ARTMIP-Early Start Comparison of Atmospheric River Detection Tools: How Many Atmospheric Rivers Hit Northern California's Russian River Watershed?

F. Martin Ralph*[1], Anna M. Wilson[1], Tamara Shulgina[1], Brian Kawzenuk[1], Scott Sellars[1], Jonathan J. Rutz[2], Maryam Asgari-Lamjiri[1], Elizabeth A. Barnes[3], Alexander Gershunov[1], Bin Guan[4], Kyle Nardi[3], Tashiana Osborne[1], and Gary A. Wick[5]

1- *Center for Western Weather and Water Extremes, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA*

2- *NOAA/NWS/Western Region Headquarters, Salt Lake City, UT*

3- *Department of Atmospheric Science, Colorado State University, Fort Collins, CO*

4- *Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, CA*

5- *NOAA/Earth System Research Laboratory/Physical Sciences Division, Boulder, CO*

Submitted to Climate Dynamics

Apr 2018

*Corresponding author address: A. M. Wilson, Center for Western Weather and Water Extremes, Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, #0224, La Jolla, CA 92093-0224

Email: anna-m-wilson@ucsd.edu

Abstract (150-250 words)

Many Atmospheric River Detection Tools (ARDTs) have now been developed. However, their relative performance is not well documented. This paper compares a diverse set of ARDTs by applying them to a single location where a unique 12-year-long time-series from an atmospheric river observatory at Bodega Bay, California is available. The study quantifies the sensitivity of the diagnosed number, duration, and intensity of ARs at this location to the choice of ARDT, and to the choice of reanalysis data set. The ARDTs compared here represent a range of methods that vary in their use of different variables, fixed vs. percentile-based thresholds, geometric shape requirements, Eulerian versus Lagrangian approaches, and reanalyses.

The ARDTs were evaluated first using the datasets documented in their initial publication, which found an average annual count of 19±7. Applying the ARDTs to the same reanalysis dataset yields an average annual count of 19±4. Applying a single ARDT to three reanalyses of varying grid sizes (0.5°, 1.0° to 2.5°) showed little sensitivity to the choice of reanalysis. While the annual average AR event count varied by about a factor of two (10-25 per year) depending on the ARDT, average AR duration and maximum intensity varied by less than ±10%, i.e., 24±2 h duration; 458±44 kg m^{-1} s^{-1} maximum IVT. ARDTs that use a much higher threshold for integrated vapor transport were compared separately, and yielded just 1-2 ARs annually on average. Generally, ARDTs that include either more stringent geometric criteria or higher thresholds identified the fewest AR events.

1. Introduction

Atmospheric rivers (ARs) are elongated, narrow regions of enhanced water vapor transport and are a major feature in the global hydrologic cycle. They are responsible for nearly 90% of poleward water vapor transport in the midlatitudes, while covering only 10% of the zonal circumference of the earth (Zhu and Newell, 1998; Ralph et al., 2004, 2017, 2018; Guan and Waliser, 2015). At the regional level, they represent an important contribution to precipitation from event to annual scales (e.g. Dettinger et al., 2011; Ralph and Dettinger, 2012; Lavers et al., 2015; Waliser and Guan, 2017; Young et al., 2017; and many others). The first papers describing ARs used perturbations to the mean flow to identify these rivers, where in the midlatitudes the perturbations are almost all directed poleward (Newell et al., 1992; Newell and Zhu, 1994; Zhu and Newell, 1992). This methodology used European Centre for Medium-Range Weather Forecasts (ECMWF) wind and humidity data at seven pressure levels and 12-hour temporal resolution. Ralph et al. (2004) pioneered the methodology to detect these features via atmospheric water vapor content observed by satellite. Building on these earlier studies, many different Atmospheric River Detection Tools (ARDT) applying automated detection methods to various datasets have been developed and described in the literature, especially during the past few years. However, there has been little assessment of how the catalogs created by these algorithms compare with each other. This paper set out to address this gap in the simplest possible way – at a single location where unique observations of AR conditions could be used as well. In addition, a community effort, organized by an ad-hoc planning committee, began developing an approach to perform such a comparison, called the Atmospheric River Tracking Method Intercomparison Project (ARTMIP; Shields et al., 2018). These efforts merged in a way such that this paper represents an early-start analysis that helps set the stage for ARTMIP, as will be described through the paper.

In context of the ARTMIP goals, the purpose of this paper is to present a cross section of these different methodologies in the specific framework of determining how many ARs strike the flood-prone Russian River in northern California each year on average. The number of ARs hitting the northern California coast made the difference between four years of severe drought [water years (WY) 2012-2015], when a lower than normal number of ARs made landfall, and the wettest year on record - WY 2017, when, by our reckoning, over 30 ARs hit the region. This latest period is one extreme example of the documented role of ARs in ending drought periods (Dettinger, 2013). Studies looking at the effect of reanalysis products on AR detection throughout the globe have found that there is generally good agreement between reanalysis- and satellite-based datasets (Guan and Waliser, 2015; Jackson et al., 2016; Brands et al., 2017). Differences in reanalysis datasets in a general sense come from their various resolutions (Guan and Waliser, 2017) as well as from different representations of important physical processes, such as the transport of moisture and energy (Trenberth et al., 2011). Differences in AR events identified using different reanalysis are seen in AR characteristics like landfall location, intensity, and spatial extent (Lavers et al., 2012; Jackson et al., 2016; Guan and Waliser, 2017; Guan et al., 2018). Jackson et al. (2016), using the Wick et al. (2013) algorithm, examined 4 datasets (CFSR, MERRA, ERA-I, and the Twentieth Century Reanalysis) during the cool season (October – March) in water years 1998 – 2012, and found that in the first three reanalyses, AR landfall detections on the U.S. West Coast (between latitudes 15° – 55° N) agreed with satellite-based detections within 5%. Lavers et al. (2012) conducted an analysis of 5 reanalysis datasets (CFSR, MERRA, ERAI, the Twentieth Century Reanalysis, and NCEP-NCAR) through the cool seasons (October – March) in water years 1980-

2010 to look for differences in results for ARs affecting Britain. They found these reanalyses to be in generally good agreement. Good agreement was also found between ERA-Interim and MERRA-2 for ARs in the northeastern Pacific (Guan et al., 2018) and over the globe (Guan and Waliser, 2017), although with NCEP/NCAR being somewhat different from the former two products based on their results. Good agreement in the identification of AR features between reanalysis datasets was found despite the datasets' very different characteristics, such as differences in resolution, differences in assimilation techniques, and differences in data assimilated.

The answer to the question of how many ARs hit the Russian River is expected to vary depending upon the ARDT and reanalysis or observations used, and, as we quantify this variation below, we will elucidate important differences between methods and their application to different datasets. The first step towards accurately estimating the number of ARs that will hit this region in a given year, and to understand how this number may shift with a changing climate, is to be able to quantify this number and understand its sensitivity to identification methods and reanalyses.

The Russian River Watershed in northern California is targeted herein for two main reasons. First, a long-term in-situ dataset that has been collected nearly continuously since 2004 at an Atmospheric River Observatory (ARO; White et al., 2013) located at the University of California Davis' Bodega Marine Laboratory (BBY) on the Sonoma County coast, just outside of the watershed, is available. The ARO provides essential, previously unavailable observations for moisture flux monitoring; its hourly temporal resolution enables highly accurate measurements of AR onset, cessation, and peak strength using a method developed by Ralph et al. (2013). However, this dataset is limited by periods of missing data that may affect its accuracy in longer term statistics.

Second, the Russian River watershed has been and continues to be an active region for AR related research activities. ARs have been shown to be associated with over half of the annual precipitation throughout this region, which has been identified as a hub of AR landfalling activity, meaning it is impacted by the most intense ARs, climatologically, along the entire west coast of midlatitude North America (Gershunov et al., 2017). While ARs provide essential water supply to this watershed (Dettinger et al., 2011; Ralph et al., 2013), they are also the main drivers of floods: the seven strongest precipitation events (resulting in flooding) here between 1997-2006 were all associated with ARs (Ralph et al., 2006). The Russian River watershed is the focus of studies on water management in northern California to leverage increasing knowledge and rapidly improving forecasting skill with respect to AR landfalls towards improved water resource management including enhanced flood protection for the benefit of agriculture, industry, municipal needs, recreation, and ecosystems.

The main objectives of this study are: 1) determine how many ARs, on average, hit the Russian River annually, and understand the variability from year to year and sensitivity to detection algorithms; 2) determine when the algorithms agree/disagree with one another; and 3) discuss implications of using different ARDTs and different reanalysis datasets.

This study uses the following language in order to refer to ARs. An AR, or an AR "object", is the coherent AR feature in space tracked throughout its temporal evolution; an AR "event" is the presence of AR conditions at a point in space over some continuous length of time (AR objects are Lagrangian, while AR events are Eulerian); an AR "timestep" is one temporal step in a given dataset (in situ or reanalysis) with AR conditions present; and "AR conditions" signifies that AR criteria for a given detection method at a given location are met at a timestep.

This paper is arranged as follows: Section 2 describes the datasets and methods used in this work; Section 3 presents the results of the ARDTs applied to their original dataset at the grid cell containing Bodega Bay; Section 4 repeats this analysis with the ARDTs applied to NASA's Modern Era Retrospective-Analysis for Research and Applications-2 (MERRA-2) reanalysis dataset at the grid cell containing BBY; Section 5 uses the ARDT developed by Rutz et al. (2014) on three different reanalysis datasets; and Section 6 provides conclusions and discussion.

2. Data and Methods

This study uses three sets of AR catalogs. The first of these sets is based on AR detection techniques (ARDTs) developed by Ralph et al. (2013), Sellars et al. (2013), Rutz et al. (2014), Guan and Waliser (2015), Mundhenk et al. (2016), and Gershunov et al. (2017), applied to the observations or reanalyses in their original, respective publications (Native Reanalysis AR catalogs hereafter). The second set is based on applying these same ARDTs plus the method of Wick et al. (2013) to observations or NASA MERRA-2 Reanalysis (Gelaro et al., 2017) 3-hourly data of IVT and/or IWV at the [38.5N, 123.125W] grid cell used to represent the Bodega Bay area for the time period of November 2004 – April 2016 (MERRA-2-based AR catalogs hereafter). The third set of catalogs are collected using the Rutz et al. (2014) ARDT on three different reanalyses, namely MERRA-2, ERA-Interim, and NCEP/NCAR (Rutzetal2014-based AR catalogs hereafter). This latter set of catalogs is used to aid in differentiating between uncertainties that arise from using different methodologies from those that arise from using different datasets. The acronyms for each method, which will be used to identify them throughout the remainder of the paper, and parameter and geometry thresholds used for AR identification in each catalog, are presented in Table 1.

These algorithms are all state-of-the-art methodologies developed to answer specific questions – the methods and datasets used vary accordingly. The observational dataset is described in detail in Section 2a. Section 2b provides a brief description of each ARDT used in Sections 3-5. Some of the algorithms were modified from the original published methodology for the purpose of this study. Those modifications and the reasons they were made will be discussed in detail below.

2a. Observations and Reanalyses

The BBY ARO observations, available from NOAA's Earth System Research Laboratory, Physical Science Division website (https://www.esrl.noaa.gov/psd), provide hourly records of AR landfalls near California's Russian River Basin northwest of San Francisco, California, between 13 November 2004 – present (in this study, data through April 30, 2016 are included). The BBY ARO is part of an extended network of observing stations throughout the western U.S. (White et al., 2013). The ARO includes a wind profiler to observe a vertical profile of horizontal winds, a GPS-Met sensor to record the integrated water vapor (IWV) in the atmosphere, a radio acoustic sounding system (RASS) to observe the vertical profile of potential temperature, and surface meteorological instrumentation.

AR events were reconstructed based on currently available data from NOAA ESRL's Hydrometeorology Testbed archive and data available from the catalog provided in Ralph et al. (2013). The hourly ARO data records were first compiled in order to maximize the amount of available data, due to some missing data in the archives currently available online. First, the hourly records were collected from the IWV flux table data output from the ARO. If the flux table data were missing, then these data were filled in with raw observations from the profiler and the GPS-Met. If both of those were missing, the AR catalog provided in Ralph et al. (2013) was used. The

Ralph et al. (2013) AR catalog is assumed to contain all of the AR hours from water years 2005 – 2010. One limitation of the ARO dataset in this paper is the amount of missing data for the period of record, during which over 36000 hours are missing, including 77% of October - March hours in water year 2013. Outside of 2013, most of the missing data is from water years 2005 – 2010, and while there are some differences in the AR criteria used in Ralph et al. (2013) and this study (see Section 2b), it is expected that most ARs during those years were counted in that original catalog. Excluding periods with ARO missing data reduces each catalog presented in this study by 24% of all AR time steps and 24% of the event counts, on average across the catalogs used. For only much stronger ARs (identified with a much higher IVT threshold, see section 2b), excluding these periods reduces AR time steps by 15% and AR event counts by 17%, on average. In spite of these missing hours, the observational catalog provides unique and detailed information at high temporal resolution regarding landfalling AR onset, cessation and peak strength, which is not detectable with any other in situ data at this scale.

2b. Algorithm Descriptions

Ralph et al. (2013) – Ralphetal2013

In this study, AR events in the two observational catalogs (Ralphetal2013-Obs and Ralphetal2013-Obs47 in Table 1; see Section 2a for a detailed description of the observational dataset) were identified as a period lasting at least 12 hours for which each hourly integrated water vapor (IWV) amount was greater than or equal to 2 cm. Additionally, for Ralphetal2013-Obs, each IWV flux amount was required to exceed or equal 20 cm m s^{-1}, and for Ralphetal2013-Obs47, each IWV flux amount was required to exceed or equal 47 cm m s^{-1}. These criteria were adapted from those used in Ralph et al. (2013) assessing ARs impacting the coastal mountains of northern

California (Cazadero). The two main differences in the method used in the current analysis are that total IWV flux is used instead of the upslope component directed at Cazadero, and the event duration is required to be 12 hours instead of 8. The 20 cm m s^{-1} value for IWV flux roughly corresponds to 250 kg $m^{-1}s^{-1}$ of integrated vapor transport (IVT), and the 47 cm m s^{-1} value corresponds to 500 kg $m^{-1}s^{-1}$. These values are based on 130 radiosonde releases at BBY during 2016-2017. Using the total flux instead of just one directional component is closer to the concept of identifying ARs based on their IVT, which is the method employed by all other ARDTs evaluated here. The AR event duration requirement was adjusted in this study so that AR events based on ARO observations are comparable with AR events defined based on reanalysis datasets, some of which are available at a maximum temporal resolution of 6-hour time steps.

Sellars et al. (2013) – SGS2013

An object-oriented detection algorithm was developed by SGS2013 to better understand and analyze massive amounts of spatiotemporal data. In the publication, the algorithm was used on precipitation data from the Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks (PERSIANN) dataset (Hsu et al., 1997). Sellars et al. (2017a) adapted the algorithm for application to IVT using MERRA-2, with an IVT threshold of 750 kg $m^{-1}s^{-1}$. The algorithm identifies an ‘object’ as a feature over the selected threshold that is connected in space and time and lasts for at least 24 hours. The results are organized in a database of objects and their characteristics that are ready for further statistical analysis (Sellars et al., 2017b). In this study, the algorithm was applied to MERRA-2 IVT data with an IVT threshold of 500 kg $m^{-1}s^{-1}$.

Wick, Neiman, and Ralph (2013) – WNR2013

WNR2013 developed a methodology for objective and automated detection and characterization of ARs based on the detection originally described in Ralph et al. (2004), using fields of IWV from satellite observations and model outputs, and was verified against results from that study. A primary purpose of this ARDT was the validation and comparison of forecast fields with observational data as performed in Wick et al. (2013b). The basic objective criteria used in this algorithm are: 1) IWV content > 20 mm, 2) length > 2000 km, and 3) width < 1000 km. Standard image processing techniques such as thresholding and skeletonization are used to identify ARs based on these criteria. This algorithm has been extended for use with IVT (Wick et al. 2014; Mahoney et al. 2016) and is used in this study with two IVT thresholds; one including all ARs above 250 kg $m^{-1}s^{-1}$, and the other detecting stronger ARs, with a threshold of 500 kg $m^{-1}s^{-1}$. This approach was not included in the first set of catalogs as its native application was to satellite IWV data which, due to intermittent inter-swath gaps, did not lend itself to detailed comparison with the reanalysis products.

Rutz, Steenburgh, and Ralph (2014) – RSR2014

RSR2014 has been applied to a number of different reanalysis datasets, including National Center for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR, used in Section 3; Kalnay et al., 1996), ERA-Interim (Dee et al., 2011), and MERRA-2. The motivation for development of this algorithm was to be able to identify ARs in reanalysis over not just the coastal western US but also the interior. They identified ARs as features ≥ 2000 km in length (without accounting for curvature of features) with IVT ≥ 250 kg $m^{-1}s^{-1}$ throughout. They imposed no width requirement on these features, and their identified AR events are not dependent on duration.

Guan and Waliser (2015) – GW2015

GW2015 developed a methodology for global detection of ARs based on ERA-Interim 6-hourly, 1.5° resolution IVT fields for the period of 1997-2014. The motivation behind GW2015's development was to assess ARs objectively and consistently on a global scale using IVT and geometric characteristics. The thresholds used in this methodology include requiring IVT intensities to be higher than the climatological 85th percentile computed for each season and grid cell and a fixed minimum limit of 100 kg $m^{-1}s^{-1}$, mean IVT direction to be within 45° of the AR shape orientation, length of AR features to be longer than 2000 km, and the length to width ratio to be greater than 2. Using the percentile method, the threshold at the Bodega Bay grid point ranges from 166-254 kg $m^{-1}s^{-1}$ depending on the season. No minimum duration requirement is considered for AR events in this methodology.

