When Will Humanity Notice Its Influence on Atmospheric Rivers?

Kai-Chih Tseng[1] [ID], **Nathaniel C. Johnson[2]** [ID], **Sarah B. Kapnick[3]** [ID], **William Cooke[2]** [ID], **Thomas L. Delworth[2]** [ID], **Liwei Jia[2,4]** [ID], **Feiyu Lu[1,2]** [ID], **Colleen McHugh[2,5]** [ID], **Hiroyuki Murakami[2,4]**, **Anthony J. Rosati[2,4]**, **Andrew T. Wittenberg[2]** [ID], **Xiaosong Yang[2]** [ID], **Fanrong Zeng[2]** [ID], and **Liping Zhang[2,4]** [ID]

[1]Program in Atmospheric and Oceanic Science, Princeton University, Princeton, NJ, USA, [2]Geophysical Fluid Dynamics Laboratory, National Oceanic and Atmospheric Administration, Washington, DC, USA, [3]J.P. Morgan, New York, NY, USA, [4]University Corporation for Atmospheric Research, Boulder, CO, USA, [5]Science Applications International Corporation (SAIC), Reston, VA, USA

Abstract Quantifying the response of atmospheric rivers (ARs) to radiative forcing is challenging due to uncertainties caused by internal climate variability, differences in shared socioeconomic pathways (SSPs), and methods used in AR detection algorithms. In addition, the requirement of medium-to-high model resolution and ensemble sizes to explicitly simulate ARs and their statistics can be computationally expensive. In this study, we leverage the unique 50-km large ensembles generated by a Geophysical Fluid Dynamics Laboratory next-generation global climate model, **S**eamless system for **P**rediction and **EA**rth system **R**esearch, to explore the warming response in ARs. Under both moderate and high emissions scenarios, increases in AR-day frequency emerge from the noise of internal variability by 2060. This signal is robust across different SSPs and time-independent detection criteria. We further examine an alternative approach proposed by Thompson et al. (2015), showing that unforced AR variability can be approximated by a first-order autoregressive process. The confidence intervals of the projected response can be analytically derived with a single ensemble member.

Plain Language Summary An "Atmospheric River" (AR) is a weather phenomenon characterized by strong, narrow moisture transport that brings heavy rainfall to land. They serve as a critical water resource but also can cause damaging flash floods and high winds. Thus, knowing how AR activity will change in the future climate can help us to mitigate potential AR-related disasters and promote effective water resource management. However, this task is challenging due to uncertainty in how fast the climate will warm and in how much the noise of natural climate variability can obscure the signal from global warming. In addition, several different definitions of AR exist, which raises questions about the sensitivity of AR changes to AR definition. In this study, we use a next-generation global climate model to evaluate the influence of these uncertainties and to determine the time when humanity will notice a discernible change in AR activity. We find that the response to global warming can be robustly identified by 2060 across all explored methods of computation. We further examine a less expensive approach, which enables us to quantify the uncertainty in warming signals and estimate the time of signal emergence in a single realization of nature.

1. Introduction

Atmospheric rivers (ARs), river-like plumes of intense atmospheric moisture transport (Newell et al., 1992) typically associated with extratropical cyclones (Zhang et al., 2019), are known for their great socioeconomic impacts worldwide. ARs occur less than 10% of the time but account for more than 30% of the annual precipitation over some populated mid-latitude regions, such as the West Coast of the United States and Western Europe, which indicates both their water resource value and the potential hazards of landfalling ARs (Dettinger, 2013; Slinskey et al., 2020). For example, in 2016–2017, a series of AR events terminated a multi-year drought in California (White et al., 2019). However, ARs also contribute to more than 84% of flash flood damage in the West Coast of the United States (Corringham et al., 2019), which is one of the most costly natural disasters. In addition to the local impacts, ARs play an important role in global climate. ARs determine about 90% of pole-ward moisture transport (Zhu & Newell, 1998), which is one potential driver of Arctic amplification through modulation of the energy and moisture budgets around polar regions (Nash et al., 2018). Due to the important role of ARs

in weather and climate extremes and the global hydrological cycle, it is imperative to understand how ARs may change under global warming and when such a signal may emerge from the noise of internal climate variability.

