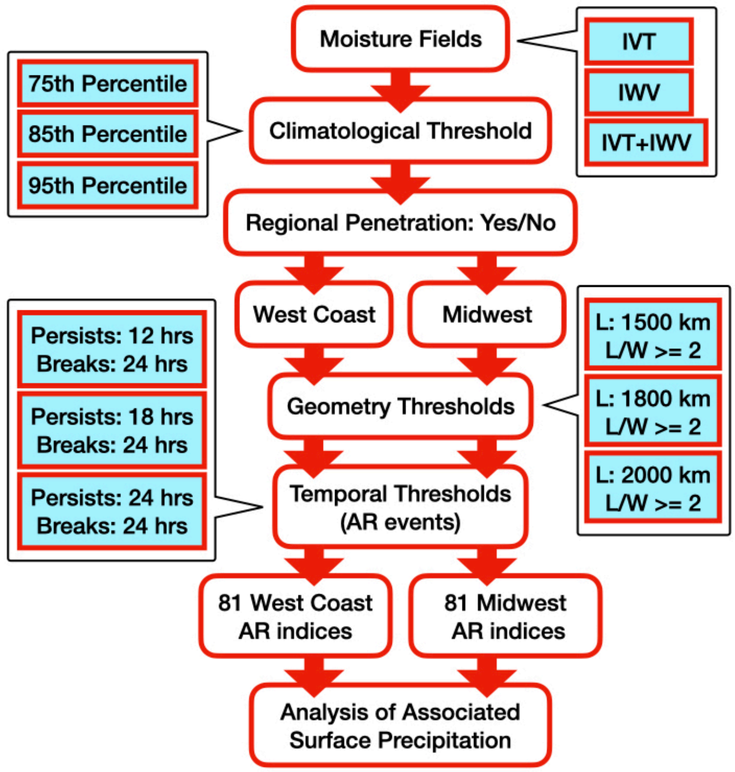

Figure 1. Schematic diagram illustrating the AR detection algorithm.

by 1.0° resolution from 1 October 1996 to the delayed present. We used GPCP Version 1.3 downloaded from NCAR Research Data Archive (NCAR RDA, 2018) to investigate the surface precipitation effects of the AR indices. These AR indices, as described in the next subsection, were defined by various AR detection criteria applied to the ARTMIP MERRA-2 data. We applied bilinear interpolation to the GPCP data to match their spatial resolution to that of the identified ARs. In order to match the time steps of the AR indices, each GPCP daily precipitation measurement was divided and evenly distributed over the twenty-four hours within a day.

2.2 AR Detection Algorithm

As shown in Figure 1, we used 4 factors—moisture fields, climatological thresholds, shape criteria, and temporal thresholds—to generate an ensemble of 81 AR indices for the US West Coast and 81 for the Midwest. First, we used IVT, IWV, or $IVT{+}IWV$ as the moisture field. Then, for each grid point, we selected moisture field values at 1200 UTC every day during neutral or weak El Niño–Southern Oscillation (ENSO) events from

January 1980 to June 2017. We called these test values. Here, we adopted the bi-monthly NOAA Multivariate ENSO index (MEI.v2, e.g., Wolter & Timlin, 1993) and preserved only test values in the months when the MEI.v2 index was within ±1. Three monthly climatological thresholds were calculated for each set of test values—IVT, IWV, or $IVT+IWV$—at each grid point. In addition to the common 85th percentile (e.g., Lavers et al., 2012; Lavers & Villarini, 2013; Guan & Waliser, 2015; Eiras-Barca et al., 2016), we also used the 75th and the 95th percentiles as thresholds. Consequently, at any given grid point and time, a moisture value equal to or exceeding a threshold suggests the potential presence of AR.

Figure 2 plots these three levels of climatological thresholds of IVT and IWV fields over North America for January and August. The threshold at each grid point elevates successively, increasingly restricting AR detection, from 75th to 95th percentile. The IVT maxima corresponded to extratropical storm tracks and ITCZ over the North Pacific and the North Atlantic. The IWV maxima co-located with tropical and extratropical warm oceans as well as maritime tropical air mass. Consistent with Clausius–Clapeyron equation, IVT and IWV thresholds were generally higher in the summer (August) than in the winter (January). The regions of the West Coast (32°–52.5° N, 124.375°–114.375° W) and Midwest (37°–47° N, 94°–84° W) are outlined in red. This seasonal difference was more evident in the Midwest than the West Coast. Regardless, the IVT maximum over the Northeast Pacific Ocean expanded towards the West Coast in January, then retreated in August.

Figure 2 also compares the monthly percentile thresholds in this study against the absolute thresholds used by Gershunov et al. (2017), defined as 250 kg m^{-1}s^{-1} in IVT and 15 kg m^{-2} in IWV. The monthly percentile thresholds exhibit more spatial details as well as seasonal variability than the absolute ones. Through visual inspection, one can infer the different outcomes of AR detection if solely based on these thresholds. In January, absolute IVT and IWV thresholds resulted in fewer instances of landfalling ARs in both West Coast and Midwest compared with the respective 75th and 85th percentile thresholds. This is because the majority of the monthly percentile threshold values in these two regions are below the absolute thresholds. The absolute thresholds, however, permitted more frequent January AR detection along the West Coast and southern Midwest than the 95th percentile threshold values. These were also true for West-Coast AR detection using IVT in August; however, in the Midwest, the absolute IVT threshold

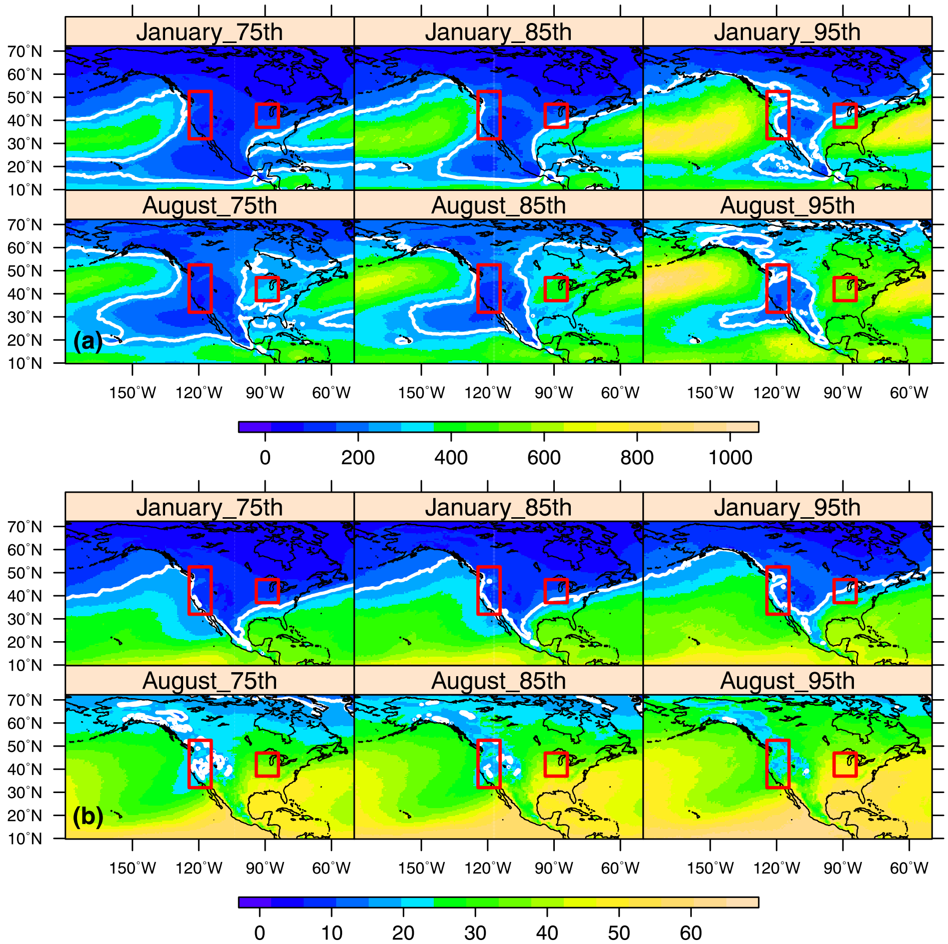

Figure 2. Three levels of climatological thresholds of IVT (a, kg m^{-1}s^{-1}) and IWV (b, kg m^{-2}) over North America for January and August derived from neutral or weak ENSO events between January 1980–June 2017. The red boxes outline the West Coast and Midwest regions in this study. White lines (250 kg m^{-1}s^{-1} in IVT and 15 kg m^{-2} in IWV) are the absolute thresholds in Gershunov et al. (2017).

was more permissive to August AR detection than all IVT monthly percentiles. The absolute IWV threshold allowed overall more August AR detection than the monthly percentiles in the plotted domain.