Mundhenk, Barnes, and Maloney (2016) – MBM2016

MBM2016 used the MERRA-1 reanalysis dataset with 0.5°x0.667° spatial resolution and 6 hourly temporal resolution (Rienecker et al., 2011) to establish an occurrence-based algorithm to detect ARs from 1979-2014. In this study, the catalog was created using MERRA-1 at 1.25° resolution, up through the end of the last full water year for which MERRA-1 is available (2015). The motivation for development of this algorithm was to enable further investigation of AR dynamics and variability over the North Pacific region throughout the year. This algorithm is based on fields of anomalous IVT intensities, so that AR features are required to have anomalous IVT fields over a given threshold. The published version of the MBM2016 algorithm uses an IVT anomaly threshold of 250 kg $m^{-1}s^{-1}$ for AR detection. The published methodology in MBM2016 requires at least 25 grid points (~1400 km) along the major axis of the AR and a ratio of 1.6:1 between the major and minor axis of the AR feature. After this, AR features go through another filter to remove weak features with mean anomalous IVT intensities < 305 kg $m^{-1}s^{-1}$ and the features that have west-east direction with center of mass southward of 20 N and orientation off the parallel of less than 0.95 radians. Well-developed tropical cyclone features, such as intense circular features or those that include eyelike holes are also removed. After this publication, the MBM2016 algorithm was modified in order to make it applicable to a wide variety of reanalysis datasets, as opposed to MERRA1 alone. The process of generalizing the algorithm resulted in changes to both intensity and geometric criteria, though the modified algorithm retained the ability to isolate long, narrow plumes of anomalous water vapor transport. In this study, the intensity criteria used is the 94^{th} percentile of IVT anomaly over the entire spatial domain, which is 180.86 kg $m^{-1}s^{-1}$ for MERRA-1, in Section 3; and 186.51 kg $m^{-1}s^{-1}$ for MERRA-2, in Section 4. Geometric criteria were modified as follows for this application: 1) the aspect ratio is 1.4:1 instead of 1.6:1; 2) the latitude threshold was moved from 20 N to 16 N; 3) the weak feature threshold was changed from 305 kg $m^{-1}s^{-1}$ to a threshold based upon the 95^{th} percentile of IVT anomaly (201.73 kg $m^{-1}s^{-1}$ for MERRA-1, Section 3; 208.75 kg $m^{-1}s^{-1}$ for MERRA-2, Section 4); 4) the orientation off the parallel of less than 0.95 radians criteria was modified to be a mean IVT orientation between 170° and 230° from horizontal, where 0° is pointing to the east; and, 5) multiple intensity peaks are not segmented into separate ARs in the version used in this study.

Gershunov et al. (2017) – GSR2017

GSR2017 developed an automated algorithm to detect AR landfalls using the NCEP/NCAR reanalysis dataset with 2.5º horizontal resolution and 6 hourly temporal resolution for the period of 1948-2017. The motivation behind development of their catalog was to include both IVT and IWV, to demonstrate the association of ARs with heavy precipitation, and to apply their algorithm to a dataset with a long enough record to resolve interdecadal variability. According to their methodology, landfalling AR features are required to have minimum IVT intensities of 250 kg $m^{-1}s^{-1}$ and IWV in excess of 15 mm, cross the North America West Coast between 20°N-60°N, and be at least 1500 km long. Movement of the center of the AR, which is defined as the grid cell with maximum IVT intensity along the coast, is allowed between each pair of time steps, but no more than 5º (north/south). ARs making landfall simultaneously are distinct if their centers are at least 7500 km (7.5° north-south) apart from each other. Based on this methodology, AR events are required to last for at least 18 hours (3 consecutive analysis time steps). In the catalogs created for this study, however, this requirement is adjusted as follows: if the required atmospheric conditions are observed at the grid cell used to represent BBY during at least 12 of 18 consecutive hours, then an AR event is counted.

Overall, there is considerable variability between the different algorithms. Some employ just a few criteria beyond intensity (e.g. RSR2014) in order not to exclude any relevant events, and others are much more complicated, in order to try and constrain the sample to a stricter definition of atmospheric river without allowing overlap with other features (e.g. MBM2016). Detection techniques also vary in their use of anomalies vs. static thresholds for intensity. The detection technique is fundamentally entangled with quantitative algorithm requirements beyond the descriptive, qualitative "plume of moisture", e.g.: How long does it need to be to qualify as a plume?; How wide is it before it is something else?; How strong or anomalous does the IVT within the plume need to be to separate it from its surroundings? In other words, the algorithms must not only define what an AR is, but what it is not, and this scope and specificity is different for each algorithm depending upon the science question addressed during its development.

3. Comparing ARDT Results at Bodega Bay Using Native Reanalyses

In this section, the ARDTs are applied to the datasets and resolutions used in their original publication at the closest grid point to Bodega Bay (Figure 1). All AR events were required to last at least 12 hours in both the ARO (hourly data would require 12 timesteps) and reanalysis datasets (3-hourly data would require four timesteps for MERRA-1 and MERRA-2; 6-hourly data would require two timesteps for NCEP/NCAR and ERA-Interim). In order to minimize the differences between these different time steps, the hourly and 3-hourly datasets were sampled at 0, 6, 12, and 18 UTC and those instantaneous values were used to determine the presence of an AR event. AR events were considered to be distinct if separated by one 6-hour period. Note that without this adjustment, the different temporal resolutions of the datasets would cause differences in AR characteristics. This could result either from ARs no longer meeting the duration criteria (e.g., one of the timesteps falls below the thresholds for a given ARDT), or it could result from multiple distinct ARs being reported in a given time period, with below threshold periods separating a longer duration AR identified with the lower resolution dataset. Similarly, the different temporal resolutions would also affect AR duration and intensity characteristics.

The period of time considered was November 2004-September 2015, in order to use the maximum amount of data where all datasets were available. This will be referred to as water years 2005-2015; however, water year 2005 is missing the month of October. Water year 2016 was not included in this section, because one of the datasets, MERRA-1, was unavailable during part of the peak of that year, as it was discontinued in favor of MERRA-2. During water years 2005 – 2015 an average of 19 ARs of at least weak strength were detected at Bodega Bay each year with an average duration of 24 hours (Table 2). For the stronger ARs identified by the three detection tools using an IVT threshold of 500 kg m^{-1} s^{-1} (see Table 1), the average number per year between 2005 – 2015 was 2, with an average duration of 17 hours. To investigate ARDT agreement on an individual AR basis, one particularly active peak season, Dec-Feb 2006, was further explored (Figure 2). During periods of longer lasting and higher IVT and IWV values there is agreement among at least five of the catalogs in most cases, however no extreme event had agreement across all seven catalogs. During periods of lower IVT and IWV there is more disagreement among ARDTs. One potential reason for this disagreement is that the reanalyses, in particular the coarser grids of MERRA-1 (in this study used at 1.25° resolution), ERA-Interim, or NCEP/NCAR, may not have recorded the same timing or magnitude of the peak in IVT or IWV that are presented from MERRA-2.

Analysis of individual detected AR timesteps shows relatively large disagreements as well, even within strong ARs. For example, several timesteps with IVT greater than 850 kg $m^{-1}s^{-1}$ were identified as ARs by as few as four ARDTs (Figure 3a). This could be due to the difference in IVT magnitude in different reanalysis datasets (Figure 3a is presented using MERRA-2 data, which may have a different magnitude than the IVT of the reanalysis dataset used for detection), as well as the potential for missing data in the observational catalogs. It is also likely that the geometric constraints present in some ARDTs play a role in the disagreement between catalogs, even when high IVT was observed. However, in general, the agreement between catalogs increases during timesteps with higher observed IVT. The IVT in the MERRA-2 appears to exhibit a seasonal cycle, with many more timesteps with higher values of IVT observed during the cool season, which is when all of the extreme ($\geq$ 1000 kg $m^{-1}s^{-1}$) and most strong ($\geq$ 750 kg $m^{-1}s^{-1}$) ARs were observed (Figure 3a and b).

The number of ARs detected per year varies between datasets and annually (Figure 4). The average standard deviation, calculated over the different years and then averaged over the different methods, between detected ARs per year is 7, which is nearly as large as some of the catalog counts (see MBM2016 during years 2014 and 2015). The largest range of detected ARs occurred in 2010, where counts range from 13 (MBM2016-MERRA1) – 30 (RSR2014-NCEP) AR events (this excludes the Ralphetal2013-Obs, which counted 5, but may be suffering from missing data). The total number of AR events detected by each ARDT during the period ranges from 114 to 279 (Table 2). Overall, the interannual variability pattern is similar between ARDTs (excluding the ARO with missing data). For example, the ARDTs all detect more AR events during 2006 and 2011, which were relatively wet years, and detect lower AR event counts during the drought period 2012-2015. These results indicate that very different answers may be found to answer the overarching question of how many AR landfalls of what strength were observed at a given location per year. In general, the stronger the AR is and the longer duration it is, the more likely it is to be identified by all of the ARDTs in all reanalyses or observations.

Throughout the next two sections of the paper, we will first isolate the differences in the detection tools by applying each tool to the MERRA-2 reanalysis, and then, we will isolate differences in the reanalysis datasets by applying the RSR2014 method to MERRA-2, ERA-Interim, and NCEP/NCAR reanalysis datasets. This design will help to pinpoint the reasons behind differences in the catalogs.

4. Comparing ARDT Results at Bodega Bay Using MERRA-2

To isolate uncertainties in AR occurrence and their characteristics caused by distinct attributes of individual ARDTs, in this section we compare AR characteristics in the vicinity of the Russian River watershed for the ARs detected by ARDTs as originally designed, but applied to the same dataset, the MERRA-2 reanalysis, with either its native grid or an interpolated grid required by one of the ARDTs. These MERRA-2-based catalogs are also compared to the observational catalogs created using the Bodega Bay ARO. The MERRA-2 dataset is chosen because it is a state of the art, easily accessible dataset, with high spatial and temporal resolution (the highest among the reanalysis products used in the native catalogs discussed earlier).

Accordingly, eight MERRA-2-based and two observational catalogs (Table 3) are compared in this section based on their identification of AR activity at the MERRA-2 grid cell containing the ARO (38.5N, 123.125W), or directly at the ARO (38.3191N, 123.0728W) for observational catalogs. Five of the catalogs were created using the MERRA-2 native grid, which is 0.5 x 0.625º. Three of the catalogs, using variations on the WNR2013 algorithm, require a grid with equal step sizes in latitude and longitude, and so MERRA-2 was interpolated to a 0.5 x 0.5º grid. The last two catalogs were created using the ARO observations. AR events are required to last for at least 4 consecutive 3-hourly timesteps (12 hours), and are separated by at least one 3-hourly timestep below AR conditions. The hourly ARO observations were transformed into 3-hourly time steps, and only instantaneous 3-hourly records were taken into account. Recall that the ARO has a significant amount of missing data that may affect results here.

During water years 2005 – 2016, the MERRA-2 based and observational catalogs identified an average of 18 ARs at Bodega Bay each year with an average duration of 23 hours (Table 3). For the stronger ARs identified by the three detection tools using an IVT threshold of 500 kg m^{-1} s^{-1} (see Table 1), the average number per year between 2005 – 2016 was 1, with an average duration of 18 hours. Differences in the number of AR events and their characteristics are associated here with differences in the methodologies. First, there is a difference between constant and percentile-based IVT threshold magnitudes used for AR identification. GW2015 applies an IVT threshold magnitude that varies with the season and ranges from 166 – 254 kg $m^{-1}s^{-1}$ (see Table 1 and Section 2); this allows weaker AR conditions to register, especially during the warm season. Three of the catalogs in this section (WNR2013-IVT500; SGS2013; Ralphetal2013-Obs47) are aimed at identifying much stronger ARs and have significantly lower counts than the other 7 catalogs throughout this period. It is important to note that these lower counts are also a reflection of the requirement for the ARs to consistently keep their higher strength (over 500 kg m^{-1}s^{-1}; see Table 1 throughout the entire minimum 12-hour AR duration, instead of just reaching that value at the peak of the AR.

Geometric constraints may also account for observed differences in the number of detected AR events. While the ARDTs have comparable AR length thresholds, five of the catalogs consider other geometric requirements as well, such as length/width ratio, shape, width, and orientation (Table 1). The most restrictive geometric criteria are found in WNR2013 and MBM2016 (see Section 2), and these consistently identify fewer AR events, a result consistent with the sensitivity analysis in Guan and Waliser (2015; their Figure 5).

Even with the variability in the number of AR events, there is good agreement between all catalogs with similar IVT thresholds in terms of average event duration (Table 3). Depending on the AR identification methodology, the number of AR events for the period of record varies from 131 to 268 (for stronger ARs, this range is 13 to 29). Annually, there are 18 AR events per year on average at this grid cell, lasting an average of 23 hours. There is on average 1 stronger AR event per year at this grid cell, lasting an average of 18 hours. The output of ARDTs with similar parameters and geometric characteristics show good agreement with regard to relevant AR characteristics such as frequency, duration, and intensity (GSR2017, GW2015, RSR2014; see Table 1). The WNR2013-IVT AR catalog is constructed using the same IVT threshold magnitude as the four mentioned earlier; however, the number of detected AR events is relatively small and the ARs are shorter in duration, which could be a result of the more restrictive geometric criteria used in this methodology (Table 1).

The 2006 water year peak season (Dec 2005 – Mar 2006) shows increased agreement with the MERRA-2 dataset compared to the ARDTs applied to their native datasets (Figure 5 and Figure 2, respectively). Seven ARDTs are applied to all ARs, and 3 are detecting only stronger ARs, with a minimum threshold of 500 kg $m^{-1}s^{-1}$ (see Table 1). Similar to Figure 2, the stronger and longer duration ARs show good agreement across the ARDTs, while weaker and shorter duration events that are closer to defined thresholds have much less agreement. Some of the differences between Figure 2 and 5 are due to the fact that in Figure 2, ARDTs are applied to different datasets, but presented with MERRA-2 reanalysis values for IVT and IWV, whereas in Figure 5 the ARDTs, excluding the observational datasets, are both applied to and presented with MERRA-2 reanalysis, with two different spatial resolutions. This results in more agreement between the ARDTs in Figure 5 than in Figure 2. Due to the higher resolution of the time steps in the ARDTs in Figure 5 (3-hourly time steps instead of 6-hourly), more distinct events are counted during the end of December 2005.

Over the period analyzed, AR timesteps show a similar pattern to Section 3, where agreement increases with IVT intensity (Figure 6). Here, in contrast to Figure 3, there are few ARDTs identifying AR timesteps with IVT below 250 kg $m^{-1}s^{-1}$ (Table 1), due to MERRA-2 being the sole reanalysis dataset in use in this section. The reason behind some of these timesteps still appearing are the ARO identified timesteps and the WNR2013-IWV presented with associated MERRA-2 IVT. As the ARs get stronger, there is more agreement between different methods; for example, IVT in 3-hour time steps between 500-549 kg $m^{-1}s^{-1}$ shows 5 or more (out of 7) methods agreeing almost 80% of the time; 700-749 kg $m^{-1}s^{-1}$ shows 7 or more (out of 10) methods agreeing almost 80% of the time. Separating the AR timesteps into cool season (October – April) and warm season (May – September) highlights that the stronger ARs occur almost entirely in the cool season, and result in more agreement between ARDTs during this season. (Figure 6b, c).

In terms of interannual variability, the catalogs in different categories (regular AR strength vs. stronger ARs with a minimum threshold of 500 kg $m^{-1}s^{-1}$ (see Table 1)) follow roughly the same patterns (Figure 7). For example, wet years such as 2006 receive higher numbers of ARs detected by almost all catalogs. The BBY ARO records less AR activity; but this is due primarily to missing periods, in particular in 2013. This shows one particular limitation of the in situ observations; however, the agreement between years with no missing data (e.g. 2014-2015) with reanalysis datasets also provides some confidence in the reanalysis datasets to capture the features observed on the ground. The MBM2016 and WNR2013-IVT catalogs consistently record fewer AR events than the others, and this may be related to their stricter geometric requirements, as discussed earlier. The WNR2013-IWV catalog includes similar geometric criteria, but does not consider IVT and does not employ several related geometric criteria (e.g. the aspect ratio), and this may be part of the reason why its AR event counts are higher.

It is notable that the overall results between Sections 3 and 4 are so similar in terms of AR event count. In this section, considering all MERRA-based catalogs (i.e., excluding observational catalogs) using the thresholds meant to include all ARs, the average number of AR events per year is 19. This value is the same as that found in Section 3. However, the standard deviation is reduced from 7 (in Section 3) to 4 (in Section 4) by confining the reanalysis choice to MERRA.