As the climate warms, it is anticipated that ARs will intensify, as moisture transport increases following the Clausius-Clapeyron equation (C-C equation), which dictates a water vapor increase of about 7% per degree Celsius of warming (Held & Soden, 2006)). However, previous studies have demonstrated that the change in AR intensity (i.e., integrated water vapor transport (IVT); see next section for definition) can exceed the scaling given by the C-C equation at a rate of about 7%–14% per degree Celsius (Gao et al., 2015) with location dependence. This difference can be attributed to the dynamical changes in ARs, which are more complicated than the thermodynamic changes since they are subject to a few different factors, such as the meridional shift of the eddy-driven jet (Shaw & Voigt, 2015; Shepherd, 2014), the increase of static stability (i.e., stronger warming aloft) and the change in stationary wave patterns (Wills et al., 2019). The change in integrated water vapor transport (IVT) is also reflected in the change in AR-day frequency (i.e., number of AR days in a given period at a fixed location). Previous research based on analyses of global climate model projections from Phase 5 of the Coupled Model Intercomparison Project (CMIP5) has shown that the AR-day frequency will increase by 30%–300% with a slight decline in the total number of global ARs in a warmer climate (Gao et al. (2015), Espinoza et al. (2018)). These studies also noted that the change mentioned above is sensitive to the definitions of AR-day in different detection algorithms.

Although previous studies indicate robust changes to ARs in response to increasing greenhouse gases, the forced signal is embedded within a climate system that undergoes substantial multidecadal internal variability. Therefore, we seek to understand when the signal of forced AR changes may emerge from the noise of internal climate variability. One way to address this question is to quantify the time of emergence (ToE), which is defined as the point in time when the signal of climate change emerges from the underlying noise of internal climate variability. In this study, we especially focus on when the forced response of AR-day frequency is significantly higher/lower than that of a chosen reference year (i.e., the first year of the record, see next section for definition). Informally, the ToE estimates the timing when societies can expect significant and detectable impacts of the climate change (Hawkins & Sutton, 2012; Thompson et al., 2015). An accurate estimation of AR ToE may benefit diverse sectors ranging from: water resource management, building design, policy decision making, and disaster preparedness. However, the uncertainty of estimating AR ToE may result from several sources in addition to internal climate variability, including model and scenario uncertainty and the choice of AR detection algorithm (O'Brien, Payne, et al., 2020; O'Brien, Risser, et al., 2020). It is therefore of great importance to examine the degree to which these uncertainties affect the ToE of AR changes. This study will especially focus on the uncertainty from internal climate variability, shared socioeconomic pathways (SSPs), and the AR detection threshold, which have been considered as the major uncertainty sources when studying AR climate change.

While there is no single metric or method for estimating ToE according to the IPCC Fifth Assessment Report (IPCC (2013)), all methods require an accurate estimate of internal climate variability, which typically requires large ensembles (≥30 ensemble members) with perturbed initial conditions from individual global climate models (Deser et al., 2020). In single-model large ensembles, the only process that can cause ensemble spread is the internal climate variability. Thus large ensembles provide a means of quantifying the uncertainty in forced responses caused by internal climate variability. Thompson et al. (2015) and Li et al. (2017) further proposed a computationally efficient way of estimating ToE by using a single ensemble member (or the observed realization of nature) with the assumption that the internal climate variability is roughly stationary and can be approximated by a Gaussian first-order autoregressive process (AR-1). By connecting ToE to t-statistics, their approach reconciles the definitions of ToE with a well-defined and widely used metric, which makes the inter-comparison of different studies possible. In this study, we leverage the large ensemble data generated by a next-generation climate model (see next section for details) developed at National Oceanic and Atmospheric Administration/ Geophysical Fluid Dynamics Laboratory (GFDL) and answer the following questions:

1. When can we expect that radiatively forced changes in AR statistics will emerge from the noise of internal climate variability?
2. Is AR ToE sensitive to the climate projection scenario or criteria used in AR detection algorithms?
3. Is the method proposed by Thompson et al. (2015) (i.e., using a single ensemble member for estimating ToE) applicable to ARs?