The variability of AR detection across various thresholds above attests to the necessity of additional constraints. At each time step from 2006 to 2015, we identified the grid points whose IVT or IWV values exceeded their corresponding climatological thresholds and kept only the data of those making landfall in the West Coast or penetrating into the Midwest. Then, we used the principal curves method (Hastie & Stuetzle, 1989) to determine the length of the curvy patterns formed by aggregating the maximum IVT or IWV values at each latitude and longitude. The width was calculated as the total Earth surface area of the identified grid points divided by the length. The geometry thresholds were further applied. A subset of potential AR data was extracted if a length was greater or equal to 1500, 1800, or 2000 km while the ratio of length to width was greater or equal to 2 (Figure 1). It is noted that, for the detection of West Coast land-falling ARs, this length was estimated using only the segment of data over the Pacific; for the ARs penetrating into the Midwest, it was estimated using the entire segment. The subsets of data were further filtered and aggregated into AR events that persisted for equal to or more than 12, 18, or 24 hours with breaks shorter than 24 hours within an event. The length of break criterion was based on Lavers and Villarini (2013).

At this point, 81 members of AR indices for each of the West Coast and Midwest regions from January 2006 to December 2015 were completed. Each index identifies the spatial and temporal information of AR events that satisfied one of the 81 combinations of the criteria form by the four factors. We proceeded with systematic analysis of the relationships between these ARs and surface precipitation in the West Coast and the Midwest. The AR detection and detailed analysis were executed via distributed-parallel computing on a high-performance computing cluster with Hadoop system in the backend and the R language-based DeltaRho software in the frontend (Cleveland & Hafen, 2014; Tung et al., 2018).

2.3 Coarse- to Fine-Grain Two-Way Summary Table

We built a two-way summary table (Figure 3) to explore the relationships between ARs identified by the indices and the surface precipitation in the West Coast and the

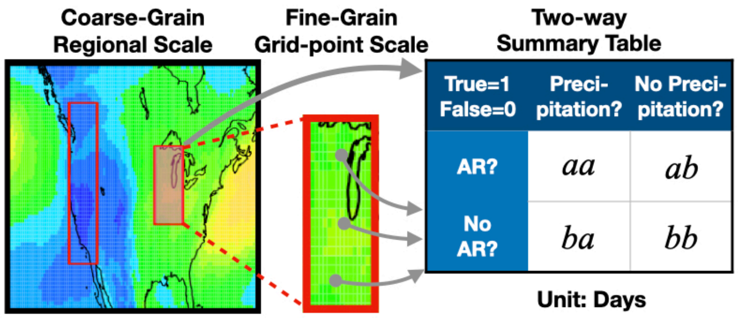

Figure 3. A two-way summary table with the terms for evaluating AR indices' relationships with surface precipitation on coarse- and fine-grain scales.

Midwest. We took two spatial scales into account: regional coarse-grain scale and grid-point fine-grain scale. On the coarse-grain scale, we regarded either West Coast or Midwest as one entity. Within each entity, days with at least one AR time step identified in a 3-hourly AR index were defined as AR days. Days without any AR time steps were considered as no AR days. Precipitation was the spatially-interpolated daily GPCP data. The aa in the summary table was total AR days with precipitation; the ab was AR days without precipitation; the ba was days with no ARs but with precipitation; and the bb was days with no ARs and no precipitation.

From the summary table, four statistics were derived: *AR Related Precipitation*, *Precision*, *Accuracy*, and $F1$ *score*. The names loosely follow those in statistical classification (Hastie et al., 2001). They were used to compare the AR indices' performance of relating to surface precipitation effects. *AR Related Precipitation* is defined as

$$\frac{aa}{aa+ba} = \frac{aa}{D_P}, \tag{3}$$

with D_P the total days with precipitation. It specifies how often surface precipitation, if existed, was related to the ARs identified by an index. *Precision* is defined as

$$\frac{aa}{aa+ab} = \frac{aa}{D_{AR}}, \tag{4}$$

with D_{AR} the total days with ARs according to an index. It describes how often the detected ARs were actually related with precipitation. *Accuracy* is defined as

$$\frac{aa+bb}{aa+ab+ba+bb} = \frac{aa+bb}{D}, \tag{5}$$

with D the total 3652 days in the data. For each AR index, it measures how often days with/without ARs were correctly associated with precipitation/no precipitation. The $F1$

score,

$$\frac{2 * AR\ Related\ Precipitation * Precision}{AR\ Related\ Precipitation + Precision}, \tag{6}$$

is the harmonic mean of *AR Related Precipitation* and *Precision*. An AR index with a low $F1$ *score* has both poor *AR Related Precipitation* and poor *Precision*, therefore an overall poor AR-precipitation relation.

Each of these four statistics had one resultant value for each index on the coarse-grain scale in either West Coast or Midwest. On the fine-grain scale, they were multiplied by the number of grid points inside a region: 714 in the West Coast and 336 in the Midwest. The different sample sizes were taken into account in interpreting the results (section 3.2).

3 Analysis and Results

3.1 Identified AR occurrence summary statistics

Figures 4 and 5 visualize three summary statistics of AR occurrence obtained with 162 AR indices. These figures are Cleveland dotplots (Cleveland & McGill, 1984) created in the Trellis display framework (Becker et al., 1996). The number of AR events (Figure 4a), the accumulated time of these events measured in 3-hourly time steps (Figure 4b), and the average duration per event in days (Figure 5) are plotted on each panel, conditional on 18 combinations of regions (West Coast or Midwest), moisture fields (IVT, IWV, or IVT+IWV), and AR length criteria (1500, 1800, or 2000 km). The results are 18 packets, or subsets, of values. Each packet has 9 paired values of a summary statistic in the y-axis and one of the AR persistent duration thresholds (12, 18, or 24 hours along the x-axis), grouped with color by climatological thresholds (75th, 85th, or 95th percentiles).

Figure 4a shows that, from 2006 to 2015, each IVT- or IWV-based AR index captured $O(100)$ events in either West Coast or Midwest regions, except for a few IVT-based ones with the most restrictive combinations of length and duration criteria in the West Coast. In Figures 4b and 5, IWV-based indices identified the most AR time steps and longest average per-event duration; IVT+IWV-based indices identified the least and the shortest. Increasing the restrictiveness of climatological threshold from 75th to 95th