5. Comparing ARDT Results on Different Reanalyses Using the RSR2014 ARDT

In this section, we attempt to isolate the effect of using different reanalysis datasets on AR detection by applying the RSR2014 algorithm to three reanalysis datasets with different spatial and temporal resolutions; these include NCEP/NCAR (6-hour temporal and 2.5° spatial resolution), ERA-Interim (6-hour temporal and 1.5° spatial resolution), and MERRA-2 (3-hour temporal and 0.5°x0.625° spatial resolution, see Figure 1 and Table 4). Since NCEP/NCAR and ERA-Interim data are available in 6-hour time steps, this section uses MERRA-2 3-hourly data sampled every 6 hours, similar to section 3, to facilitate comparison. The RSR2014 algorithm results for ERA-Interim are only available during the cool season, between November – April, and so results for this section are presented for this 6-month period of each water year. Statistics computed for subperiods within 1990-2015 for NCEP/NCAR and MERRA-2 indicate that over 70% of all counted ARs per year occur during this part of the year. In addition, most strong ARs ($750 < IVT < 1000$ kg $m^{-1}s^{-1}$) and all extreme ARs ($IVT > 1000$ kg $m^{-1}s^{-1}$) considered during the study period of 2005 – 2016 occur during the cool season (Figures 3 and 6).

Rutzetal2014 catalogs based on different reanalyses show excellent agreement for AR event identifications during the peak season of the 2006 water year (Figure 8). Most of the disagreements occur at the start and end times of the events. Allowing the time step to be either ±6 hours from the other catalogs increases the agreement between the reanalysis datasets by about 18% (Table 4). The IVT time steps identified agree over 80% of the time for MERRA-IVT values greater than 400 kg $m^{-1}s^{-1}$, and 100% of the time for MERRA-IVT values greater than 700 kg $m^{-1}s^{-1}$ (Figure 9).

Differences between the three reanalyses show that AR time step frequency, number of events per year, and event duration decrease slightly with increasing resolution, while higher resolution reanalyses observe greater peak IVT. Correlations of AR and non-AR timesteps between different catalogs are lowest for MERRA-2 and NCEP, which are the reanalyses with the most different resolutions (Table 4).

The entire time series for all available reanalyses show that there is substantial agreement during the overlapping years (Figure 10). AR events, particularly the stronger and longer events, are identified consistently regardless of which reanalysis dataset is used. Differences between the reanalysis datasets are primarily in the intensity and in the timing of the AR. These results are consistent with other studies that have investigated and compared landfalling ARs in different reanalysis datasets (Lavers et al., 2012; Jackson et al., 2016; Guan and Waliser, 2017; Guan et al., 2018). Here, we can consider what information can be provided by the relative agreement between reanalyses. The results from this section provide confidence in the idea that, at least in northern California, even coarse-resolution datasets such as NCAR/NCEP are excellent resources for understanding AR activity through time, especially as this dataset goes back in time to 1948, much further back than the others. In the NCEP/NCAR record, drought years correspond with very low AR counts (1977), and record flood years correspond with very high AR counts (1983), with the overall range from under 5 ARs to over 30 ARs due to interannual and interdecadal variability. While it is true that new data sources (e.g., satellite) became available for assimilation during the NCEP/NCAR period of record, Gershunov et al. (2017) validated their catalog with respect to possible discontinuities stemming from satellite data assimilation, and found none.

6. Conclusions and Discussion

This study set out to answer a specific question: how many ARs per year hit the Russian River, a vulnerable coastal watershed in northern California, as well as to assess the sensitivity of the answer to different AR detection algorithms and reanalysis datasets. The results highlight the benefits and challenges in using specific ARDTs to study ARs. Each method is based on expertdeveloped criteria that must be understood to fully appreciate and compare the AR frequency, intensity, and duration results for different ARDT catalogs.

The importance of ARs in regions throughout the globe has been well documented, and understanding the differences in detection methods is essential. Individual ARDTs were produced in order to address specific research objectives (e.g., different regions vs. global scale, ocean vs. overland), and to take advantage of different data sources (e.g., satellite, reanalysis, in situ observations), and these objectives informed the criteria that were applied to detecting ARs. Therefore, no single method should be expected to be perfect for every application. This work provides additional context when selecting or designing an ARDT for future studies. It also helps set the stage for the recently developed Atmospheric River Tracking Method Intercomparison Project (ARTMIP), which aims to quantify uncertainties in AR climatology and impacts on a global scale as a result of differences in AR identification and tracking methods, and which is described in more detail in Shields et al. (2018). This study begins to address the physical origins behind the broad variability in counts found so far in ARTMIP.

The study presented here focuses the comparison on one geographic location that is a focus for land-falling ARs and has a unique, high temporal resolution, long-term in situ observational dataset. The ARO dataset generally suffers from too much missing data each year to be a completely reliable tool for yearly totals and overall statistics. However, the high temporal resolution information it provides on individual storms in real time, while not used in this study where the ARO was sub-sampled at the same temporal resolution as the reanalyses, is particularly valuable. The hourly observations can better resolve AR onset at the ARO location, evolution to peak and through the end of AR conditions, and provide high vertical resolution horizontal wind measurements throughout the column. The period compared in this study, WY2005-2016, includes

24

4 years of severe drought as well as two anomalously wet years, (2006 and 2011), which are representative of California's volatile hydrology. Identified ARs can be categorized, as in this work, by strength measured using IVT intensities and durations, or impacts measured by precipitation and streamflow.

These results provide important information in the context of much foundational work that has been completed on ARs, which has shown that a few big ARs in a year can make the difference between drought and a wet water year (e.g. Dettinger et al., 2011). Therefore, knowing how many ARs to expect on average, and what the variance and range can be, is essential from both water management and emergency preparedness standpoints. Previous studies have shown the connection between moisture flux or IVT strength and significant precipitation (e.g., Lavers et al., 2016; Gershunov et al., 2017). Other studies have shown the importance of AR duration on impacts, where the duration may matter as much or more than AR intensity (Ralph et al., 2013; Lamjiri et al., 2017). Orientation at landfall also drives AR impacts, as recently shown by work done in the Russian River watershed (Guirguis et al., in review). While it is out of the scope of this paper to directly consider impacts of ARs such as on precipitation and streamflow totals in depth, we estimated the contribution of precipitation from ARs detected using different AR tracking schemes to total annual precipitation accumulated at BBY. The set of ARDTs with less stringent geometric criteria such as RSR2014, GSR2017 and GW2015 ranged from 55-60% of AR contribution per year, while WNR2013-IVT250 and MBM2016 ranged from 45-53% per year. The ARDTs focused on much stronger ARs, Ralphetal2013-OBS47, SGS2013 and WNR2013-IVT500 contribute roughly 10% of AR-related precipitation per year. Moreover, more than 40% of heavy precipitation and 80% extreme precipitation events are associated with ARs (see Appendix I, Fig. AI-2).

In Section 3, we first applied each ARDT to the dataset used in original publication, and determined the average annual AR count. Excluding those with high-IVT thresholds, the average annual count was 19±7. In Section 4, applying these ARDTs to a single reanalysis yielded an average annual count of 19±4. Moving to a single reanalysis in this exercise did not change the average annual count but did reduce the variability. Including the ARO observations along with MERRA-2 produces an average annual count of 18±5. Using a single ARDT (Rutz et al., 2014) on three different reanalyses of different resolutions resulted in an average November-April (limited season) count of 17±1 ARs (for comparison, the average November-April AR count for different ARDTs applied to one reanalysis is 13±3; this excludes the high-IVT ARDTs). Therefore, a major conclusion of this work is that the choice of reanalysis has much less of an effect on the AR count than does the choice of ARDT. Specifically, analysis of sensitivity of ARs to the detection method (Section 4) and the reanalysis datasets (Section 5) shows that AR catalogs based on different ARDTs applied to the same reanalysis share 70% of interannual AR variability, whereas AR catalogs based on the same detection methodology applied to different reanalyses share 84% of AR variability (see Appendix II, Table AII-1, for details).

When assessing differences between reanalyses, higher temporal resolution generally decreased AR event counts overall (see Section 4 results compared to Section 3) because there was more opportunity for an AR to fall below threshold or not to meet geometric constraints given the same duration requirement. Other studies have reported similar results with respect to spatial resolution (Guan and Waliser, 2017; Blamey et al., 2018). In this work, all higher resolution timesteps were sub-sampled to be equivalent to the coarsest resolution, meaning that the hourly ARO dataset was sampled at every 6 hours in Sections 3 and 5, and every 3 hours in Section 4, while the MERRA-2 3-hourly dataset used in Section 4 was sampled at every 6 hours in Sections 3 and 5. While the sub-sampling helps to address different temporal resolutions of the datasets, it does not alleviate all of the differences impacting instantaneous AR detections at sub-sampled timesteps.

Differences in ARDTs were found predominantly during weaker storms, and in both ARDT and reanalysis comparisons there were timing differences at the beginning and ends of storms. The differences in AR event counts were much larger as a percentage of the mean than differences in overall AR properties such as strength and duration. This study also shows that the detection algorithms used here can be broken into groups or clusters, based upon geometric criteria and intensity. Focusing on these "clusters" of algorithms within the MERRA-2 based catalogs results in average counts of: 21±4 (for those methods with less limiting geometric criteria), 14±3 (for those methods with strict geometric criteria), or 1±1 (for those methods aiming to identify much stronger ARs) (Table 5). The relatively large effect of the geometric constraints has also been documented by algorithm developers (Guan and Waliser, 2015). The catalogs also perform very well and agree with regard to interannual variability.

The fact that the largest difference in ARDT catalogs are between those with stricter geometric requirements points to fundamental differences in the way that ARs are defined. There has been significant discussion in the literature regarding the definition of ARs since the term was first introduced by Zhu and Newell (1994). For example, the differences and relations between ARs, warm conveyor belts, and tropical moisture exports, are important considerations (Dettinger et al., 2015; Ralph et al., 2017). ARs have recently been defined in a general sense in the Glossary of Meteorology (AMS, 2017) after an extensive community input process and discussion of the aforementioned and other considerations (Ralph et al., 2018). In new studies, the specific research question being asked may determine how narrowly one may want to define these features.

However, a contextual understanding of how ARs are defined in the literature through the use of different types of ARDTs is required in order to understand important and relevant findings such as ARs increasing in frequency and/or intensity in future climate (Dettinger, 2011; Lavers et al., 2013; Warner et al., 2015), or comparing conclusions from studies using various ARDTs over different regions of the globe (Baggett et al., 2016; Hagos et al., 2016; Ramos et al., 2016; Lora et al., 2017; DeFlorio et al., 2018; many others), is essential.

To this end, the authors would like to stress that a single ARDT cannot be recommended universally. Beyond what is described here, other ARDTs continue to be developed, including those based on machine learning, and these techniques should be evaluated as well to understand how they compare with other objective methods. Different ARDTs perform differently by design, and the ARDT should be selected with thoughtful consideration of the particular application. For example, to study changes in the future precipitation regime, it is reasonable to choose ARDTs designed to catch landfalling ARs and their geometric characteristics at landfall. Moreover, it is preferable to use those ARDTs whose outputs were validated on precipitation over various historical periods (e.g. Gershunov et al., 2017; Rutz et al., 2014). If the main focus of the study is moisture transport from the tropics to high latitudes, ARDTs developed for global applications should be used. Ensemble methods, using either perturbations of ARDT thresholds and/or a range of ARDTs, may also be appropriate.

In terms of climate change, additional analysis of GCM realism with respect to simulated AR activity is needed. The authors note that work applying ARDTs to GCMs is ongoing, and that the results from using different ARDTs may be used to illuminate details of how ARs may change in the future. For example, Lavers et al. (2015) shows that IVT increases everywhere thermodynamically as expected in a warmer climate. This means that qualitatively any of the ARDTs will result in the same signal, but details may vary.

Appendix I. Estimating the Precipitation Contribution from ARs Identified with Different ARDTs

To quantify the impact of AR activity detected using different AR tracking schemes on precipitation regime at BBY, the contribution of AR-related precipitation to total annual precipitation accumulated at the area (certain grid cell) during the water years of 2005 – 2013 (Table AI-1) was estimated using Livneh's (2013) precipitation dataset. Precipitation during AR days, defined as days with at least one 3-hour time step associated with AR conditions, and the day after an AR day are counted. The set of ARDTs with the least strict criteria such as RSR2014, GSR2017 and GW2015 ranged from 55-60% of AR contribution per year, while WNR2013-IVT250 and MBM2016 ranged from 45-53% per year. The ARDTs focused on much stronger ARs, Ralphetal2013-OBS47, SGS2013 and WNR2013-IVT500 contribute roughly 10% of AR-related precipitation per year. The annual behavior of AR-related precipitation contribution is illustrated in Figure AI-1. In particular, during wet years such as 2006 the contribution of AR-related precipitation was as much as 70% for ARDT outputs with the least strict criteria, whereas strict AR detection schemes account for up to 30% of the contribution. During dry years both AR activity (Figure 7 from the main text) and AR precipitation contribution (Figure AI-1) are about 25% lower.

Proving the statement on the connection of AR strength and precipitation intensity (Section 6) we estimated the contribution of AR precipitation to all precipitation summed in the different percentile categories (Figure AI-2). The results show that in general moderate to extreme precipitation accumulations are most likely to be associated with AR events. Namely, more than 40% of heavy precipitation and 80% extreme precipitation events are associated with ARs. The catalogs focused on much stronger ARs (SGS2013, WNR2013-IVT500 and Ralphetal2013-OBS47) tend to catch predominantly heavy and extreme precipitation cases. The set of ARDT outputs based on simpler (or no) geometric characteristics (GSR2017, RSR2014, GW2015) cover a wider spectrum of precipitation events.

Table AI-1. Annual average contribution of AR-related precipitation to all precipitation.

AR data source	**Annual average AR precipitation contribution (%)**
GSR2017-MERRA2	56 %
GW2015-MERRA2	55 %
MBM2016-MERRA2	45 %
Ralphetal2013-OBS	39 %
RSR2014-MERRA2	60 %
WNR2013-IVT-MERRA2	53 %
WNR2013-IWV-MERRA2	37 %
Summary	***49 %***

Ralphetal2013-OBS47*	7 %
SGS2013*-MERRA2	15 %
WNR2013-IVT500*-MERRA2	8 %
Summary*	***10 %***

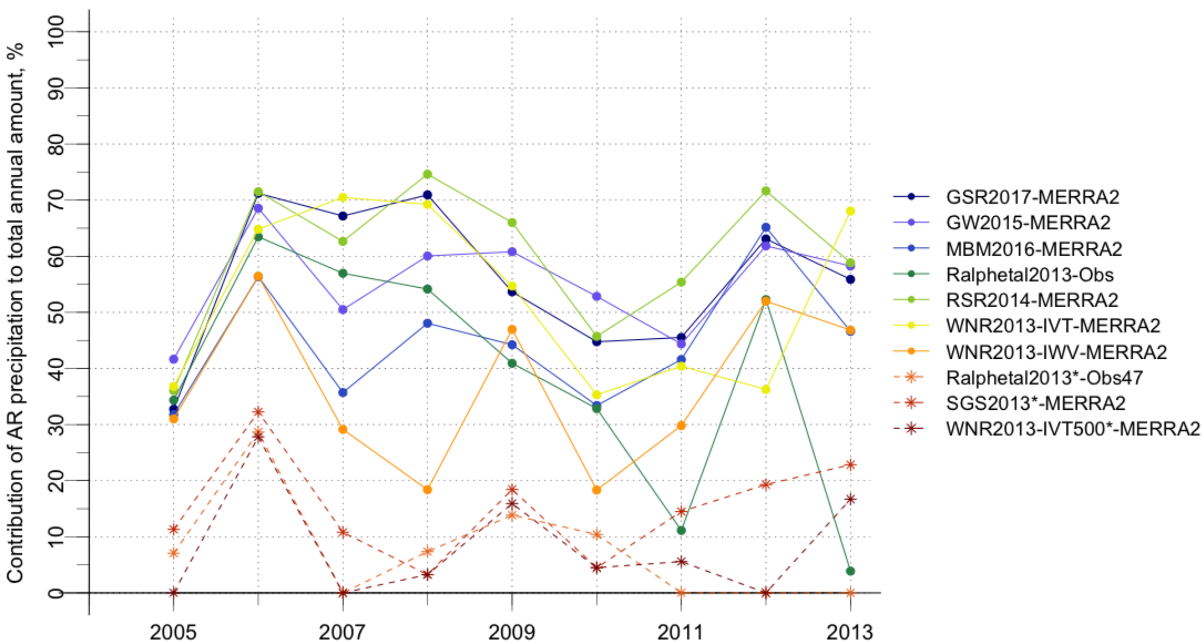


Figure AI-1. Annual average contribution of precipitation associated with AR events counted by each MERRA2-based AR catalog at the grid cell containing BBY during water years 2005-2013.