We first show a general picture of how the AR-day frequency responds as the climate warms in large-ensemble simulations by a GFDL next-generation climate model. We then define the AR ToE and test the fidelity of the approach proposed by Thompson et al. (2015) and Li et al. (2017). We demonstrate the ToE of AR changes is generally insensitive to SSPs or time-independent detection criteria and that increased frequencies of AR occurrence due to increasing greenhouse gases emerge from the noise of internal climate variability by the middle of the 21st century over most extratropical regions.

2. Data and Methods

2.1. Global Climate Model Large Ensemble

We analyze a 30-member initial-condition large ensemble from the **S**eamless system for **P**rediction and **EA**rth system **R**esearch (SPEAR-LE, Delworth et al. (2020)), which is the GFDL next-generation coupled GCM optimized for the study of seasonal to multidecadal variability, predictability, and projection. SPEAR incorporates GFDL's most recently developed atmospheric (AM4), oceanic (MOM6), sea-ice (SIS2) and land (LM4) component models. SPEAR shares the same component models as GFDL's CMIP6 model, Global Climate Model version 4 (CM4, Held et al. (2019)), but with configuration and physical parameterization choices optimized for climate prediction and projection from seasonal to multidecadal timescales. The version of SPEAR we analyze here has horizontal resolution of 0.5° in the atmosphere and land components and 1° in the ocean and sea ice components. The 0.5° atmosphere resolution is unprecedented for available coupled GCM large ensembles and provides a unique opportunity for investigating the role of internal variability in the projected changes of atmospheric extreme events, including ARs. Previous research has examined ARs in SPEAR and AM4 showing its capability of reproducing the observed climatology, including frequency, numbers and geometry (Zhao, 2020). In addition, SPEAR also shows the capability of forecasting multiseasonal AR activity and capturing interannual variability in the present climate (Tseng et al., 2021), which gives us additional confidence for projecting future AR activity.

In this study, we examine simulations with historical forcing, natural forcing, and projected forcing following two SSPs: SSP2-4.5 and SSP5-8.5. The historical simulation follows the observed radiative forcing from 1921 to 2014, while the all-natural forcing simulation, which can be considered as the control simulation, contains the same radiative forcing but with anthropogenic greenhouse gases excluded. The all-natural forcing simulation is only used to provide a reference state and to confirm that all significant trends are not caused by climatological drift in SPEAR; the natural forcing results are shown in Supporting Information S1. The SSP2-4.5 and SSP5-8.5 pathways represent "middle of the road" and "upper end" of anthropogenic greenhouse gas emissions scenarios, respectively, and cover the period of 2014–2100. For the 30 ensemble members, the initial conditions were chosen from a long control simulation with 20-year spacing to sample different phases of internal climate variability. SPEAR large ensemble data is described further and available at: https://www.gfdl.noaa.gov/spear_large_ensembles/. The pre-processed data used in this study is stored at: https://doi.org/10.5281/zenodo.6366550.

2.2. Observational Data

The European Centre for Medium-Range Weather Forecasts reanalysis (ERA5, Hersbach et al. (2020)) from 1995 to 2018 is used as the observational reference. The selection of the period is due to the availability of satellite observation to constrain the moisture field. The ERA5 data, including wind and specific humidity, are interpolated from their 30-km resolution native grid to the same common grid as SPEAR (0.5° and daily temporal resolution) before applying the AR detection algorithm to the data (see next section for details). ERA5 is published within 3 months of real time and is available at: https://cds.climate.copernicus.eu/#!/search?text=ERA5&type=dataset.