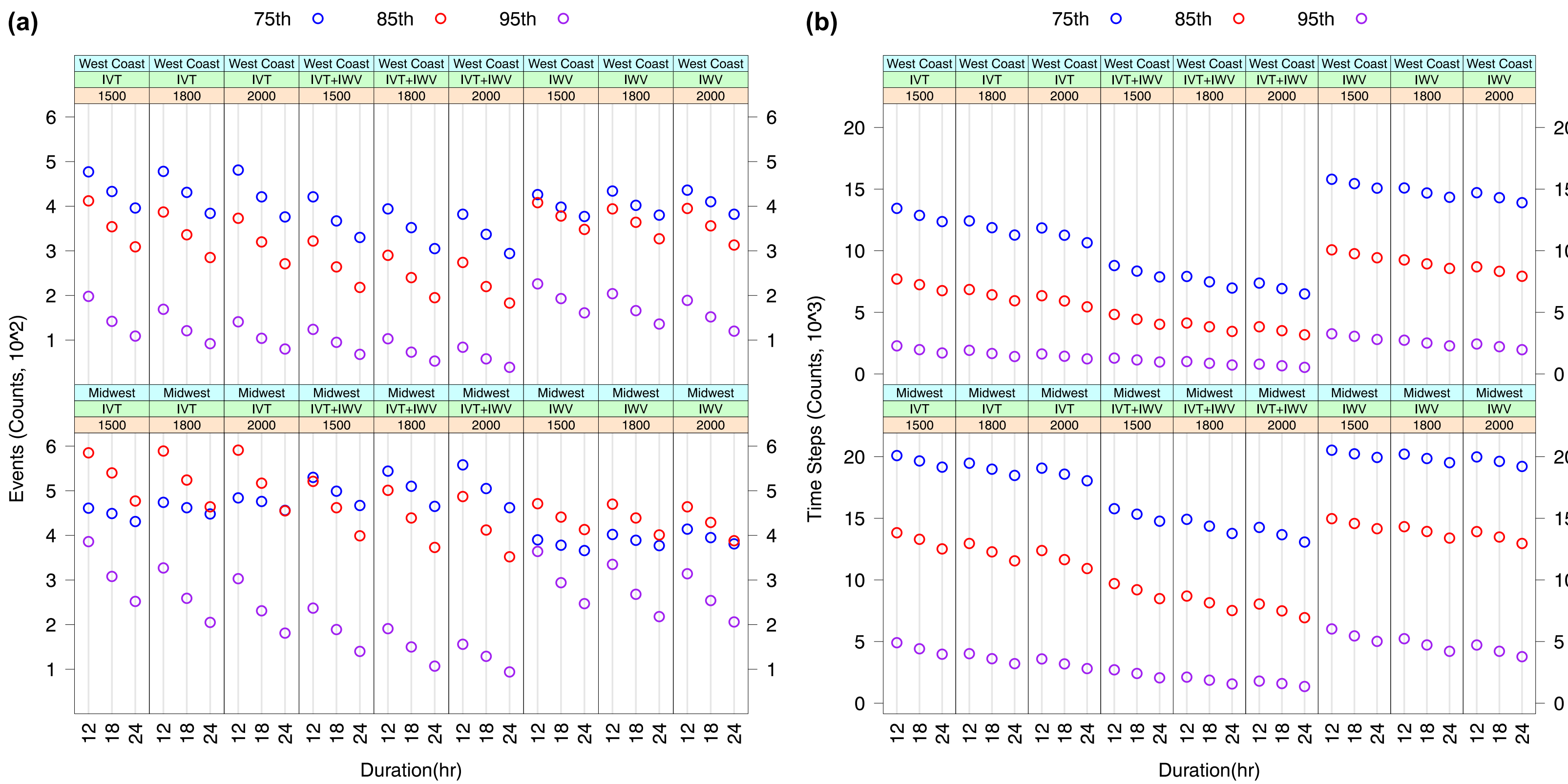

Figure 4. Numbers of (a) AR events (counts) and (b) accumulated AR time steps (counts) identified by 81 West-Coast (top row) and 81 Midwest (bottom row) AR indices.

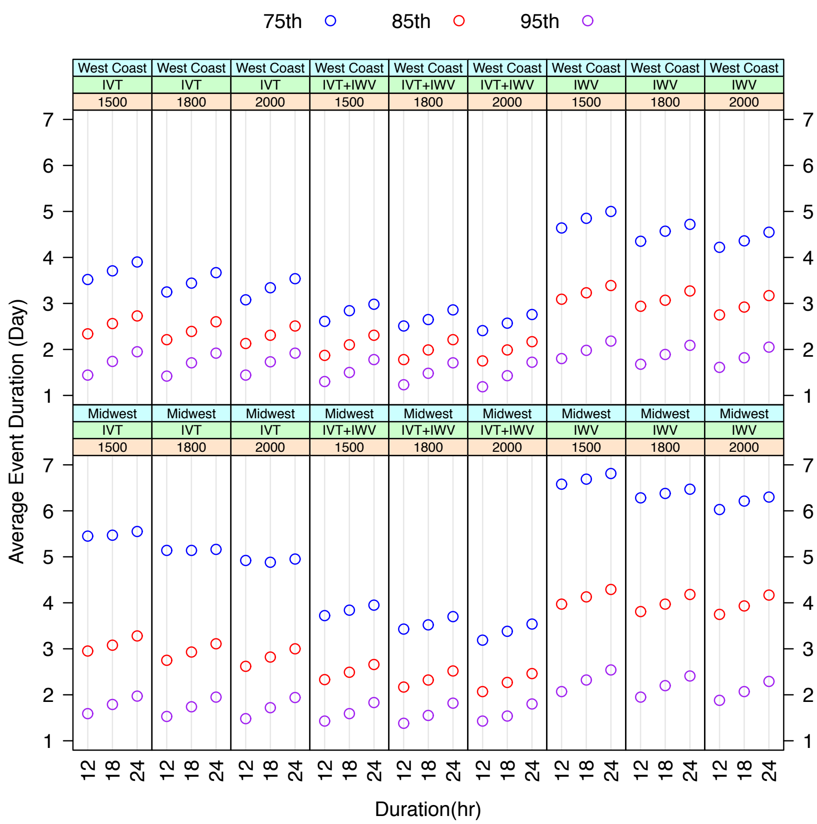

Figure 5. Average per-event duration (unit: Day) identified by 81 West-Coast (top row) and 81 Midwest (bottom row) AR indices.

percentile while holding other factors constant, the number of identified AR time steps decreased dramatically, so did the average duration per event.

However, more restrictive climatological thresholds did not always yield fewer AR events (Figure 4a). Among IVT- and IWV-based Midwest AR indices, the 85th percentiles permitted more AR events but fewer time steps than the 75th percentiles, owing to the latter's tendency to yield longer per-event durations (Figure 5). Furthermore, the identified Midwest ARs had overall 1.5-2 times the total time steps of West-Coast ARs from 2006 to 2015 (Figure 4b). Midwest ARs had longer average per-event durations than those in the West Coast; the differences were the largest at the 75th percentiles and the least at the 95th percentiles (Figure 5).

In Figures 4 and 5, the effects of length criteria were only secondary to climatological thresholds. However, increasing the thresholds of AR persistent duration from 12 to 24 hours resulted in shorter accumulated time steps (Figure 4b) and, in most cases, longer average per-event duration (Figure 5) of the identified ARs. It also led to decreasing AR event counts (Figure 4a).

3.2 Coarse- to fine-grain daily AR-precipitation occurrence relation analysis

3.2.1 Coarse-grain analysis

Figure 6a shows the coarse-grain *Accuracy* in dotplots. Ideally, detected ARs should represent the precipitation occurrence as complete and correct as possible. An *Accuracy* of 1 means there was precipitation if and only if ARs were detected by an index. In general, indices associated with more AR time steps (Figure 4b) also exhibited higher *Accuracy* at the coarse-grain scale. Indeed, Midwest ARs bore higher *Accuracy* than the West Coast ARs given otherwise the same factors. The IWV-based AR indices yielded the highest *Accuracy* in both regions. Among them, indices using the 75th percentile climate threshold had *Accuracy* exceeding 0.7 in the Midwest. More restrictive climatological thresholds resulted in lower *Accuracy*. The lowest values were within the 95th-percentile-based IVT+IWV indices—below 0.08 for the West Coast ARs. More restrictive length and temporal criteria that detected fewer AR events or time steps also depressed *Accuracy* values, while the effect of length was minor in comparison to other factors.

Figure S1 shows the *AR Related Precipitation*, *i.e.,* the fraction of total days with precipitation attributable to identified ARs. It has a similar pattern to Figure 6a. In particular, when 75th-percentile IWV-based indices were used, more than 58% and 77% of precipitation days occurred in the presence of ARs in the West Coast and Midwest, respectively. However, 95th-percentile $IVT + IWV$-based indices could only capture less than 8% and 19% of precipitation days in the respective regions. On the other hand, *Precision* values in Figure S2 display a very different pattern from Figures 6a or S1. For the West Coast landfalling ARs, 63 out of 81 indices had *Precision* equal to 1, with the rest approximately 1. That means each index very precisely associated AR days with precipitation. For ARs influencing the Midwest, the *Precision* values were slightly smaller but still larger than 0.975.

The $F1$ *score*s in Figure 6b summarizes for each AR index the combined performance of relating to the presence of precipitation (*Precision*) and explaining the occurrence of precipitation (*AR Related Precipitation*) at the coarse-grain scale. Unlike *Accuracy*, $F1$ *score* does not consider days with no AR and no precipitation, expressed as the *bb* term in (5). In practice, we are more concerned about the relationship between the presence of AR and that of precipitation than the absence of both. Therefore, $F1$ *score* is a more sensible measurement than *Accuracy*. Furthermore, the score could be considered as adjusted *Precision*, with which indices gained high *Precision* via narrowing to extreme samples are penalized. The adjustment differentiated the overall high *Precision* values (Figure S2) to the pattern of $F1$ *score*s (Figure 6b), which resembles Figures 6a and S1 but have larger magnitudes across the board.

3.2.2 Fine-grain analysis

We established for each index a two-way summary table for each individual grid point in West Coast and Midwest for fine-grain analysis. The distributions of fine-grain $F1$ *score*s are summarized using boxplots for the 81 West Coast AR indices, each with 714 points (Figure 7a) and 81 Midwest indices with 336 points (Figure 7b).