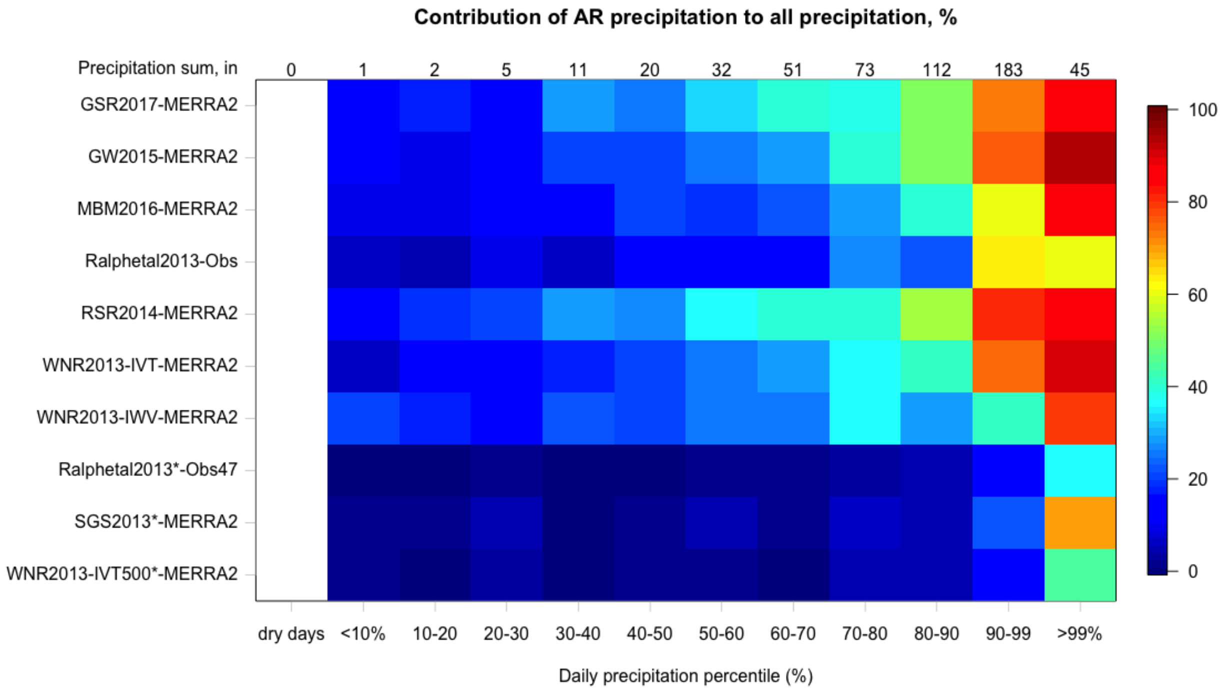


Figure AI-2. Contribution of precipitation associated with AR days in different daily precipitation percentile categories counted by each MERRA2-based AR catalog at the BBY grid cell during water years 2005-2013.

Appendix II. Quantification of the Difference between Choice of ARDT and Choice of Reanalysis

Sensitivity of AR frequency, duration and intensity to the detection methodology (the reanalysis datasets) is quantified by the amount of shared variance in AR catalogs obtained from applying different (the same) detection algorithms to the same (different) reanalysis dataset. The percentage of shared variance represented by square of average correlation coefficient between pairs of AR catalogs shows the amount overlap variation of those catalogs. Two sets of catalogs are considered: six MERRA2-based AR catalogs developed using GSR2017, GW2015, RSR2014, MBM2016, WNR2013-IVT and WNR2013-IWV with solid/percentile-based IVT/IWV thresholds and with/without geometry characteristics at AR detection schemes (see Section 4), and three AR catalogs obtained from applying RSR2014 algorithm to NCEP/NCAR, ERA-Interim and MERRA-2 reanalysis datasets (see Section 5) with different spatial and temporal resolutions. The number of AR events, their average duration and IVT intensity were computed from November through April during 2005-2010 water years according to data availability in considered AR catalogs. The results (Table AII-1) show that the AR catalogs based on different ARDTs applied to the same reanalysis share 70% of interannual AR variability, whereas AR catalogs based on the same detection method applied to different reanalyses share 84% of AR variability. This illustrates that the choice of reanalysis has about 14% less of an effect on AR frequency than does the choice of ARDT. Shared variations in average duration and IVT intensity of different reanalysis based ARs are 14% and 20% higher, respectively.

Table AII-1. Shared variance in AR catalogs obtained from applying GSR2017, GW2015, RSR2014, MBM2016 and WNR2013-IVT AR detection algorithms to MERRA2 Reanalysis dataset and RSR2014 algorithm to NCEP/NCAR, ERA-Interim and MERRA-2 reanalysis datasets.

Shared Variance in AR catalogs from:	**Seasonal number of AR events**	**Seasonal average duration of AR event**	**Seasonal average of AR event MERRA2-IVT**
Different methods to single Reanalysis	70 %	71 %	25 %
Single method to Different reanalysis	84 %	85 %	46 %
Difference in shared variance	***14 %***	***14 %***	***20 %***

Acknowledgements

This research was supported by grant number W912HZ-15-2-0019 from the US Army Corps of Engineers. It was partially supported, with authors Alexander Gershunov and Tamara Shulgina, by Department of the Interior via the Bureau of Reclamation (USBR-R15AC00003, Seasonal and extended-range predictability of atmospheric rivers and their associated precipitation) and by the California Department of Water Resources (4600010378 UCOP2-11, Development of seasonal outlooks for Atmospheric Rivers). The Atmospheric River Observatory data were provided by NOAA's Physical Sciences Division, Earth System Research Laboratory, from their website at www.esrl.noaa.gov/psd/. The authors would like to thank UC San Diego Qualcomm/Calit2 and Pacific Research Platform (ACI-1541349) for supporting the PRP/CONNECT pilot project with network data transfer and storage support for the Sellars et al., 2017b dataset. The authors would also like to thank two anonymous reviewers for their comments that helped us to strengthen the paper.

References

AMS, 2017: Definition of "atmospheric river" in the Glossary of Meteorology, http://glossary.ametsoc.org/wiki/Atmospheric_river.

Baggett, C., S. Lee, and S. Feldstein, 2016: An investigation of the presence of atmospheric rivers over the north Pacific during planetary-scale wave life cycles and their role in Arctic warming. *J. Atmos. Sci.*, **73**, 4329-4347.

Blamey, R., A. Ramos, R. Trigo, R. Tomé, and C. Reason, 2018: The influence of atmospheric rivers over the South Atlantic on winter rainfall in South Africa. *J. Hydrometeor.*, doi:10.1175/JHM-D-17-0111.1.

Brands, S., J. M. Gutiérrez, and D. San-Martín, 2017: Twentieth-century atmospheric river activity along the west coasts of Europe and North America: Algorithm formulation, reanalysis uncertainty and links to atmospheric circulation patterns. *Clim. Dyn.*, 48, 2771-2795.

Dee, D.P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q.J.R. Meteorol. Soc.*, **137(654)**, 553-597.

DeFlorio, M., D. Waliser, B. Guan, D. Lavers, F. Ralph, and F. Vitart, 2018: Global assessment of atmospheric river prediction skill. *J. Hydrometeor.*, **19,** 409-426.

Dettinger, M.D., F.M. Ralph, and D. Lavers, 2015: Setting the stage for a global science of atmospheric rivers. *Eos, Trans. Amer. Geophys. Union*, **96**, doi:10.1029/2015EO038675.

Dettinger, M.D., 2013: Atmospheric rivers as drought busters on the U.S. West Coast. *J. Hydrometeor.*, **14**, 1721-1732.

Dettinger, M.D., 2011: Climate change, atmospheric rivers, and floods in California – A multimodel analysis of storm frequency and magnitude changes. *J. Am. Water Resour. Assoc.*, **47**, 514-523.

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*, **3**(2), 445-478.

Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *J. Climate*, **30**, 5419-5454.

Gershunov A., T.M. Shulgina, F.M. Ralph, D. Lavers, and J.J. Rutz, 2017: Assessing climate-scale variability of atmospheric rivers affecting western North America. *Geophys. Res. Lett.*, **44**, doi:10.1002/2017GL074175.

Guan, B., and D.E. Waliser, 2017: Atmospheric rivers in 20 year weather and climate simulations: A multi-model, global evaluation. *J. Geophys. Res. Atmos.*, **122**, 5556-5581.

Guan, B., and D.E. Waliser, 2015: Detection of atmospheric rivers: Evaluation and application of an algorithm for global studies. *J. Geophys. Res. Atmos.*, **120**, 12514-12535.

Guan, B., D. E. Waliser, and F. M. Ralph, 2018: An inter-comparison between reanalysis and dropsonde observations of the total water vapor transport in individual atmospheric rivers. *J. Hydrometeor.,* **19**, 321–337.

Guirguis, K., A. Gershunov, T.M. Shulgina, R.E.S. Clemesha, and F.M. Ralph, 2018: Atmospheric rivers impacting northern California and their modulation by a variable climate. *Clim. Dyn.*, **in review.**

Hagos, S.M., L.R. Leung, J.-H. Yoon, J. Lu, and Y. Gao, 2016: A projection of changes in landfalling atmospheric river frequency and extreme precipitation over western North America from the Large Ensemble CESM simulations. *Geophys. Res. Lett.*, **43**, 1357-1363.

Hsu, K., X. Gao, and S. Sorooshian, 1997: Precipitation estimation from remotely sensed information using artificial neural networks. *J. Appl. Meteor.*, **36**, 1176-1190.

Jackson, D.L., M. Hughes, and G.A. Wick, 2016: Evaluation of landfalling atmospheric rivers along the U.S. West Coast in reanalysis data sets. *J. Geophys. Res. Atmos.*, **121**, 2705-2718.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteorol. Soc.*, **77(3)**, 437-471.

Lamjiri, M.A., M.D. Dettinger, F.M. Ralph, and B. Guan, 2017: Hourly storm characteristics along the U.S. West Coast: Role of atmospheric rivers in extreme precipitation. *Geophys. Res. Lett.*, **44**, doi:10.1002/2017GL074193.

Lavers, D.A., D.E. Waliser, F.M. Ralph, and M.D. Dettinger, 2016: Predictability of horizontal water vapor transport relative to precipitation: Enhancing situational awareness for forecasting Western U.S. extreme precipitation and flooding. *Geophys. Res. Lett.*, **43**, 2275-2282.

Lavers, D.A., F.M. Ralph, D.E. Waliser, A. Gershunov, and M.D. Dettinger, 2015: Climate change intensification of horizontal water vapor transport in CMIP5. *Geophys. Res. Lett.*, **42**, 5617-5625.

Lavers, D.A., and G. Villarini, 2015: The contribution of atmospheric rivers to precipitation in Europe and the United States. *J. Hydrol.*, **522**, 382-390.

Lavers, D.A., R.P. Allan, G. Villarini, B. Lloyd-Hughes, D.J. Brayshaw, and A.J. Wade, 2013: Future changes in atmospheric rivers and their implications for winter flooding in Britain. *Environ. Res. Lett.*, **8**, 034010.

Lavers, D.A., G. Villarini, R.P. Allan, E.F. Wood, and A.J. Wade, 2012: The detection of atmospheric rivers in atmospheric reanalyses and their links to British winter floods and the large-scale climatic circulation. *J. Geophys. Res.*, **117**, D20106.

Livneh, B., E.A. Rosenberg, C. Lin, B. Nijssen, V. Mishra, K.M. Andreadis, E.P. Maurer, and D.P. Lettenmaier, 2013: A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: update and extensions. *J. Climate*, **26**, 9384-9392.

Lora, J.M., J.L. Mitchell, C. Risi, and A.E. Tripati, 2017: North Pacific atmospheric rivers and their influence on western North America at the Last Glacial Maximum. *Geophys. Res. Lett.*, **44**, 1051-1059.

Mahoney, K., D.L. Jackson, P.J. Neiman, M. Hughes, L. Darby, G. Wick, A.B. White, E. Sukovich, and R. Cifelli, 2016: Understanding the role of atmospheric rivers in heavy precipitation in the southeast United States. *Mon. Wea. Rev.*, **144**, 1617-1632.

Mundhenk, B.D., E.A. Barnes, and E.D. Maloney, 2016: All-season climatology and variability of atmospheric river frequencies over the North Pacific. *J. Climate*, **29**, 4885-4903.

Ralph, F.M., P.J. Neiman, and G.A. Wick, 2004: Satellite and CALJET aircraft observations of atmospheric rivers over the eastern North Pacific Ocean during the winter of 1997/98. *Mon. Wea. Rev.*, **132**, 18721-1745.

Ralph, F.M., P.J. Neiman, G.A. Wick, S.I. Gutman, M.D. Dettinger, D.R. Cayan, and A.B. White, 2006: Flooding on California's Russian River: Role of atmospheric rivers. *Geophys. Res. Lett.*, **33**, L13801, doi:10.1029/2006GL026689.

Ralph, F.M., and M.D. Dettinger, 2012: Historical and national perspectives on extreme West Coast precipitation associated with atmospheric rivers during December 2010. *Bull. Amer. Meteor. Soc.*, **93**, 783-790.

Ralph, F.M., T. Coleman, P.J. Neiman, R.J. Zamora, and M.D. Dettinger, 2013: Observed impacts of duration and seasonality of atmospheric-river landfalls on soil moisture and runoff in coastal northern California. J. Hydrometeor., **14**, 443-459.

Ralph, F. M., S. F. Iacobellus, P. J. Neiman, J. M. Cordeira, J. R. Spackman, D. E. Waliser, G. A. Wick, A. B. White, and C. Fairall, 2017: Dropsonde observations of total water vapor transport within North Pacific atmospheric rivers. *J. Hydrometeor*. **18**, 2577-2596.

Ralph, F. M., M. D. Dettinger, M. M. Cairns, T. Galarneau, and J. Eylander, 2018: Development of the definition of the term "atmospheric river" for the Glossary of Meteorology. *Bull. Amer. Meteor. Soc.,* **99**, 837-839.

Ramos, A.M., R. Tome, R.M. Trigo, M.L.R. Liberato, and J.G. Pinto, 2016: Projected changes in atmospheric rivers affecting Europe in CMIP5 models. *Geophys. Res. Lett*., **43**, 9315-9323.

Rienecker, M.M., and Coauthors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624-3648.

Rutz, J.J., W.J. Steenburgh, and F.M. Ralph, 2014: Climatological characteristics of atmospheric rivers and their inland penetration over the western United States. *Mon. Wea. Rev*., **142**, 905-921.

Sellars, S., P. Nguyen, W. Chu, X. Gao, K-I. Hsu, and S. Sorooshian, 2013: Computational Earth science: Big data transformed into insight. *Eos, Trans. Amer. Geophys. Union*, **94**, 277-278.

Sellars, S.L., B. Kawzenuk, P. Nguyen, F.M. Ralph, and S. Sorooshian, 2017a: Genesis, pathways, and terminations of intense global water vapor transport in association with large-scale climate patterns. *Geophys. Res. Lett*., **44**, 12465-12475.

Sellars, S.L., P. Nguyen, and B. Kawzenuk, 2017: The CONNected object, or CONNECT algorithm applied to National Aeronautics and Space Administration (NASA) Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA V2) – integrated water vapor from 1980 to 2016. UC San Diego Library Digital Collections. https://doi.org/10.6075/J01834P8.

Shields, C., and Coauthors, 2018: Atmospheric river tracking method intercomparison project (ARTMIP): Project goals and experimental design. *Geosci. Model Dev.*, **11**, 2455-2474.

Trenberth, K.E., J.T. Fasullo, and J. Mackaro, 2011: Atmospheric moisture transports from ocean to land and global energy flows in reanalyses. *J. Clim.*, **24(18)**, 4907-4924.

Waliser, D., and B. Guan, 2017: Extreme winds and precipitation during landfall of atmospheric rivers. *Nature Geosci.*, **10**, 179-183.

Warner, M.D., C.F. Mass, and E.P. Salathe Jr., 2015: Changes in winter atmospheric rivers along the North American West Coast in CMIP5 climate models. *J. Hydrometeor.*, **16**, 118-128.

White, A.B., and co-authors, 2013: A Twenty-First-Century California observing network for monitoring extreme weather events. *J. Atmos. Oceanic Technol.*, **30**, 1585-1603.

Wick, G.A., 2014: Implementation and initial application of an atmospheric river detection tool based on integrated vapor transport. *2014 Fall Meeting*, San Francisco, CA, Amer. Geophys. Union, Abstract A34E-06.

Wick, G.A., P.J. Neiman, and F.M. Ralph, 2013: Description and validation of an automated objective technique for identification and characterization of the integrated water vapor signature of atmospheric rivers. *IEEE Trans. Geosci. Remote Sens.*, **51**, 2166-2176.

Wick, G.A., P.J. Neiman, F.M. Ralph, and T.M. Hamill, 2013: Evaluation of forecasts of the water vapor signature of atmospheric rivers in operational numerical weather prediction models. *Wea. Forecasting*, **28**, 1337-1352.

Young, A.M., K.T. Skelly, and J.M. Cordeira, 2017: High-impact hydrologic events and atmospheric rivers in California: An investigation using the NCEI Storm Events Database. *Geophys. Res. Lett.*, **44**, 3393-3401.

Zhu, Y., and R. E. Newell, 1994: Atmospheric rivers and bombs. *Geophys. Res. Lett.,* **21,** 1999–2002, doi:10.1029/94GL01710.

Zhu, Y., and R.E. Newell, 1998: A proposed algorithm for moisture fluxes from atmospheric rivers. *Mon. Wea. Rev.*, **126**, 725-735.

TABLES

Table 1. The variables and geometric thresholds used in each AR detection method considered throughout the study. In methods 1 and 6 (Gershunov et al. 2017; Sellars et al. 2013), the AR object must persist for 24 hours, but may not last that long in an individual grid cell. Starred lines indicate that the method was modified for use in this study. Please see the text for details.