2.3. AR Detection

ARs are identified with a detection algorithm (Mundhenk et al., 2016) that tests if the gridded IVT meets a set of criteria regarding intensity (85% or 94% of the global domain in historical simulation) and geometry (>1,400 km in length and with an aspect ratio ≥1.4). The IVT is defined as:

$$\text{IVT} = \frac{1}{g}\sqrt{\left(\int_{1000}^{250} q\mathbf{v}\,dp\right)^2} \tag{1}$$

where g, q, v and p are gravitational acceleration, specific humidity, horizontal wind and pressure coordinate respectively. We remove the first three harmonics of the IVT calendar day means (over the historical period) for each grid point to compute the anomaly field. To examine the sensitivity of ToE to the AR intensity criteria, we apply the detection algorithm to IVT anomalies with two different thresholds: 85% and 94%. The thresholds are defined by aggregating the global IVT over the entire historical period and ensemble members and calculating the 85% and 94% percentiles. These two criteria represent a reasonable range that spans most of the threshold criteria range used in previous research (Zhao, 2020). Each IVT map is scrutinized independently and a daily AR mask is generated (0 for non-AR and 1 for AR days for a given grid cell). The seasonal AR-day frequency for a grid cell is then defined as the number of AR days that are labeled in a given season. The detection algorithm is available at: https://mountainscholar.org/handle/10217/170619.

In addition to the Mundhenk et al. (2016) method, we also test two other detection algorithms, Toolkit for Extreme Climate Analysis, Bayesian AR Detector (TECA-BARD, O'Brien, Payne, et al. (2020), O'Brien, Risser, et al. (2020)) and Tempest (Ullrich and Zarzycki (2017); O'Brien et al. (2021)). Different from the fixed IVT thresholds used in Mundhenk et al. (2016), TECA-BARD and Tempest have a relative/time-dependent threshold. TECA BARD has a percentile threshold P that constrains AR area coverage to be no larger than approximately $1 - P$, and since most of the ARDTs run within TECA-BARD have $P = 0.95$ (posterior probability based on subjective expert knowledge), the maximum area covered is typically close to 5%, in historical simulations and SSPs simulations, as described by O'Brien et al. (2021). On the other hand, Tempest implements a Laplacian operator to identify the local maximum of IVT. Here we test two thresholds in Tempest, $-10,000\ \mathrm{kg\ m^{-1}s^{-1}\ rad^{-2}}$ and $-3,000\ \mathrm{kg\ m^{-1}s^{-1}\ rad^{-2}}$, which are roughly equivalent to 300–100 $\mathrm{kg\ m^{-1}s^{-1}}$ difference in IVT from the core of an AR to its environment, respectively. One should note that the use of time-dependent thresholds can automatically remove the effect of increased background moisture caused by global warming, because the IVT threshold is defined by percentile calculated in space rather than in time. O'Brien et al. (2021) found that the ARs detected by these two algorithms behave differently from other algorithms in a warmer climate (see discussion in Section 3.4). While the main focus is the fixed IVT threshold used in Mundhenk et al. (2016), we also discuss ToE sensitivity to these different algorithms and the impact of time-dependent thresholds in Section 3.4.

2.4. Estimating the Time of Signal Emergence (ToE)

While there is no single metric for estimating ToE, the most widely-used definition is the time when the ratio of the signal from forced response to the noise from the internal climate variability exceeds a predefined threshold. In this study, we especially focus on the timing when the anthropogenically forced response is significantly greater or smaller than the forced response from a reference year. For example, considering the case where the deviation of AR seasonal frequency from a reference year (i.e., x_{n_t}, where n_t is the time step) is determined by two components:

$$x_{n_t} = f(n_t) + \xi_{n_t} \tag{2}$$

where $f(n_t)$ is the forced response of AR-day frequency and ξ_{n_t} is the contribution from internal climate variability (i.e., unforced component). Since we are interested in when the trend in AR-day frequency is significantly different from 0, it is equivalent to the timing when the null hypothesis H_0: $f(n_t) = 0$ is rejected at a certain confidence level. The corresponding confidence interval (CI) is:

$$CI = f(n_t) \pm e_{n_t} \tag{3}$$

where e_{n_t} is the spread of the forced response. In an ideal case, e_{n_t} will be 0 if ξ_{n_t} is decorrelated with $f(n_t)$. However, in a short time period, the internal climate variability can project onto the trend, which can obscure the forced response. One way to estimate $f(n_t)$ is with the ensemble mean from a large ensemble of radiatively forced global climate simulations, with the assumption that the ensemble size is large enough to sample the different realizations of internal climate variability. Then the CI can be determined by regressing detrended x_{n_t} from each ensemble member onto $f(n_t)$ (see Supporting Information S1 for details). Note that considering the forced response in the context of specific values of the natural variability (i.e., $|\frac{f(n_t)}{\mathrm{std}(\xi_{n_t})}| \geq n$, n is a positive integer and $\mathrm{std}(\xi_{n_t})$ is the standard deviation of ensemble spread) is not the same criteria as considering when $f(n_t)$ is

significantly different from 0. The former tends to overestimate the ToE because it is a more difficult criterion to achieve (see discussion in Li et al. (2017)).