In Figure 7a, the interquartile ranges (IQR) of the 81 $F1$ distributions, as indicated by the box lengths, vary from $\sim$0.03 to $\sim$0.1 for the West Coast AR indices. Spatial inhomogeneity of precipitation captured by different indices contributed to this variation. Another important influencer was the different AR days, D_{AR}, as inferred by the AR

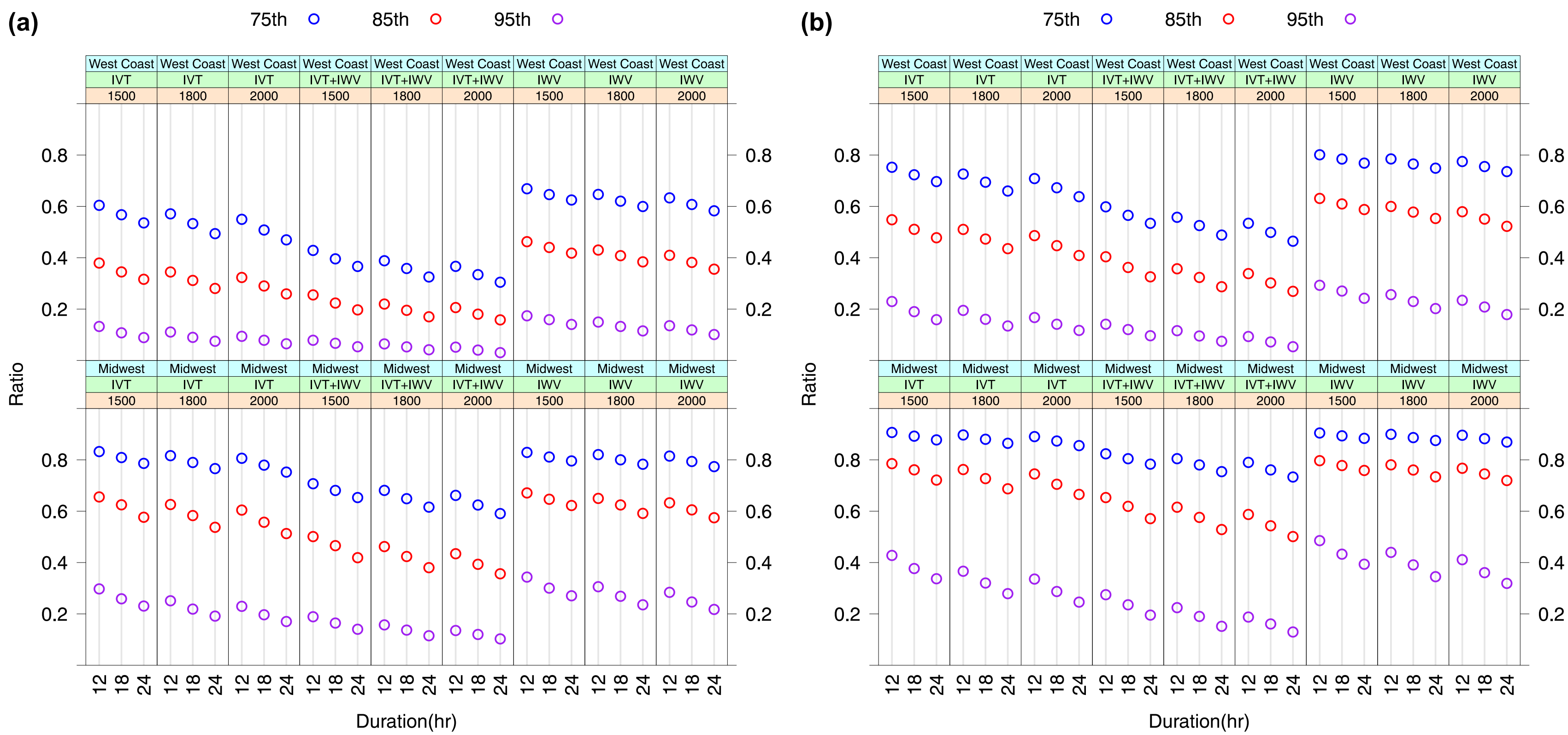

Figure 6. Dotplots of coarse-grain (a) *Accuracy* and (b) *F*1 *scores* for 81 West-Coast (top-row) and 81 Midwest (bottom-row) AR indices.

time steps (Figure 4b), resulted from different indices. Indeed, the smaller IQRs are seen among the most restrictive indices with the fewest AR time steps, such as the 95th-percentile $IVT{+}IWV$-based ones. Moreover, the minimum, first quartile (Q1), second quartile/median (Q2), third quartile (Q3), and maximum of each subset of $F1$ *score*s decrease with more restrictive criteria. This is consistent with the coarse-grain analysis (Figure 6b). When the climate threshold, length, and time criteria were fixed, the IWV-based indices slightly outperformed IVT-based ones and were significantly better than $IVT + IWV$-based ones. The 75th-percentile IWV-based indices yielded the largest median $F1$ *score*s, all exceeding 0.5.

The IQRs of fine-grain $F1$ *score* distributions for the 81 Midwest AR indices (Figure 7b) are smaller than those for West Coast AR indices (Figure 7a). This is most certainly due to the ∼50% smaller sample size in the Midwest than that of the West Coast. The differences among the $F1$ *score* distributions in the Midwest are qualitatively similar to those in the West Coast. Nevertheless, the $F1$ *score*s in the Midwest are overall higher. The 75th-percentile IWV-based indices struck the highest median $F1$ *score*s at ∼0.7. These are consistent with the coarse-grain $F1$ analysis (Figure 6b).

In both West Coast and Midwest, the IQRs of *Accuracy* values (Figures S3 and S4) are larger than those of the $F1$ *score*s (Figure 7). This is attributed to the effects of no AR and no precipitation days, *bb*, in the calculation of *Accuracy* (5). Regardless, IWV-based indices had slightly higher median *Accuracy* compared to IVT and $IVT{+}IWV$-based indices.

3.3 Deep Analysis at the Finest Granularity

In section 3.2, we studied the presence or absence of ARs in relation to those of precipitation, as reflected by the ensembles of indices in the North American West Coast and the US Midwest. Past studies consistently showed that in general, ARs contributed to a fair amount of annual precipitation—up to 50% depending on the location—in the contiguous United States (Dettinger et al., 2011; Rutz & Steenburgh, 2012; Lavers & Villarini, 2015; Nayak & Villarini, 2017). Hence, in the next step, we analyzed the amount of AR-related precipitation associated with different indices. We quantified precipitation impacts with event-average rate (3.3.1) and event-accumulated precipitation (3.3.2 and 3.3.3) and compared them across the AR indices.