ARDT		**Detection parameter**	**AR detection thresholds**	
			IVT, IWV	**Spatial, temporal**
a) Ralphetal2013-Obs* b) Ralphetal2013-Obs47*	Ralph., F.M. et al., 2013	IWV, IWV flux	a) WVF ≥ 20 cm m/s, IWV ≥ 20 mm b) WVF ≥ 47 cm m/s, IWV ≥ 20 mm	**12 hours**
SGS2013*	Sellars, S. et al., 2013	IVT	IVT ≥ 500 kg/m/s	24 hours
a) WNR2013-IVT* b) WNR2013-IVT500*	Wick, G. et al., 2013	IVT	a) IVT ≥ 250 kg/m/s b) IVT ≥ 500 kg/m/s	≥ 1500 km width < 1000 km length/width ≥ 1.4
WNR2013-IWV	Wick, G. et al., 2013	IWV	IWV ≥ 20 mm	≥ 2000 km width < 1000 km
RSR2014	Rutz, J. et al., 2014	IVT	IVT ≥ 250 kg/m/s	≥ 2000 km

GW2015	Guan, B. and D. Waliser, 2015	IVT	Monthly IVT ≥ 85th percentile (~ 166-254 kg/m/s)	≥ 2000 km, Length/width > 2
MBM2016*	Mundhenk, B. et al., 2016	IVT anomaly	IVT anomaly ≥ 94th percentile (IVT ~ 209-283 kg/m/s)	≥ 1400 km length/width ≥ 1.6
GSR2017	Gershunov, A. et al., 2017	IVT, IWV	IVT ≥ 250 kg/m/s IWV ≥ 15 mm	≥ 1500 km 18 hours

Table 2. Statistics for each of the AR catalogs on their native datasets, water years 2005-2015. The baseline AR threshold is separated from those methods that identify ARs beginning at moderate strength(*). Summary statistics for native reanalysis IVT are not presented as IVT is not a default output in some of the ARDTs.

AR data source	**Number of AR events**	**Annual average number of AR events**	**Average duration of AR event (hr)** ± σ	**Average of AR event Native-IVT average (kg/m/s) per 6-hour time step** ± σ	**Average of AR event MERRA2-IVT average (kg/m/s) per 6-hour time step** ± σ	**Average of AR event Native-IVT maximum (kg/m/s) per 6-hour time step** ± σ	**Average of AR event MERRA2-IVT maximum (kg/m/s) per 6-hour time step** ± σ	**Average of AR event MERRA2-IWV average (mm) per 6-hour time step** ± σ	**Average of AR event MERRA2-IWV maximum (mm) per 6-hour time step** ± σ
GSR2017-NCEP	244	22±4	25±3	342±59	343±79	407±115	447±154	22±5	25±6
GW2015-ERAI	264	24±4	25±3	299±81	316±90	376±146	408±155	22±4	26±5
MBM2016-MERRA1	152	14±5	21±2	430±81	429±89	507±140	519±148	25±4	27±5
Ralphetal2013-OBS	114	10±4	24±2	389±126	389±126	485±175	485±175	25±4	28±5
RSR2014-NCEP	279	25±5	25±3	336±58	328±81	398±115	429±154	21±5	25±6
Summary	***210±73***	***19±7***	***24±2***	***-***	***361±47***	***-***	***458±44***	***23±2***	***26±1***
Ralphetal2013-OBS47*	13	1±1	16±1	621±164	621±164	706±191	706±191	27±4	29±4
SGS2013*-MERRA2	34	3±2	18±1	646±98	646±98	721±147	721±147	30±4	32±5

Summary*	***24±15***	**2±1**	***17±1***	***633±18***	***633±18***	***714±11***	***714±11***	***29±2***	***31±2***

Table 3. Statistics for each of the AR catalogs through the 2005-2016 water years, using the MERRA-2 reanalysis dataset. The baseline AR threshold is separated from those methods that identify ARs beginning at moderate strength(*).

AR data source	**Number of AR events (Summary includes ± σ)**	**Annual average number of AR events (Summary includes ± σ)**	**Average duration of AR event (hr) ± σ**	**Average of AR event MERRA2-IVT average (kg/m/s) per 3-hour time step ± σ**	**Average of AR event MERRA2-IVT maximum (kg/m/s) per 3-hour time step ± σ**	**Average of AR event MERRA2-IWV average (mm) per 3-hour time step ± σ**	**Average of AR event MERRA2-IWV maximum (mm) per 3-hour time step ± σ**
GSR2017-MERRA2	257	21 ± 4	24 ± 5	372 ± 74	485 ± 154	23 ± 4	27 ± 5
GW2015-MERRA2	238	20 ± 3	24 ± 5	344 ± 88	455 ± 166	23 ± 4	27 ± 5
MBM2016-MERRA2	152	13 ± 4	22 ± 5	442 ± 80	553 ± 153	25 ± 4	28 ± 5
Ralphetal2013-OBS	131	11 ± 5	23 ± 4	385 ± 120	503 ± 179	25 ± 4	28 ± 5
RSR2014-MERRA2	268	22 ± 5	25 ± 5	369 ± 75	480 ± 154	23 ± 4	27 ± 5
WNR2013-IVT-MERRA2	185	15 ± 3	20 ± 4	394 ± 96	500 ± 155	24 ± 5	27 ± 5
WNR2013-IWV-MERRA2	261	22 ± 6	20 ± 3	299 ± 159	378 ± 202	26 ± 3	28 ± 4
Summary	***213 ± 56***	**18 ± 5**	***23 ± 2***	***372 ± 44***	***479 ± 54***	***24 ± 1***	***27 ± 0.5***
Ralphetal2013-OBS47*	13	1 ± 1	16 ± 2	613 ± 159	733 ± 205	27 ± 4	30 ± 4

SGS2013*-MERRA2	29	2 ± 1	19 ± 3	654 ± 75	778 ± 137	31 ± 4	33 ± 4
WNR2013-IVT500*-MERRA2	14	1 ± 1	20 ± 4	678 ± 124	784 ± 155	31 ± 4	34 ± 5
Summary*	***19 ± 9***	***1*** ± 1	***18 ± 2***	***660 ± 16***	***765 ± 28***	***30 ± 2***	***32 ± 2***

Table 4. (Rows 2–9) Statistics for each of the RSR2014 AR catalogs, using the different reanalysis datasets during Nov-Apr 1990-2010. Resolution for each reanalysis is in parentheses. Linear correlations are based on AR timesteps identified in each reanalysis. See Figure 1 for grid points containing Bodega Bay. The 1436 value of maximum IVT on ERA-Interim and MERRA-2 is from 1200 UTC, 12 December 1995. (Rows 11–17) Overlap in identified AR steps using various time windows for each of the RSR2014 AR catalogs, using the different reanalysis datasets.

	NCEP (2.5)	ERA-Interim (1.5)	MERRA-2 (0.5)
Max IVT	1088	1436	1436
AR Timestep Frequency	10.6%	10.0%	8.5%
AR Events/Nov-Apr *	17.9	17.2	14.8
AR Duration (hr)	25.9	25.4	24.8
Linear Correlations	**NCEP**	**ERA-Interim**	**MERRA-2**
NCEP		.76	.75
ERA-Interim	.76		.81
MERRA-2	.75	.81	
AR Step Overlap	**Exact**	**+/- 6 hr**	**+/- 12 hr**
NCEP Only	16.4%	9.2%	7.7%
ERAI Only	7.8%	3.7%	3.5%
MERRA-2 Only	2.8%	1.2%	0.9%
NCEP and ERAI	62.4%	80.0%	83.4%
NCEP and MERRA-2	58.6%	74.6%	78.6%
ERAI and MERRA-2	69.4%	82.4%	84.2%
All 3 Datasets	49.9%	68.2%	73.3%

*To compare with Table 3, add 4/year (May - Oct storms)

Table 5. AR event counts with ARDT characteristics. ARDTs are sorted by criteria. * indicates those catalogs that are designed to identify only stronger storms; ** indicates observational catalogs with significant missing data during some years; *** indicates catalogs using IWV alone.

ARDT	Avg Annual AR Events	IVT Threshold (kg/m/s)	IWV Threshold (mm)	Geometric (Length, km)	Geometric (Width km, or ratio)	Geometric/ Duration (Other)

RSR2014-NCEP	22±5	250	No	>2000	No	No
GSR2017	21±4	250	15	>1500	No	No
GW2015-ERAI	20±3	166-254	No	>2000	L/W > 2	Yes
WNR2013-IVT* (Section 4)	15±3	250	No	>2000	<1000; L/W > 1.4	Yes
MBM2016	13±3	209-283	No	>1400	L/W > 1.6	Yes
Based on much higher IVT threshold						
SGS2013*	2±1	500	No	No	No	Yes
WNR2013-IVT500*	1±1	500	No	>1500	<1000	Yes
Not based on IVT						
WNR2013-IWV***	22±6	No	20	>1500	<1000	Yes
Ralphetal2013-Obs**	11±5	250 (20 cm m/s)	20	No	No	Yes
Ralphetal-Obs47*,**	1±1	500 (47 cm m/s)	20	No	No	Yes

FIGURES

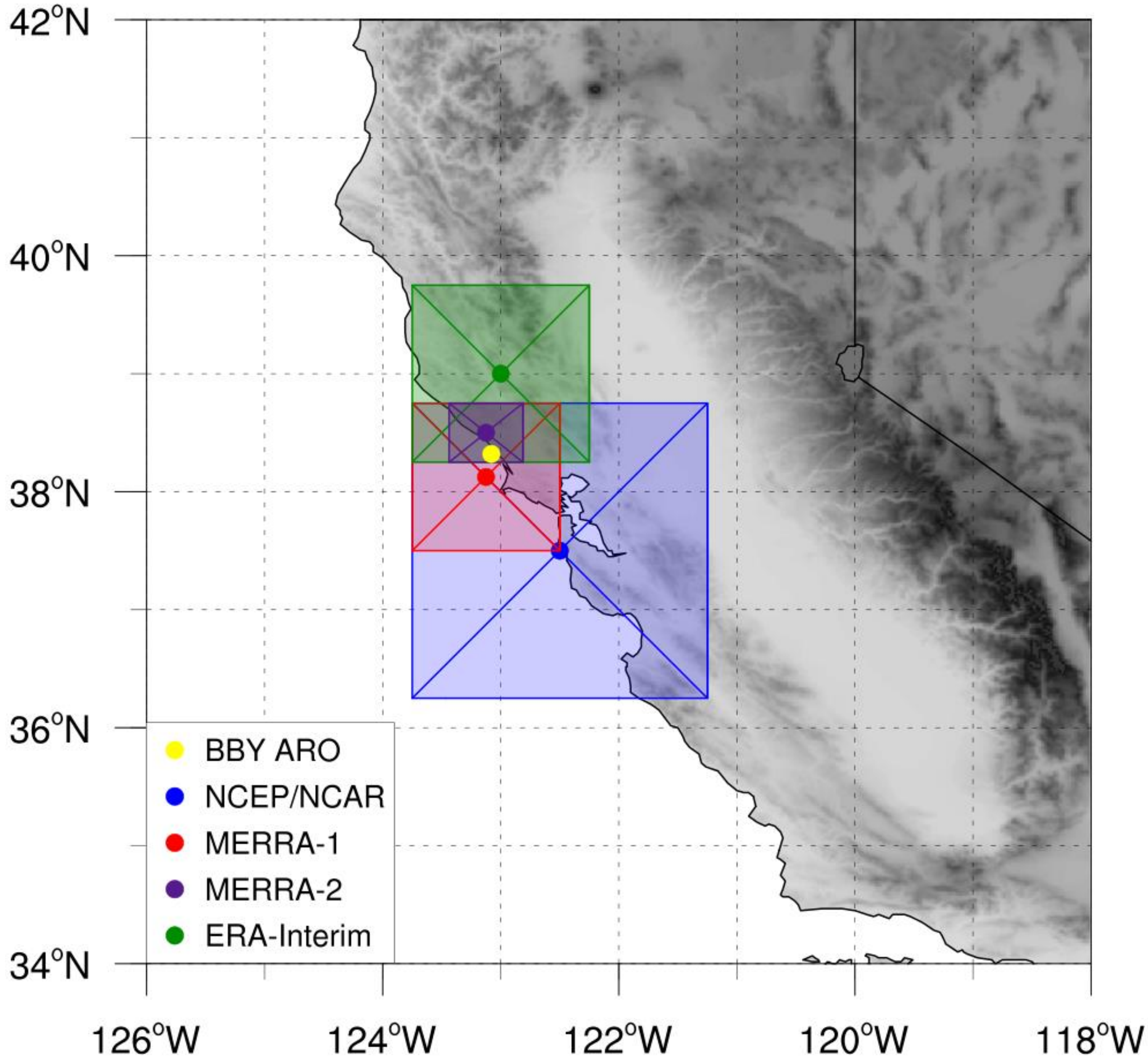


Figure 1. Map of the study region with the Bodega Bay ARO location marked in yellow. Grid center points and boxes for all of the reanalyses used in this study are presented in solid markers and shading (colored according to scale). Terrain represented by gray shading.

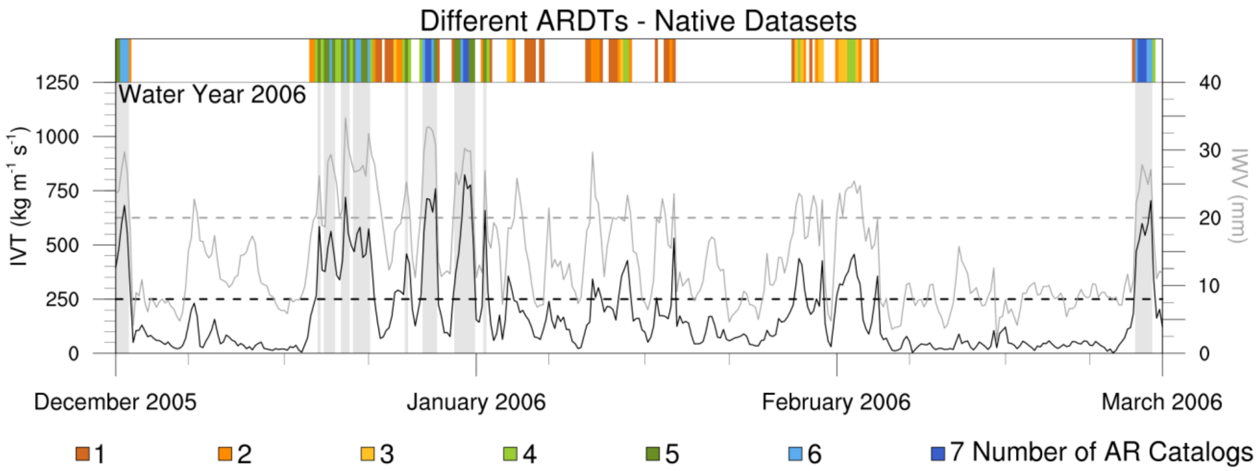


Figure 2. Time series of IVT (black line) and IWV (gray line) from MERRA-2 reanalysis during the peak of the cool season during water year 2006. Color bars at the top indicate how many of the seven considered catalogs on their native reanalysis identified an AR at a specific time. All catalogs besides the Wick catalogs are represented here (see items 1-6 in Table 1). Gray shading indicates agreement of five or more AR catalogs. Two of the seven catalogs are designed to identify moderate and stronger ARs.

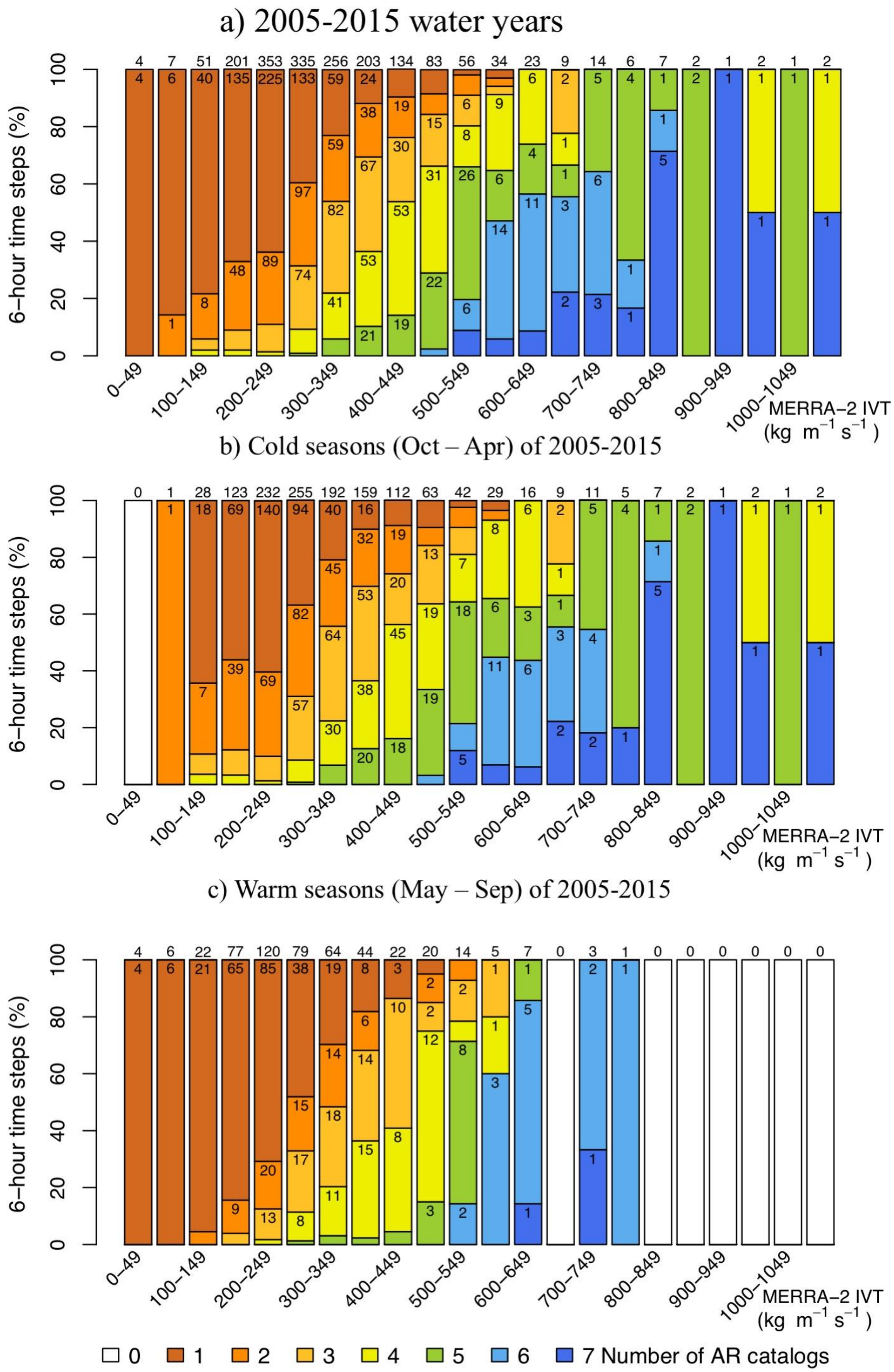

a) 2005-2015 water years
b) Cold seasons (Oct – Apr) of 2005-2015
c) Warm seasons (May – Sep) of 2005-2015
6–hour time steps (%)
MERRA–2 IVT
(kg m^{-1} s^{-1})
0–49
100–149
200–249
300–349
400–449
500–549
600–649
700–749
800–849
900–949
1000–1049
0 1 2 3 4 5 6 7 Number of AR catalogs

Figure 3. Agreement of native reanalysis-based AR catalogs expressed in terms of frequency and **<u>MERRA2-IVT</u>** intensity of 6-hour time steps associated with AR conditions in Bodega Bay during (a) 2005-2015 water years, (b) the cool (October – April) seasons of 2004-2015 and (c) the warm (May - September) seasons of 2005-2015. The number of 6-hour AR time steps is displayed on the top of each bin. The colors represent the number of AR catalogs that shared the AR time steps. Percentage is expressed in terms of the number of total identified time steps in a given IVT bin.