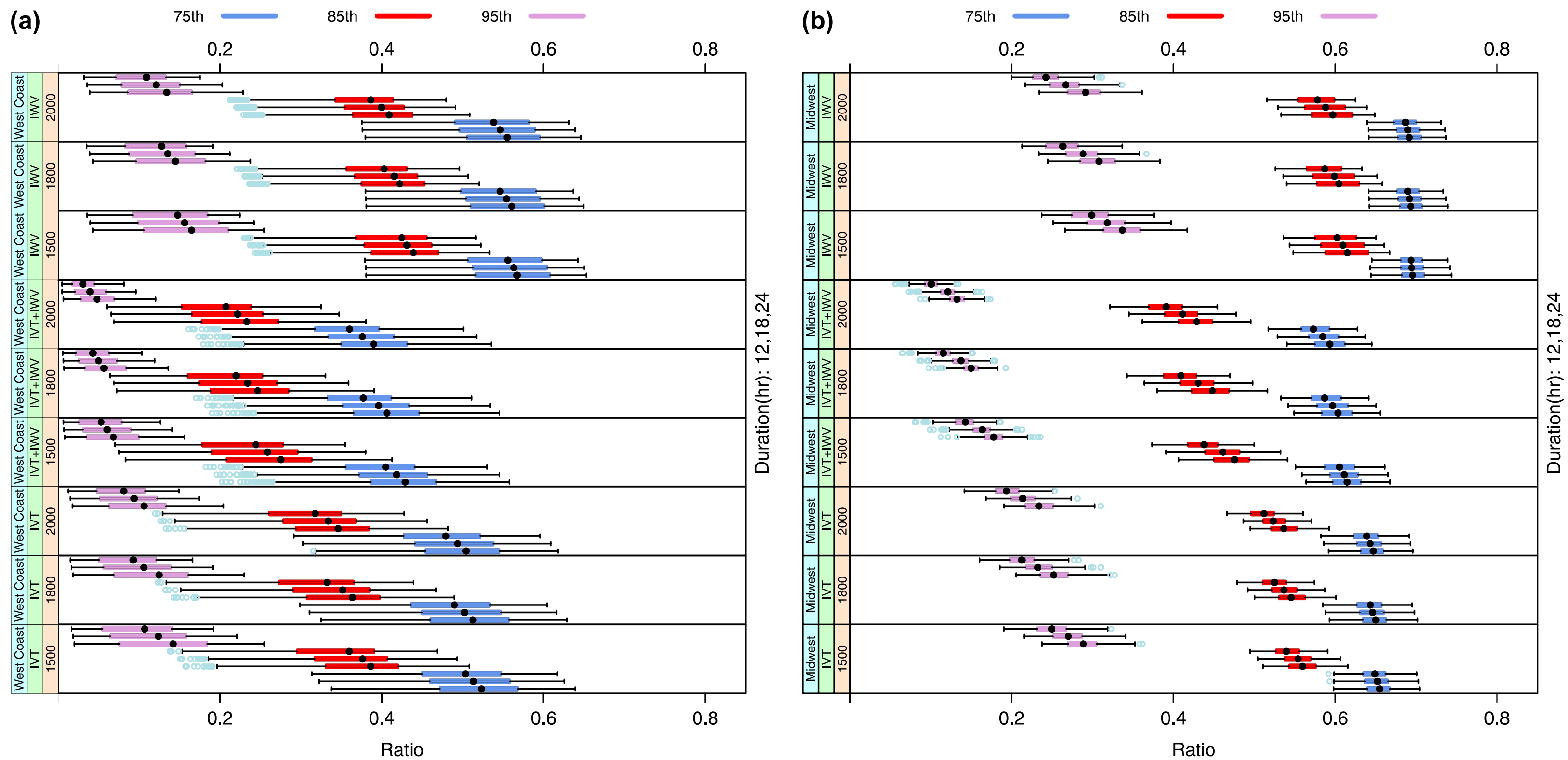

Figure 7. Boxplots of fine-grain $F1$ *score*s for the (a) West Coast and (b) Midwest AR indices. Each figure has nine packets from combinations of three moisture and three AR length (km) criteria. Each packet has nine boxplots grouped by color into three levels of climatological thresholds. Within each triplet, from bottom to top, the persistent duration thresholds increase from 12, 18, to 24 hours. Each boxplot includes the colored box spanning from Q1 to Q3 of the distribution, a black dot marking the median, and the whiskers. The whiskers extend to the most extreme data point that is no more than 1.5 times the length of the box (IQR) away from the box. Any data points outside the whiskers are marked as potential outliers in light blue.

3.3.1 Event-Average Surface Precipitation Effects

For each AR index, we tracked the surface area of an AR at each recorded time step. We then calculated the areal-averaged surface precipitation rate at each time step. The event-average surface precipitation rate was calculated as the event time-mean of areal averages. As an example, Figure 8 compares the event-average precipitation rate across a group of AR indices with the 1500-km length and 18-hr persistent duration criteria using boxplots, conditional on locations, climatological thresholds, and moisture fields. The values of precipitation rates shown are the original values plus one and transformed with base-2 logarithm to accommodate the wide range.

All indices for Midwest ARs in Figure 8 were prone to associate with more event-average precipitation than those for the West Coast ARs. As the climatological thresholds on moisture fields became increasingly more restrictive, the indices pointed to heavier event-average precipitation rates. One conspicuous feature in Figure 8 is that $IVT+IWV$-based indices are the strongest performer in both regions. As already shown in section 3.2, the combined moisture field posed the most restrictive criterion, detecting the fewest events with the shortest lifespan per event. The analysis further shows its propensity to crop out AR features with the highest precipitation rates. This is consistent with previous studies (Neiman et al., 2008; Nayak & Villarini, 2018).

Another distinct feature in Figure 8 is the disparate performance of IVT-based indices between the West Coast and the Midwest. IVT-based AR indices were associated with higher event-average precipitation in the West Coast than IWV-based ones. However, this was not the case in the Midwest. This difference is likely due to the orographic origin of precipitation on the West Coast. Compared with IWV, the horizontal transport of moisture expressed by the IVT better indicated the vertical lifting and condensation processes upon convergence at the coastal mountains' windward side. Notably, the 95th percentile IVT-based West Coast AR index captured the intense orographic precipitation that IWV missed.

The effects of shape and temporal criteria on the detected ARs' relations to event-average surface precipitation rate were inconclusive across different climatological thresholds and moisture fields (Figures S5 and S6). With the 75th percentile thresholds, more restrictive persistent duration criteria, such as 24 hours, appeared to be associated with more average precipitation. However, with the 95th percentile thresholds, more permis-

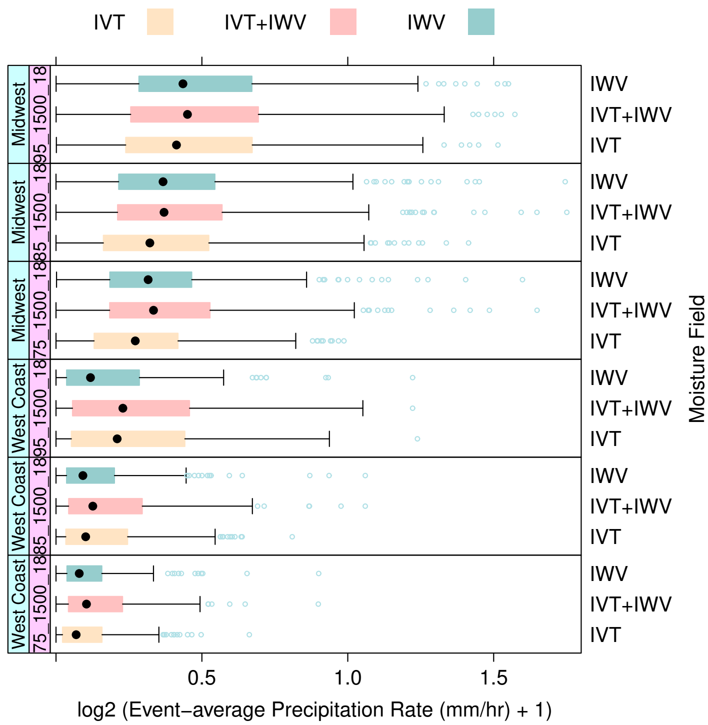

Figure 8. Boxplots of base-2 logarithmic transformation of event-average precipitation rate plus 1 (in mm hr^{-1}) over unit area.

sive duration criteria tended to identify more average precipitation. Still, the climatological thresholds and moisture fields had the first-order influences on the event-average surface precipitation rate.

3.3.2 Deep Analysis of Accumulated Precipitation at Fine Granularity

Although the event-average surface precipitation is a useful metric for an AR index's overall precipitation intensity, it is even more indicative of an AR's hydrometeorological impact when combined with total event duration. Therefore, we further quantified such hydrometeorological impact using event-accumulated precipitation averaged inside a surface area swept by a detected AR. We defined, for each AR index, this area with all grid points visited at least once by the detected AR throughout its lifetime within the West Coast or Midwest region (shown in Figure 9). Given this area, we calculated the areal average of precipitation at each time step, then summed through all time steps to obtain event-accumulated precipitation for the AR event.

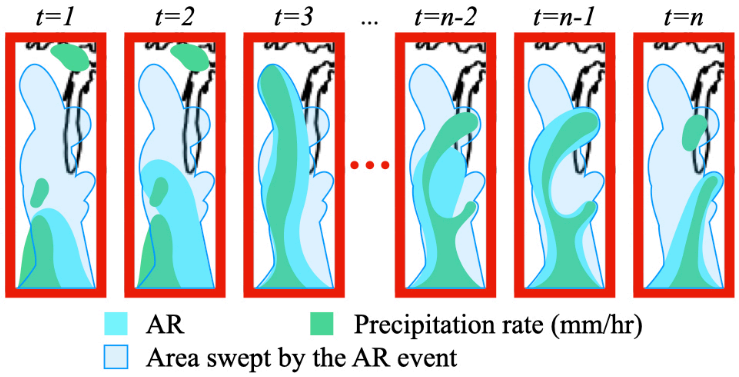

Figure 9. Schematic interpretation of spatial-averaged granule-level AR event-accumulated precipitation.

Figures S7 and S8, respectively, show the swept-area distributions resulted from West Coast and Midwest AR indices. The area of the West Coast region is about twice that of the Midwest region, as shown by the data upper bounds in these figures. As expected, these areas decreased with increasing climatological thresholds; the areas increased with more restrictive persistent duration thresholds; $IVT{+}IWV$-based indices restricted the areas to the smallest among all moisture fields, other factors being equal. Using the 75th percentile climatological thresholds, IWV-based indices tended to sweep a slightly broader area than IVT-based ones. In the Midwest region, the median areas of the 75th-percentile IWV-based AR indices were identical to the area upper bound; at least 50%—but fewer than 75%—of the AR events covered the entire Midwest region. The 75th-percentile IVT-based AR indices had median areas smaller than but very close to this upper bound. However, the areal differences between IWV- and IVT-based indices diminished at 95th percentile thresholds.

Figure 10a compares the event-accumulated precipitation per unit area, plus one and transformed with base-2 logarithm, across the 81 West Coast AR indices using boxplots. The IQRs straddle one order of magnitude, with Q2s at ∼3–7 mm and Q3s reaching as high as ∼15 mm. The climatological and persistent duration thresholds affected the resultant accumulated precipitation the most. We see that the more restrictive duration thresholds retained higher accumulated precipitation events when other factors were fixed. The effects of changing the climatological thresholds, however, are not as simple.

The AR indices based on the 75th percentile IWV performed as well as, if not better than, any other 75th percentile indices in the West Coast region. Increasing the climatological threshold of IWV beyond this point decreased accumulated precipitation (Figure 10a). Since the area swept by the ARs decreased (Figure S7) and the event-average precipitation likely increased (e.g., Figure 8), the shorter event duration (Figure 5) was responsible for this decline in accumulated precipitation. However, among the IVT- and IVT+IWV-based indices, increased climatological thresholds resulted in increased event-accumulated precipitation (Figure 10a). Even so, the event duration decreased (Figure 5). Again, this could be attributed to the orographic effect on intense precipitation, a prominent influencer of accumulated precipitation retained by IVT and IVT+IWV but missed by IWV with restrictive climatological thresholds. IVT's prowess in capturing the accumulated precipitation stands out with the 95th-percentile threshold, considering that 95th-percentile IVT- and IWV-based indices swept over similar sizes of areas (Figure S7), and IVT indices tended to have shorter event duration than IWV ones (Figure 5).

Figure 10b compares the accumulated precipitation across the 81 Midwest AR indices using boxplots. In general, detected Midwest ARs tended to bring twice the amount of event accumulated precipitation than the West Coast ARs. The Q2s, or median values, are at ∼10–20 mm and Q3s extending to ∼30 mm. Similar to the West Coast AR indices, more restrictive persistent duration thresholds led to higher accumulated precipitation. Different from the West Coast, indices based on IWV outperformed those based on IVT or $IVT + IWV$ and resulted in the most accumulated precipitation in the Midwest across all climatological thresholds.

Moreover, increasing the climatological thresholds decreased accumulated precipitation regardless of choices of moisture field. Comparison between Figures 10a and 10b shows that the choice of moisture field affected the detected AR's accumulated precipitation differently by region. AR indices with longer event duration (Figure 5) tend to be associated with more event-accumulated precipitation in the Midwest, whereas indices with larger event-average precipitation rate (Figure 8) are related to more precipitation accumulation in the West Coast. This strongly suggests that the choice of moisture field for AR indices that best expresses surface precipitation impacts on a geographical region ultimately depends on the physical understanding of the region's precipitation processes.

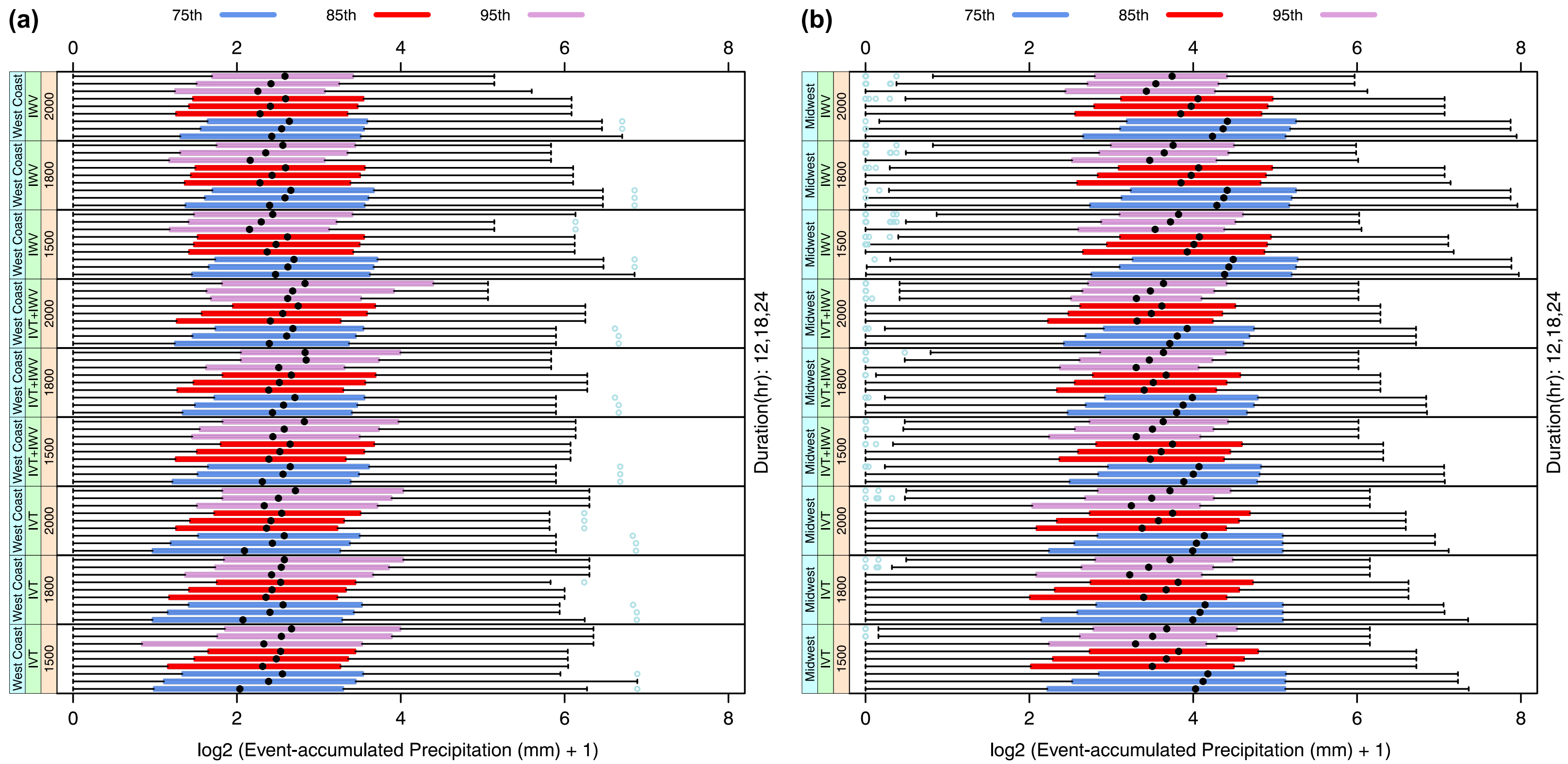

Figure 10. Boxplots of event-accumulated precipitation (mm) over unit area swept by ARs in the (a) West Coast and (b) Midwest. The results are base-2 logarithmic transformation of the original values plus 1, and are conditional on nine combinations of moisture fields and AR length criteria (km). Each resultant packet has nine boxplots grouped by color into three levels of climatological thresholds. Within each triplet, from bottom to top, the persistent duration thresholds increase from 12, 18, to 24 hours.

3.3.3 Seasonal Effects on Event-Accumulated Precipitation

Previous studies have demonstrated the seasonality of AR occurrence (Neiman et al., 2008; Lavers & Villarini, 2015; Nayak & Villarini, 2017). With seasonality as a point of departure, we further examined the event-accumulated precipitation. In particular, section 3.3.2 showed that the climatological threshold and moisture field choices for an AR index significantly affected its resultant accumulated precipitation. Figure 11, therefore, compares the accumulated precipitation across a group of AR indices using boxplots conditional on locations, climatological thresholds, seasons, and moisture fields. For simplicity, only indices with 1500-km length and 18-hr persistent duration thresholds are shown.

Among landfalling West Coast ARs, there was a clear seasonal cycle in the accumulated precipitation that maximized in the winter and minimized in the summer. The phase of this seasonal cycle remained unchanged across all climatological thresholds. This is consistent with the rainy and dry seasons in the West Coast, as well as the previous conclusion that warm seasons had less AR-related precipitation in the West Coast (Neiman et al., 2008). Moreover, the combined effects of climatological threshold and moisture field on the event-accumulated precipitation also had seasonality. In the warm spring and summer, IWV-based indices with the 75th climatological threshold led to the most accumulated precipitation. While in the fall and winter, IVT-based indices with the 95th threshold corresponded with the most precipitation accumulation. This was likely due to the significant orographic enhancement during the landfall of winter ARs but not summer ARs that Neiman et al. (2008) found.

In contrast, among the Midwest ARs, the maximum accumulated precipitation shifted from the warm spring-summer to the winter as the climatological threshold increased, while warm-season accumulated precipitation reduced by about half. This suggests a dichotomy of synoptic systems associated with Midwest ARs: In addition to extratropical cyclones, the warm-month ARs received a significant amount of precipitation from maritime tropical air masses. Unlike in the West Coast, IWV-based Midwest AR indices were associated with the most median precipitation across all climatological thresholds and seasons.