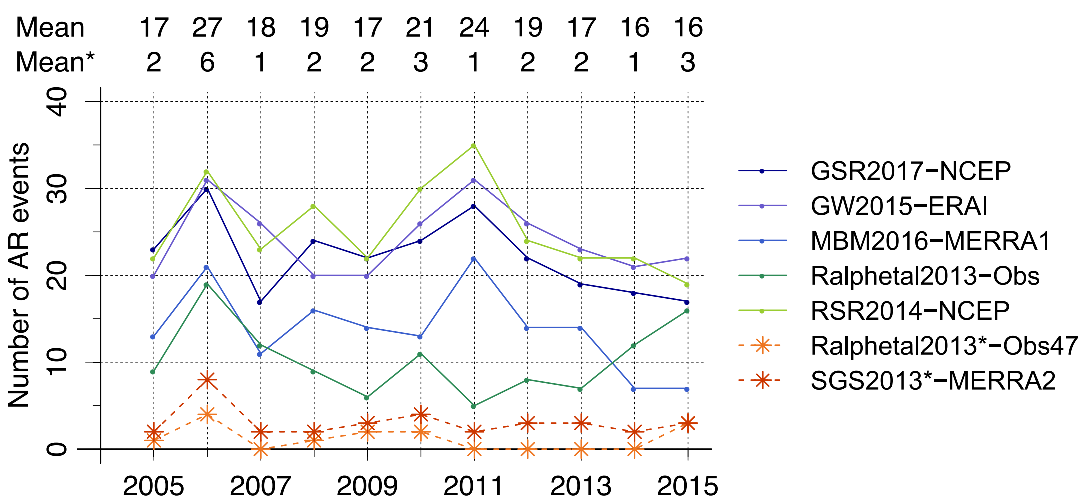


Figure 4. Number of distinct AR events counted by each native reanalysis-based AR catalog at the grid cell containing BBY (see Figure 1) during water years 2005-2015. Dashed lines(*) indicate methods that identify ARs that persist for at least 12 hours with moderate AR strength thresholds.

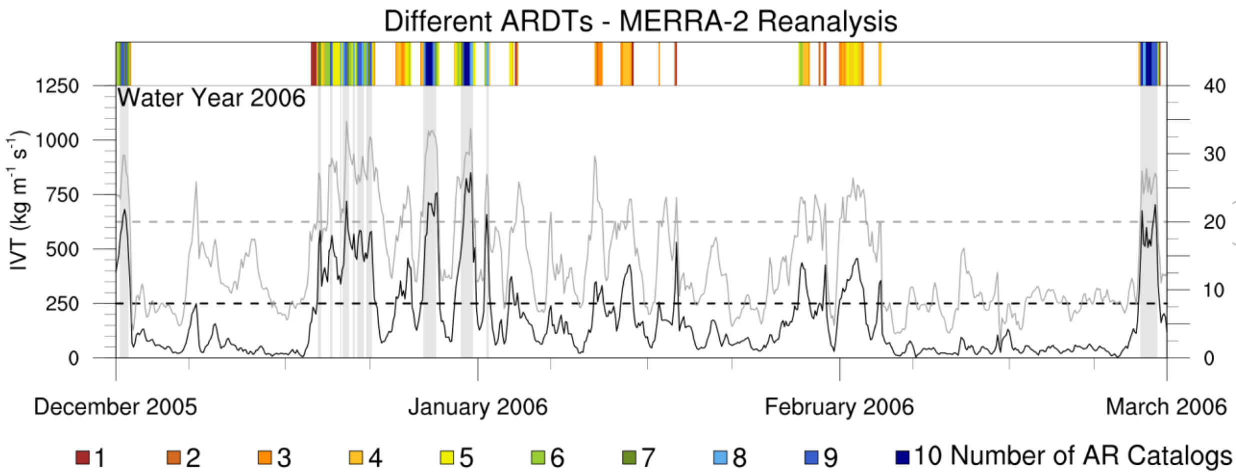


Figure 5. As in Figure 2, except for all ten ARDTs (see Table 1) as applied to MERRA-2. Three of the catalogs are designed to identify only moderate and stronger ARs. Gray shading represents agreement of at least seven catalogs.

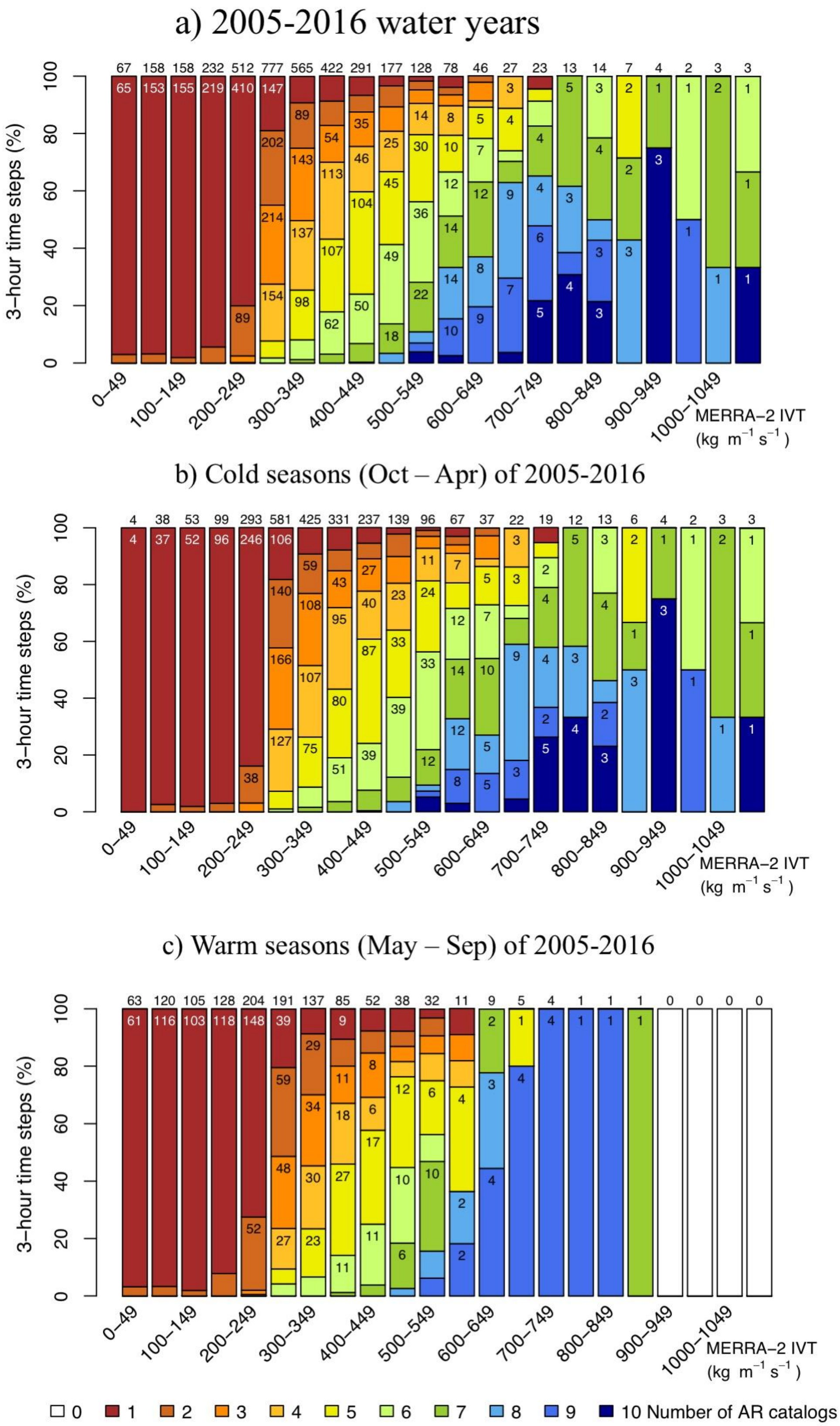

a) 2005-2016 water years
b) Cold seasons (Oct – Apr) of 2005-2016
c) Warm seasons (May – Sep) of 2005-2016
3–hour time steps (%)
0–49
100–149
200–249
300–349
400–449
500–549
600–649
700–749
800–849
900–949
1000–1049
MERRA–2 IVT
(kg m^{-1} s^{-1})
0 1 2 3 4 5 6 7 8 9 10 Number of AR catalogs

Figure 6. Agreement of MERRA-2-based AR catalogs expressed in terms of frequency and IVT-intensity of 3-hour time steps associated with AR conditions in Bodega Bay during (a) 2005-2016 water years, (b) the cool (October – April) seasons of 2004-2016 and (c) the warm (May - September) seasons of 2005-2015. The number of 3-hour AR time steps is displayed on the top of each bin. The colors represent the number of AR catalogs that shared the AR time steps.

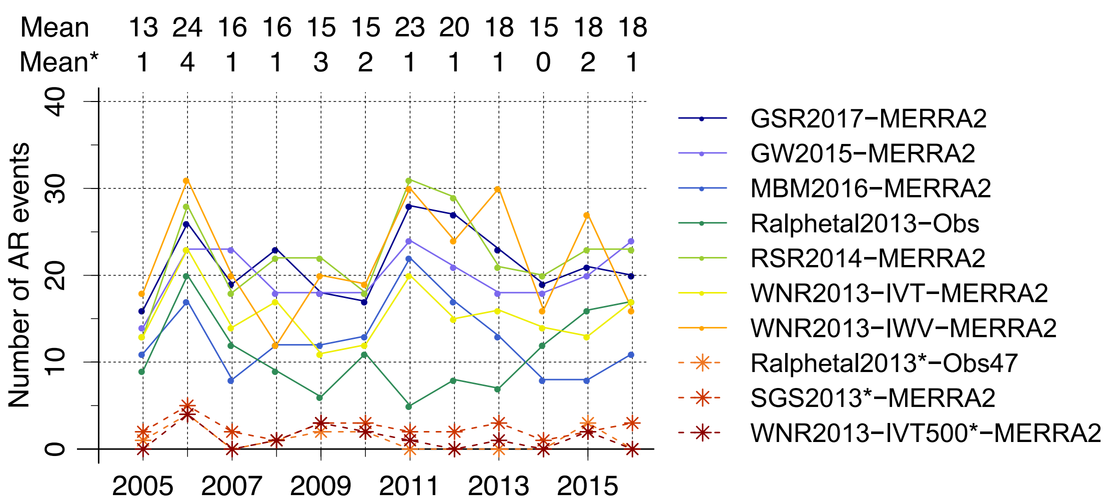


Figure 7. As in Figure 4, but with each ARDT applied to the MERRA-2 dataset.

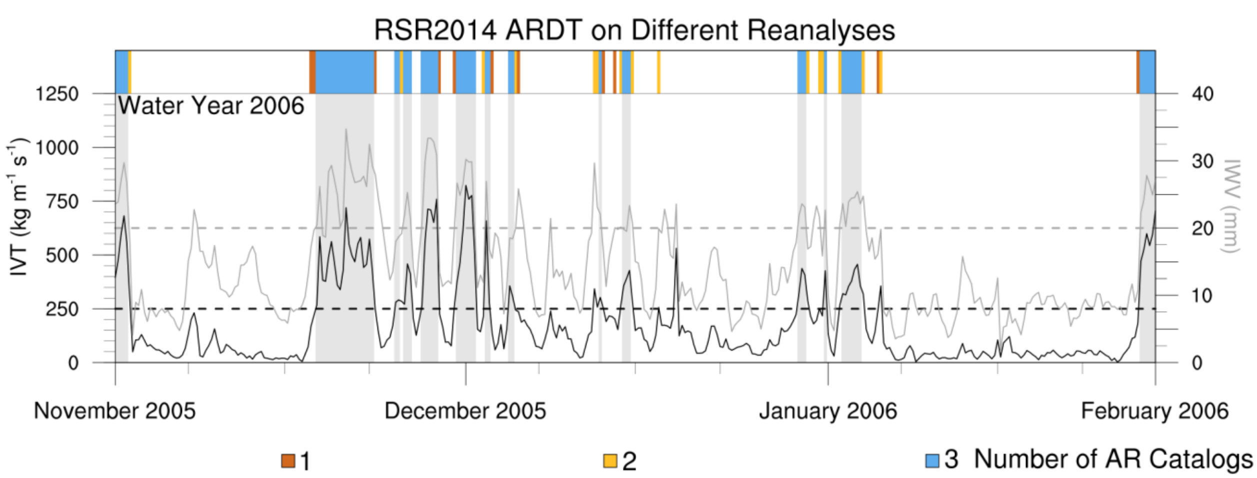


Figure 8. As in Figure 2, except for RSR2014 ARDT applied to 3 different reanalyses (ERA-Interim; NCEP; MERRA-2). Gray shading is present when all three catalogs agree.

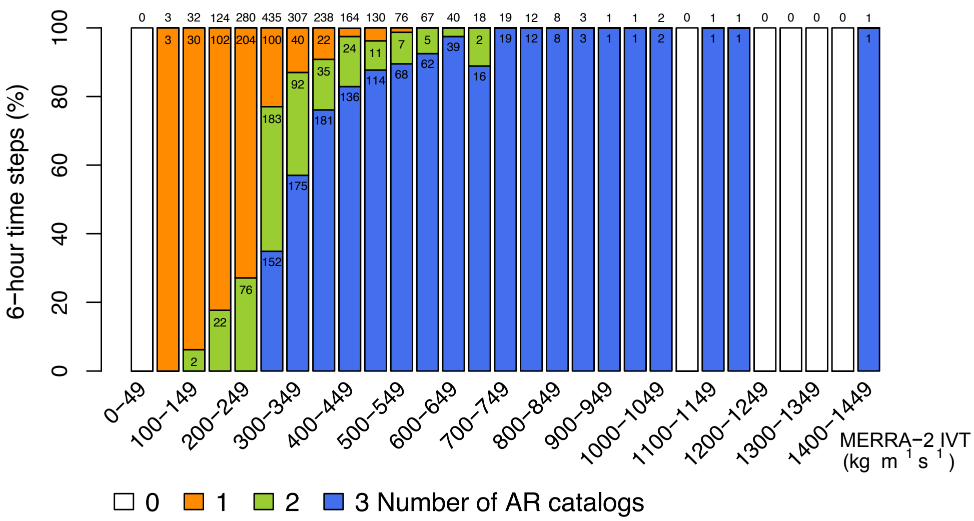


Figure 9. Agreement of Reanalysis-based AR catalogs expressed in terms of frequency and **MERRA2-IVT** intensity of 6-hour time steps associated with AR conditions in Bodega Bay during the cool (November – April) seasons of 2004-2010. The number of 6-hour AR time steps is displayed on the top of each bin. The colors represent the number of AR catalogs that shared the AR time steps.