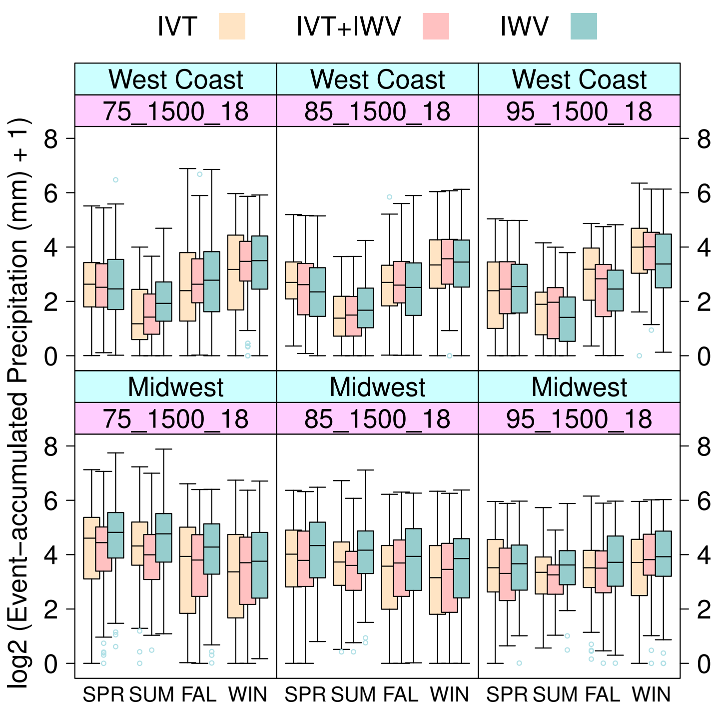

Figure 11. Boxplots of base-2 logarithmic transformation of event-accumulated precipitation (mm), plus 1, over unit area swept by AR in West Coast and Midwest during different seasons—spring (SPR: March–May), summer (SUM: June–August), fall (FAL: September–November), and winter (WIN: December–February)—according to IVT, $IVT+IWV$, and IWV-based AR indices with the 1500-km length and 18-hr persistent duration criteria, labeled as climate threshold in percentile (75, 85, or 95)_1500_18.

4 Discussions

A single optimal AR detection algorithm expressing the surface precipitation impacts does not exist. A hint of bifurcation in our analysis started in Figure 2, in which the Midwest climate thresholds underwent a greater seasonal change than that of the West Coast. In section 3.3, we further found that, with meandering south-north mountain ranges in the West Coast, IVT-based detection algorithms captured the intense orographic precipitation better than the IWV-based ones. This is consistent with the trend to use IVT-based detection algorithms (Guan & Waliser, 2015). However, in the Midwest, in the absence of prominent orographic lifting, IWV-based AR indices were associated with most event-average precipitation and event-accumulated precipitation.

Midwest ARs recruit moisture from tropical sources such as the Gulf of Mexico, Caribbean Sea, subtropical eastern North Pacific, and the Atlantic coast of Central America (Dirmeyer & Kinter, 2009, 2010). The diverse sources complicate the ARs' characteristics (Dirmeyer & Kinter, 2010). In section 3.3.3, the seasonality of event-accumulated precipitation in the Midwest shifted its peak phase from warm to cold seasons along with rising climate thresholds (Figure 11), suggesting a rolling change of moisture sources and baroclinicity as the seasons progressed. On the other hand, West Coast AR's peak phase remained the same regardless of the changing climate threshold. There is a caveat, however. The West Coast's south-north geographic features are inhomogeneous. The landfalling AR characteristics between the Pacific Northwest and California coast are different in terms of occurrence frequency, occurrence time, and distribution and intensity of related precipitation (Neiman et al., 2008). Therefore, to further refine the AR detection algorithms, the entire North American West Coast ARs could be divided into northwest and southwest ARs.

The combined $IVT{+}IWV$-based indices should be used cautiously. It is only the best of both worlds when the goal is to extract snapshots of extreme precipitating events. As seen in Figure 8, it led to the highest event-average precipitation rate in both West Coast and Midwest. This was, however, achieved through few and short events (Figures 4a, 5). In fact, they performed the worst in AR–precipitation relation metrics such as *Accuracy* and $F1$ *scores* (Figures 6, 7, S3, and S4).

Moreover, climate thresholds and moisture fields had first-order influences on the associated surface hydrometeorological impacts. However, more restrictive persistent duration thresholds can help obtain higher event-accumulated precipitation if that is the goal of detection (Figure 10).

Calculation of IVT-based indices requires height-dependent horizontal winds, so reanalysis data are indispensable. Previous studies have suggested that AR characteristics were robust across different reanalysis data (Nayak & Villarini, 2017; Ralph et al., 2019). We used MERRA2 here since Nayak and Villarini (2017) recommended high-resolution products for AR impact assessments. Nevertheless, we showed that depending on the goal, IWV could provide optimal AR indices. When IWV is useful, researchers can use satellite or radiosonde water vapor measurements in lieu of reanalysis.

5 Conclusions

This paper investigated the optimal AR detection algorithm for expressing AR's surface precipitation effects using data in MERRA data or ARTMIP. We applied a solution-driven approach by first asking which impacts, in which region, and in what time scale and period were of concern. We then used an algorithm combining climatological thresholds, image processing, and statistical methods to create large ensembles of AR indices for answering the questions with uncertainty quantification aided by detailed data visualization. Specifically, we varied the values of four factors—moisture fields, climatological thresholds, shape criteria, and duration thresholds—to generate an ensemble of 81 AR indices for the West Coast and 81 indices for the Midwest regions from 2006 to 2015 (Figure 1). With GPCP data, we examined the AR indices' association with the surface precipitation impacts, including the daily co-occurrence (section 3.2), event-average precipitation rate (section 3.3.1), and per-event accumulation (sections 3.3.2 and 3.3.3).

The identified Midwest ARs had more accumulated time steps (Figure 4b), longer average per-event durations (Figure 5), more event-average precipitation (Figures 8, S5, and S6), and more event-accumulated precipitation (figure 10 than the West Coast ARs. The results were sensitive to the selection of moisture field and climatological threshold in index generation. In West Coast and Midwest, IWV-based AR indices identified the most abundant AR event time steps and most accurately associated AR to days with precipitation. These were observed at the coarse-grain regional (Figure 6) and fine-grain grid-point scales (Figures 7, S3, and S4). A restrictive climate threshold, such as the 95th percentile, emphasized extreme instances but limited event duration; therefore, it led to higher event-average precipitation rates. The most restrictive combination of 95th percentile $IVT+IWV$-based indices yielded the highest average precipitation (Figures 8, S5, and S6).

However, it is important to use both event-average and event-accumulated precipitation as metrics for surface hydrometeorological impacts when scrutinizing the AR indices. Therefore, we defined an area swept by each AR event (Figures 9, S7, and S8) and calculated the event-accumulated precipitation per unit area for each AR index (Figure 10). On the West Coast, the 75th percentile IWV-based indices were associated with the most accumulated precipitation, while the 95th percentile IVT captured the accumulated precipitation the best (Figure 10a). This could be explained by the IVT's better representation of intense coastal orographic precipitation. IWV-based AR indices with the longest persistent duration thresholds were associated with the most accumulated precipitation in the Midwest across a range of climate thresholds (Figure 10b). Therefore, we recommend to use IWV-based algorithm to identify AR-related surface precipitation in the Midwest but IVT-based algorithm to capture the orographically-induced precipitation in the West Coast.

Even more, the AR event-accumulated precipitation showed seasonality (Figure 11). The accumulated precipitation of all West Coast landfalling ARs had a clear seasonal cycle with the maximum in the winter and the minimum in the summer. However, for the Midwest ARs, the phase of the seasonal cycle depended on the climatological threshold. Increasing the climatological threshold from the 75th to the 95th percentile shifted the maximum from the warm spring–summer to the cold winter; this reflects the effects of seasonal change of moisture sources and atmospheric baroclincity.

In conclusion, an optimal AR detection algorithm should be adaptive to the types of impact to be addressed, the associated physical mechanisms in the affected regions, timing such as the phase in the seasonal cycle, and event durations. The systematic ensemble approach we used was made possible by distributed parallel computing with data and, specifically, the divide-and-recombine approach using the R-based DeltaRho backended by a Hadoop system. This study's findings provide useful information for future creators and users of AR indices who consider surface precipitation in their decision processes. Our detection algorithms and computational approach can be applied to climate model output, such as CMIP6, to explore the changes of ARs and AR-related surface precipitation impacts in climate change scenarios.

Acknowledgments

The authors are grateful to R. Cannoodt, D. Crabill, Y. Song, M. Bowers and Purdue ITAP RCAC for their help on computing with data. We are indebted to the discussions with ARTMIP scientists, especially T. O'Brien, J. Rutz, and C. Shields. We also thank W. L. Downing and C. Shen for insightful comments. Data is available through NCAR CDG (2019) and NCAR RDA (2018); AR indices further generated will be available in NCAR CDG (2020). The work is supported by DARPA-BAA-16-43-D3M-FP-051.

References

Becker, R. A., Cleveland, W. S., & Shyu, M. J. (1996). The visual design and control of trellis display. *Journal of Computational and Graphical Statistics*, *5*(2), 123–155. doi: 10.1080/10618600.1996.10474701

Cleveland, W. S., & Hafen, R. (2014). Divide and recombine (D&R): Data science for large complex data. *Statistical Analysis Data Mining*, *7*(6), 425–433. doi: 10.1002/sam.11242

Cleveland, W. S., & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, *79*(387), 531–554. doi: 10.1080/01621459.1984.10478080

Dettinger, M. D. (2011). Climate change, atmospheric rivers, and floods in California - a multimodel analysis of storm frequency and magnitude changes. *Journal of the American Water Resources Association*, *47*(3), 514–523. doi: 10.1111/j.1752-1688.2011.00546.x

Dettinger, M. D. (2013). 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

Dettinger, M. D., Ralph, F. M., Das, T., Neiman, P. J., & Cayan, D. R. (2011). Atmospheric Rivers, Floods and the Water Resources of California. *Water*, *3*(2), 445–478. doi: 10.3390/w3020445

Dirmeyer, P. A., & Kinter, J. L. (2009). The "Maya Express": Floods in the U.S.

Midwest. *EOS Transactions*, *90*(12), 101–102. doi: 10.1029/2009EO120001

Dirmeyer, P. A., & Kinter, J. L. (2010). Floods over the U.S. midwest: A regional water cycle perspective. *Journal of Hydrometeorology*, *11*(5), 1172–1181. doi: 10.1175/2010JHM1196.1

Eiras-Barca, J., Brands, S., & Miguez-Macho, G. (2016). Seasonal variations in North Atlantic atmospheric river activity and associations with anomalous precipitation over the Iberian Atlantic margin. *Journal of Geophysical Research Atmosphere*, *121*(2), 931–948. doi: 10.1002/2015JD023379

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., ... Zhao, B. (2017). 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

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

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

Guan, B., Molotch, N. P., Waliser, D. E., Fetzer, E. J., & Neiman, P. J. (2013). The 2010/2011 snow season in California's Sierra Nevada: Role of atmospheric rivers and modes of large-scale variability. *Water Resources Research*, *49*(10), 6731–6743. doi: 10.1002/wrcr.20537

Guan, B., & Waliser, D. E. (2015). Detection of atmospheric rivers: Evaluation and application of an algorithm for global studies. *Journal of Geophysical Research Atmosphere*, *120*(24), 12,514–12,535. doi: 10.1002/2015JD024257

Hastie, T., & Stuetzle, W. (1989). Principal Curves. *Journal of the American Statistical Association*, *84*(406), 502–516. doi: 10.1080/01621459.1989.10478797

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical Learning: Data mning, inference, and prediction.* Springer, New York, NY. doi: 10.1007/978-0-387-84858-7

Huffman, G. J., Adler, R. F., Morrissey, M. M., Bolvin, D. T., Curtis, S., Joyce, R.,

... Susskind, J. (2001). 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

Lavers, D. A., & Villarini, G. (2013). Atmospheric rivers and flooding over the central United States. *Journal of Climate*, *26*(20), 7829–7836. doi: 10.1175/JCLI-D-13-00212.1

Lavers, D. A., & Villarini, G. (2015). 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

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

Leung, L. R., & Qian, Y. (2009). Atmospheric rivers induced heavy precipitation and flooding in the western U.S. simulated by the WRF regional climate model. *Geophysical Research Letters*, *36*(3), L03820. doi: 10.1029/2008GL036445

Luo, Q., & Tung, W.-w. (2015). Case study of moisture and heat budgets within atmospheric rivers. *Monthly Weather Review*, *143*(10), 4145–4162. doi: 10.1175/MWR-D-15-0006.1

Nayak, M. A., & Villarini, G. (2017). A long-term perspective of the hydroclimatological impacts of atmospheric rivers over the central United States. *Water Resources Research*, *53*(2), 1144–1166. doi: 10.1002/2016WR019033

Nayak, M. A., & Villarini, G. (2018). Remote sensing-based characterization of rainfall during atmospheric rivers over the central United States. *Journal of Hydrology*, *556*, 1038–1049. doi: 10.1016/j.jhydrol.2016.09.039

NCAR CDG. (2019). *3-hourly MERRA2 IVT, uIVT, vIVT, IWV data computed for ARTMIP.* (last access: 24 January 2019) doi: 10.5065/D62R3QFS

NCAR CDG. (2020). *ARTMIP Tier 1 Catalogues.* (last access: 12 August 2020) doi: 10.5065/D6R78D1M

NCAR RDA. (2018). *GPCP Version 1.3 One-Degree Daily Precipitation Data Set.* Boulder CO: Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. (last access: 26

August 2019) doi: 10.5065/PV8B-HV76

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

Neiman, P. J., White, A. B., Ralph, F. M., Gottas, D. J., & Gutman, S. I. (2009). A water vapour flux tool for precipitation forecasting. *Proceedings of the Institution of Civil Engineers: Water Management*, *162*(2), 83–94. doi: 10.1680/wama.2009.162.2.83

O'Brien, T. A., Payne, A. E., Shields, C. A., Rutz, J., Brands, S., Castellano, C., ... Zhou, Y. (2020). Detection uncertainty matters for understanding atmospheric rivers. *Bulletin of the American Meteorological Society*, *101*(6), E790-E796. doi: 10.1175/bams-d-19-0348.1

Payne, A. E., & Magnusdottir, G. (2016). Persistent landfalling atmospheric rivers over the west coast of North America. *Journal of Geophysical Research Atmosphere*, *121*(22), 13,287–13,300. doi: 10.1002/2016JD025549

Ralph, F. M., Coleman, T., Neiman, P. J., Zamora, R. J., & Dettinger, M. D. (2013). Observed impacts of duration and seasonality of atmospheric-river landfalls on soil moisture and runoff in coastal Northern California. *Journal of Hydrometeorology*, *14*(2), 443–459. doi: 10.1175/JHM-D-12-076.1

Ralph, F. M., Neiman, P. J., Kiladis, G. N., Weickmann, K., & Reynolds, D. W. (2011). A multiscale observational case study of a Pacific atmospheric river exhibiting tropical-extratropical connections and a mesoscale frontal wave. *Monthly Weather Review*, *139*(4), 1169–1189. doi: 10.1175/2010MWR3596.1

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

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

Ralph, F. M., Neiman, P. J., Wick, G. A., Gutman, S. I., Dettinger, M. D., Cayan, D. R., & White, A. B. (2006). Flooding on California's Russian River: Role of atmospheric rivers. *Geophysical Research Letters*, *33*(5), L13801. doi: 10.1029/2006GL026689

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

Rutz, J. J., Shields, C. A., Lora, J. M., Payne, A. E., Guan, B., Ullrich, P., … Viale, M. (2019). The Atmospheric River Tracking Method Intercomparison Project (ARTMIP): Quantifying uncertainties in atmospheric river climatology. *Journal of Geophysical Research: Atmospheres*, *124*, 13,777–13,802. doi: 10.1029/2019JD030936

Rutz, J. J., & Steenburgh, W. J. (2012). Quantifying the role of atmospheric rivers in the interior western United States. *Atmospheric Science Letters*, *13*(4), 257–261. doi: 10.1002/asl.392

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

Sellars, S. L., Gao, X., & Sorooshian, S. (2015). 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

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

Tung, W.-w., Barthur, A., Bowers, M. C., Song, Y., Gerth, J., & Cleveland, W. S. (2018). Divide and recombine (D&R) data science projects for deep analysis of big data and high computational complexity. *Japanese Journal of Statistics and Data Science*, *1*(1), 139–156. doi: 10.1007/s42081-018-0008-4

Wick, G. A., Neiman, P. J., & Ralph, F. M. (2013). 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

Wolter, K., & Timlin, M. S. (1993). *Monitoring enso in coads with a seasonally adjusted principal component index.* Paper presented at the 17th Climate Diagnostics Workshop, NOAA/NMC/CAC, NSSL, Oklahoma Clim. Survey, CIMMS and the School of Meteor., Univ. of Oklahoma, Norman, OK.

Zhu, Y., & Newell, R. E. (1998). 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