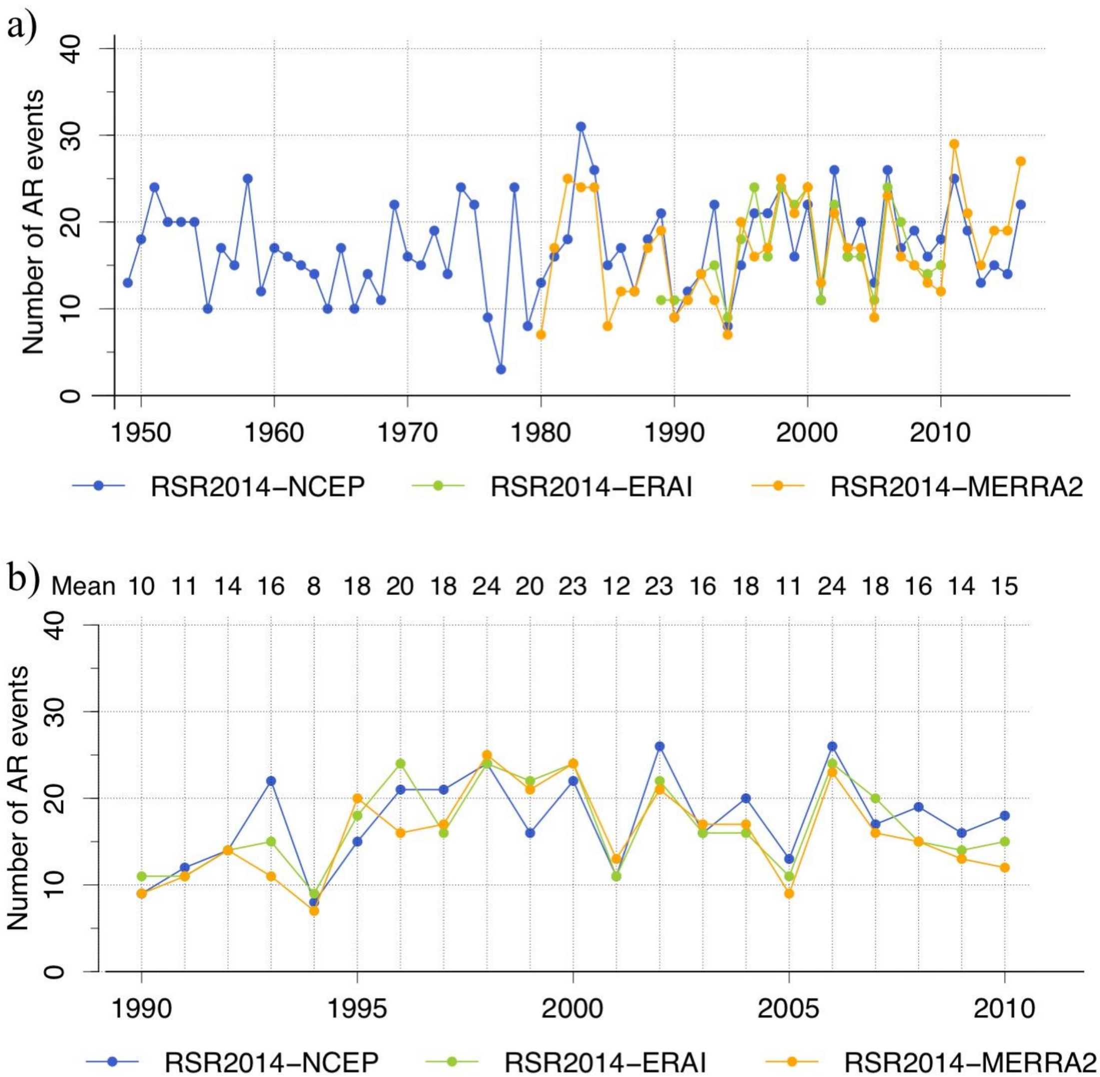


Figure 10. As in Figure 4, except with only RSR2014 applied to different reanalysis datasets during (a) all November – April for all available water years for each reanalysis and (b) November - April during water years 1990-2010, when all three datasets are available.

Figure 1

Click here to download Figure Fig1-GridPoints_Map_BW.png

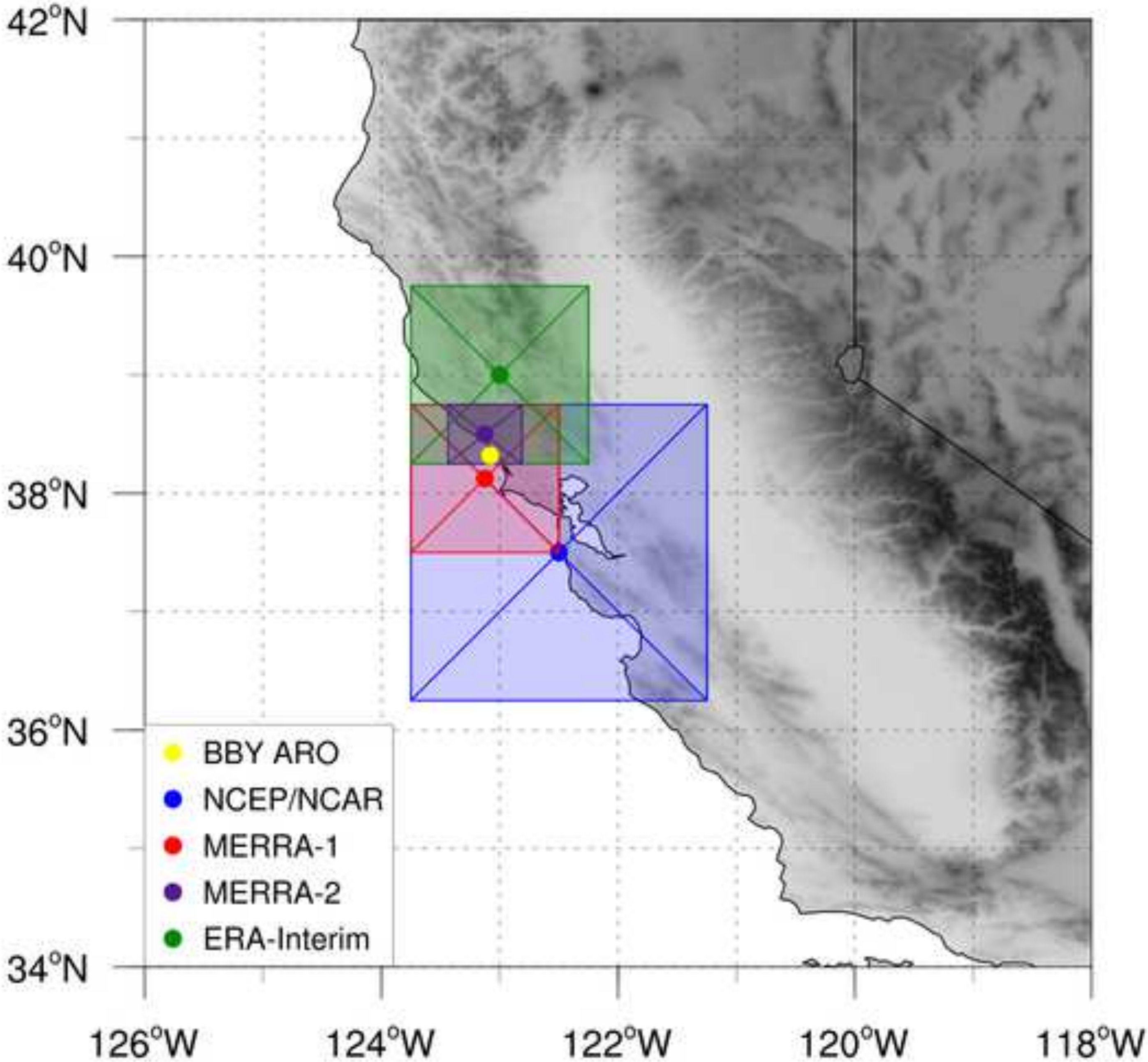


Figure 2

Click here to download Figure Fig2-ARDT_ColorBarEvents_IVT-IWV-Sec3_WY2006_DJF.png

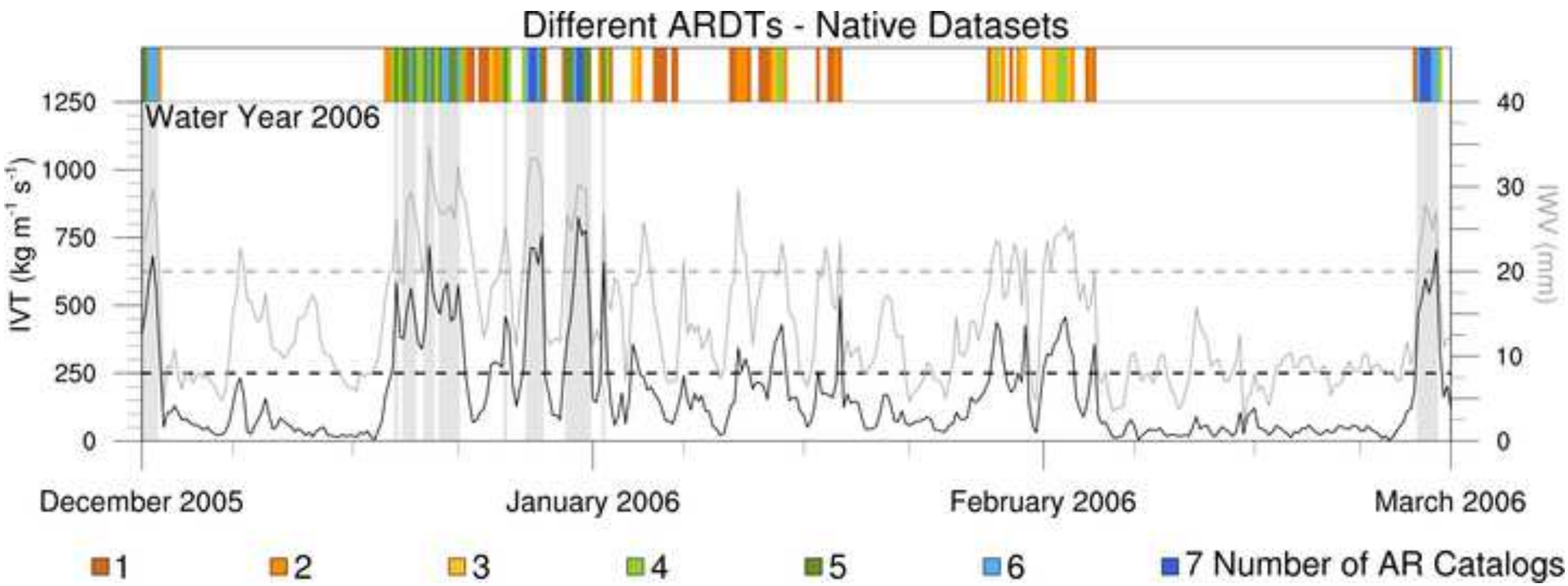


Figure 3 Click here to download Figure ARDT-Fig3.jpg

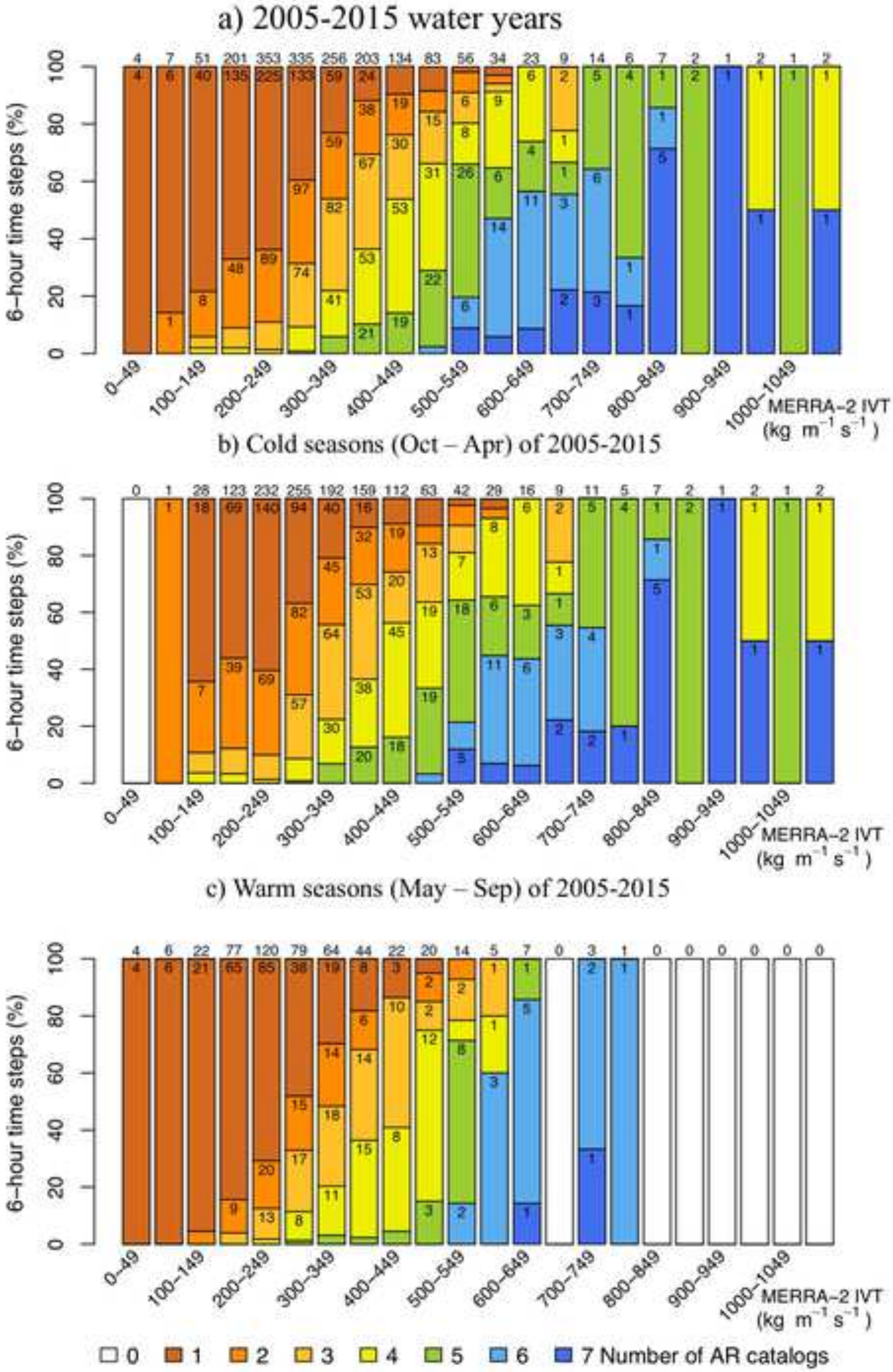


Figure 4

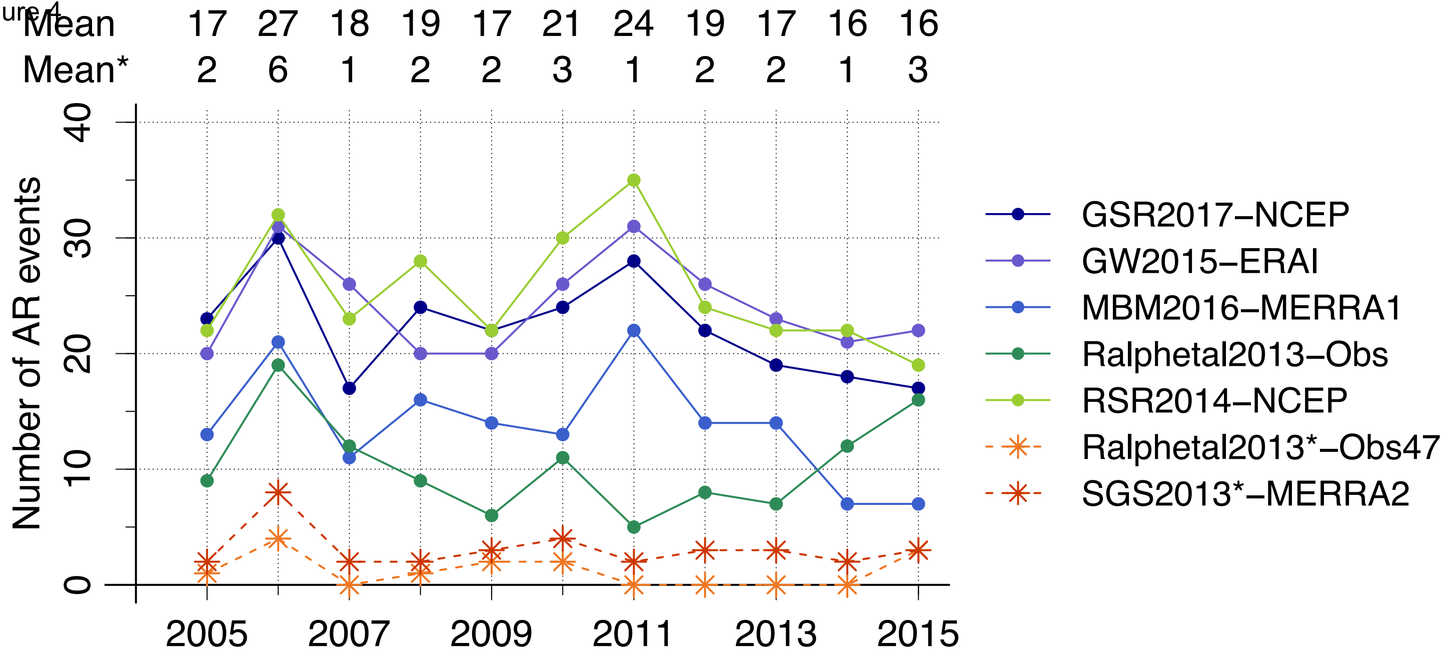

Mean 17 27 18 19 17 21 24 19 17 16 16
Mean* 2 6 1 2 2 3 1 2 2 1 3
Number of AR events
0
10
20
30
40
2005
2007
2009
2011
2013
2015
GSR2017−NCEP
GW2015−ERAI
MBM2016−MERRA1
Ralphetal2013−Obs
RSR2014−NCEP
Ralphetal2013*−Obs47
SGS2013*−MERRA2

Figure 5

Click here to download Figure Fig5-ARDT_ColorBarEvents_IVT-IWV-Sec4_WY2006_DJF.png

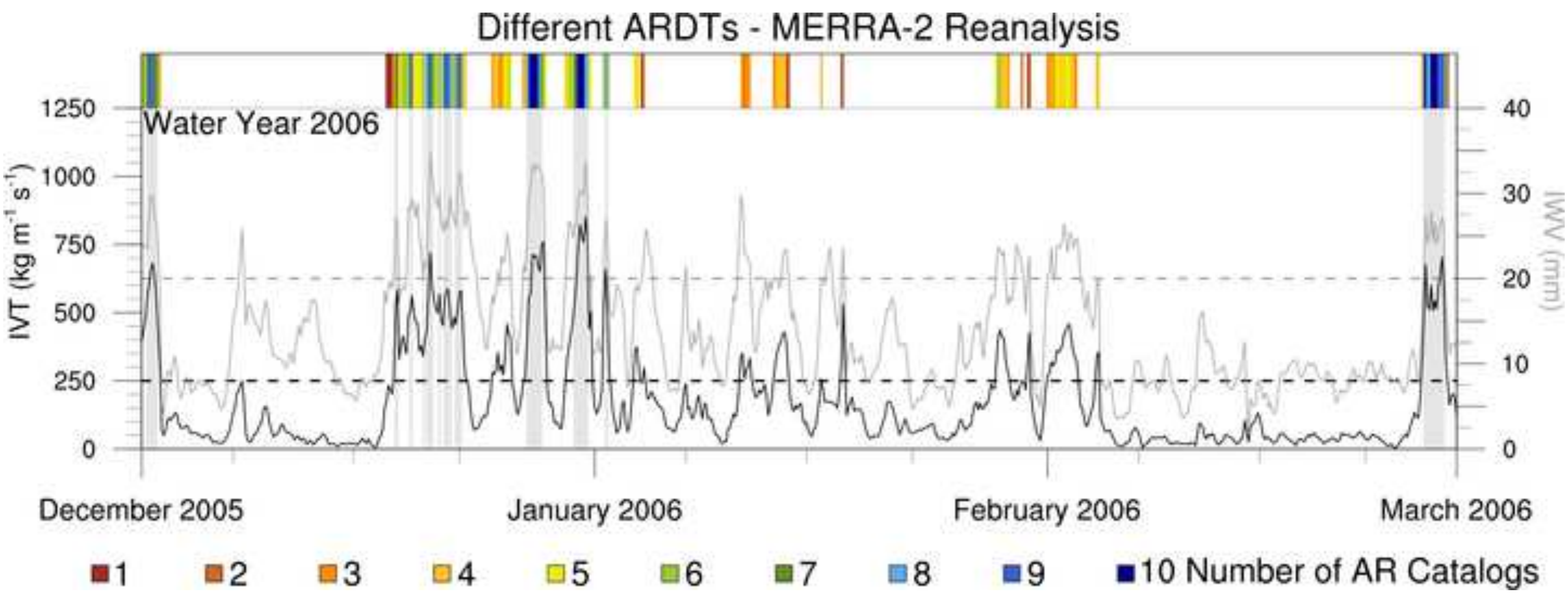


Figure 6

Click here to download Figure Fig6.jpg

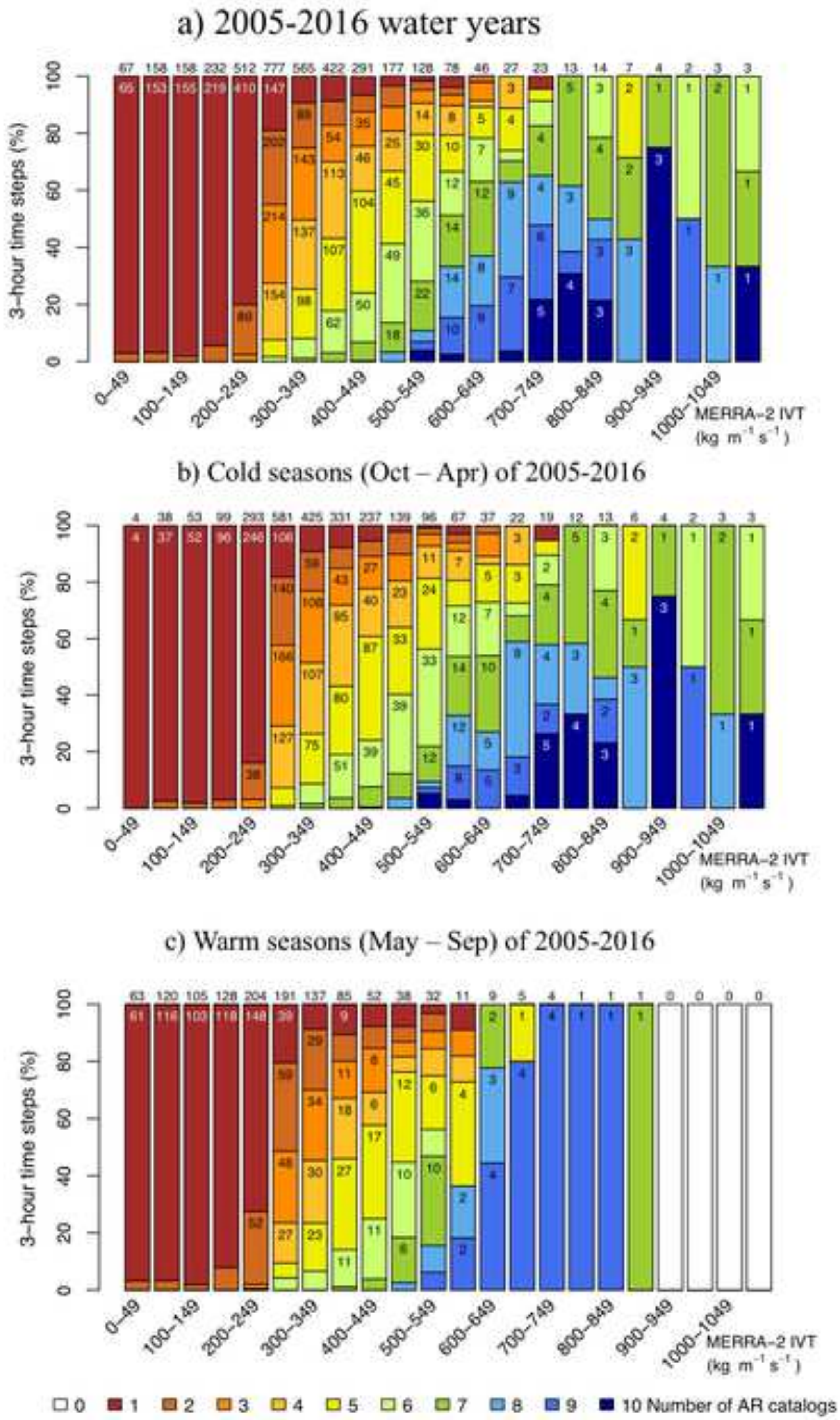
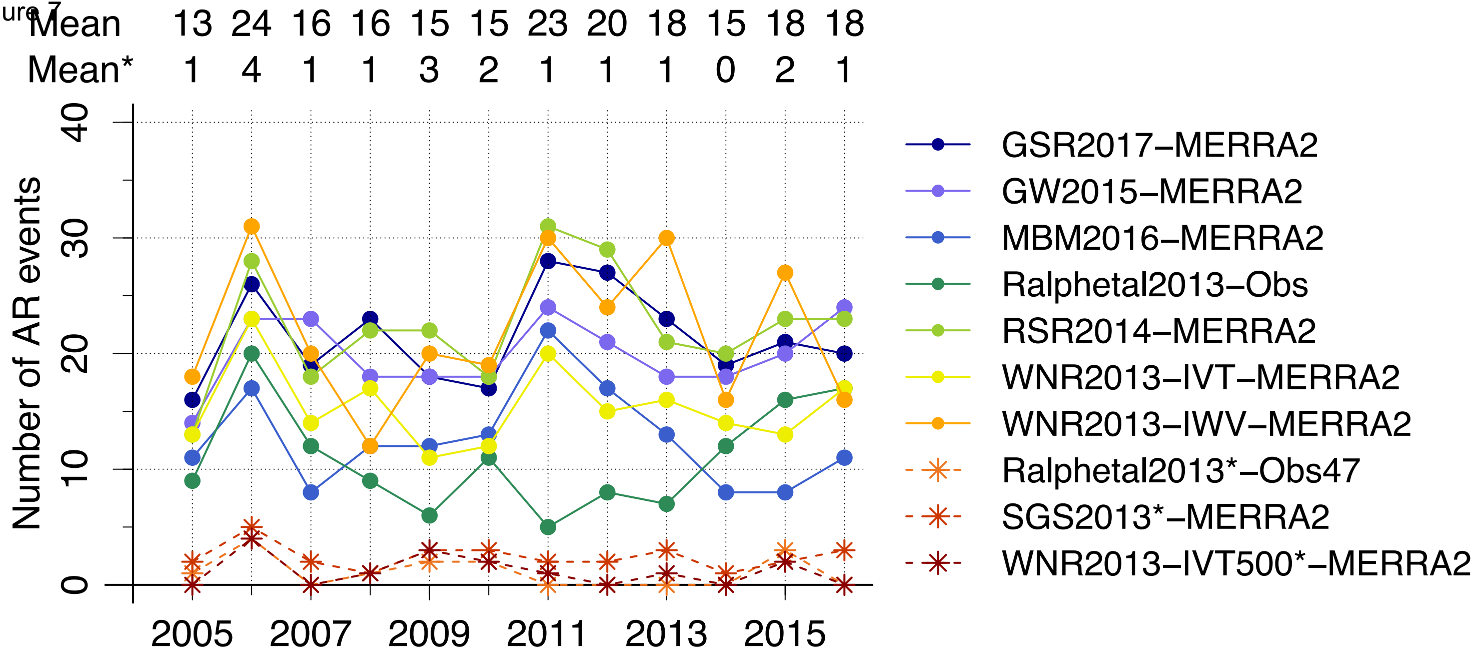


Figure 7



Mean 13 24 16 16 15 15 23 20 18 15 18 18
Mean* 1 4 1 1 3 2 1 1 1 0 2 1
Number of AR events
0
10
20
30
40
2005
2007
2009
2011
2013
2015
GSR2017–MERRA2
GW2015–MERRA2
MBM2016–MERRA2
Ralphetal2013–Obs
RSR2014–MERRA2
WNR2013–IVT–MERRA2
WNR2013–IWV–MERRA2
Ralphetal2013*–Obs47
SGS2013*–MERRA2
WNR2013–IVT500*–MERRA2

Figure 8

Click here to download Figure Fig8-ARDT_ColorBarEvents_IVT-IWV-Sec5_WY2006_DJF.png

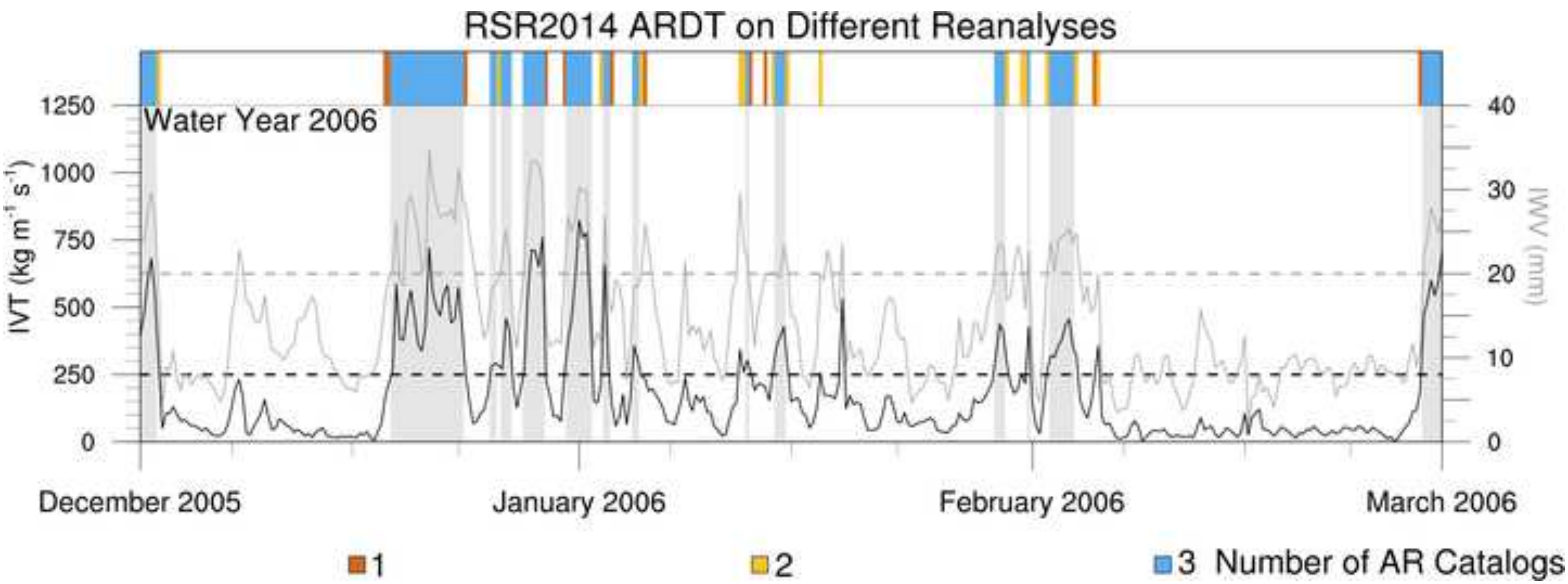
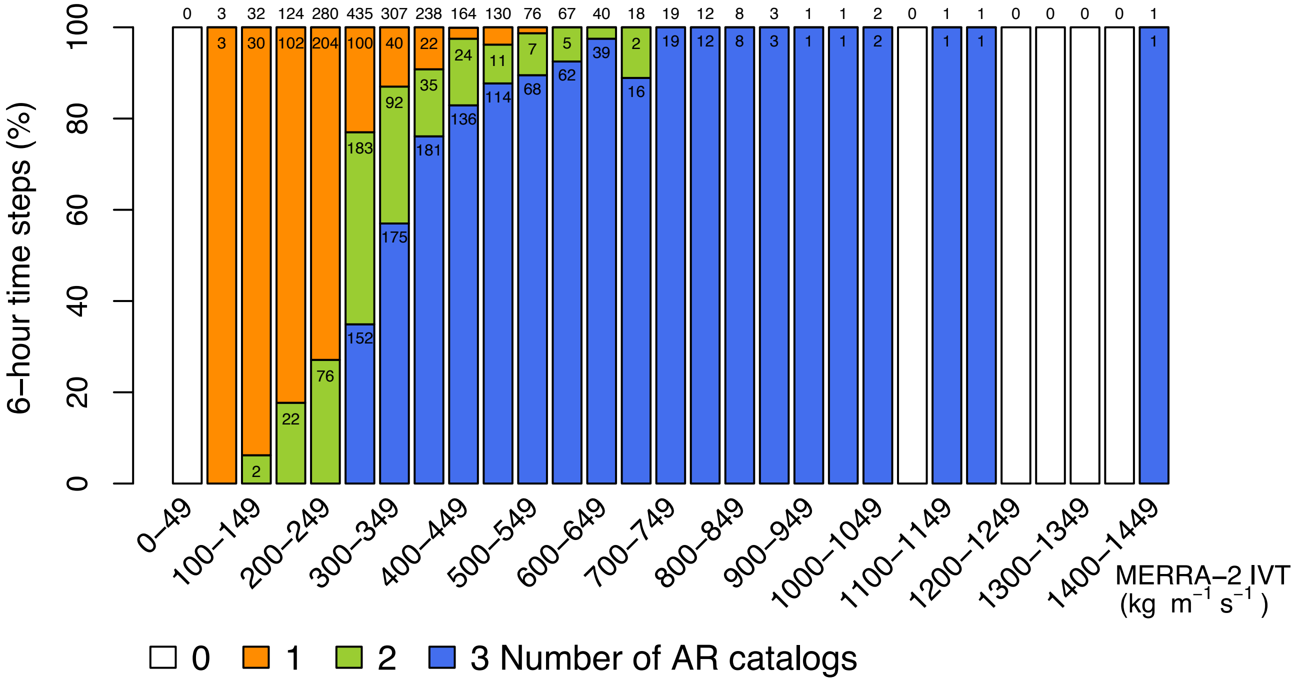


Figure 9



6–hour time steps (%)
0
20
40
60
80
100
0 3 32 124 280 435 307 238 164 130 76 67 40 18 19 12 8 3 1 1 2 0 1 1 0 0 0 0 1
0–49
100–149
200–249
300–349
400–449
500–549
600–649
700–749
800–849
900–949
1000–1049
1100–1149
1200–1249
1300–1349
1400–1449
MERRA–2 IVT
(kg m^{-1} s^{-1})
0 1 2 3 Number of AR catalogs

Figure 10

Click here to download Figure Fig10.jpg