Because generating the large-ensemble simulations is computationally expensive, Thompson et al. (2015) and Li et al. (2017) further proposed an alternative approach that can be applied to a single climate realization. They model the total change of x_{n_t} as an AR-1 process superimposed on an external forcing:

$$x_{n_t} = f(n_t) + \alpha x_{n_t - 1} + \epsilon_{n_t} \tag{4}$$

where x_{n_t} and $f(n_t)$ are identical to the definitions given in Equation 2, α is the lag-1 autocorrelation of the detrended x and ϵ_{n_t} is random white noise at n_t. In this formula, $f(n_t)$ is predetermined by a parametric fitting (e.g., linear trend or 2nd order polynomial) or the ensemble mean from anthropogenically forced climate simulations.

Following Thompson et al. (2015), if the detrended data is well modeled by an AR-1 process, the e_{n_t} in Equation 3 can be estimated based on t-statistics as follows,

$$e_{n_t} = t_c n_t s \gamma(n_t, r_1) g(n_t) \tag{5}$$

where

$$\gamma(n_t, r_1) \equiv \left\{ (n_t - 2) \middle/ \left[n_t \left(\frac{1 - \alpha}{1 + \alpha} \right) - 2 \right] \right\}^{1/2} \tag{6}$$

and

$$g(n_t) \equiv \sqrt{\frac{12}{n_t^3 - n_t}} \tag{7}$$

In Equation 5, t_c is the t value for the desired CI and s is the standard deviation of detrended values of x_{n_t}. $\gamma(n_t, r_1)$ is the function which accounts for the effect of memory in detrended data, where α is the lag-1 autoregression of detrended x and $g(n_t)$ accounts for the degrees of freedom. For example, if the detrended x_{n_t} has internal climate variability with long memory (large r_1), $\gamma(n_t, r_1)$ tends to be large, suggesting the influence of noise can last longer and modulate x_{n_t} in the following time steps. On the other hand, like the reduced spread of t-distribution with the increase of sample size, $g(n_t)$ and e_{n_t} are reduced when we use more data (i.e., bigger n_t) to model the underlying distribution of detrended x_{n_t}. The derivations of Equations 3–7 can be found in the appendix of Thompson et al. (2015). In this study, we focus on December–February (DJF) averaged AR statistics and each time step represents one season with 1-year spacing to the next time step.

3. Results

3.1. Changes in AR-Day Frequency and Numbers

We first illustrate the *epoch difference*–that is, the difference between the average values of a time series from two selected windows (e.g., 2015–2100 minus 1921–2014), and the time evolution of DJF AR-day frequency in the two warming and the historical simulations. The left column of Figure 1 shows the AR-day frequency as a function of time and location (diamond marks in the middle column) based on the 94% IVT threshold and the right column shows the AR numbers in given regions (red boxes in the middle column). The selection of the four locations (diamond marks) is based on the large amplitude of AR-day frequency change in response to global warming, while the regional average shows similar features but with a smaller amplitude (not shown). The shading is the range of two standard deviations of the ensemble spread and the color scatters are the ensemble means from the SPEAR historical, SSP2-4.5, and SSP5-8.5 simulations. We also include ERA5 data (black scatters) as an observational reference. The AR-day frequency from ERA5 is generally bounded within the range of 2 standard deviations of the ensemble spread, showing the capability of SPEAR in reproducing the observed climatology. The middle column shows the epoch difference between the historical and two warming simulations. Most regions show an increase in AR-day frequency under both SSPs, especially in the vicinity of the extratropical storm tracks.
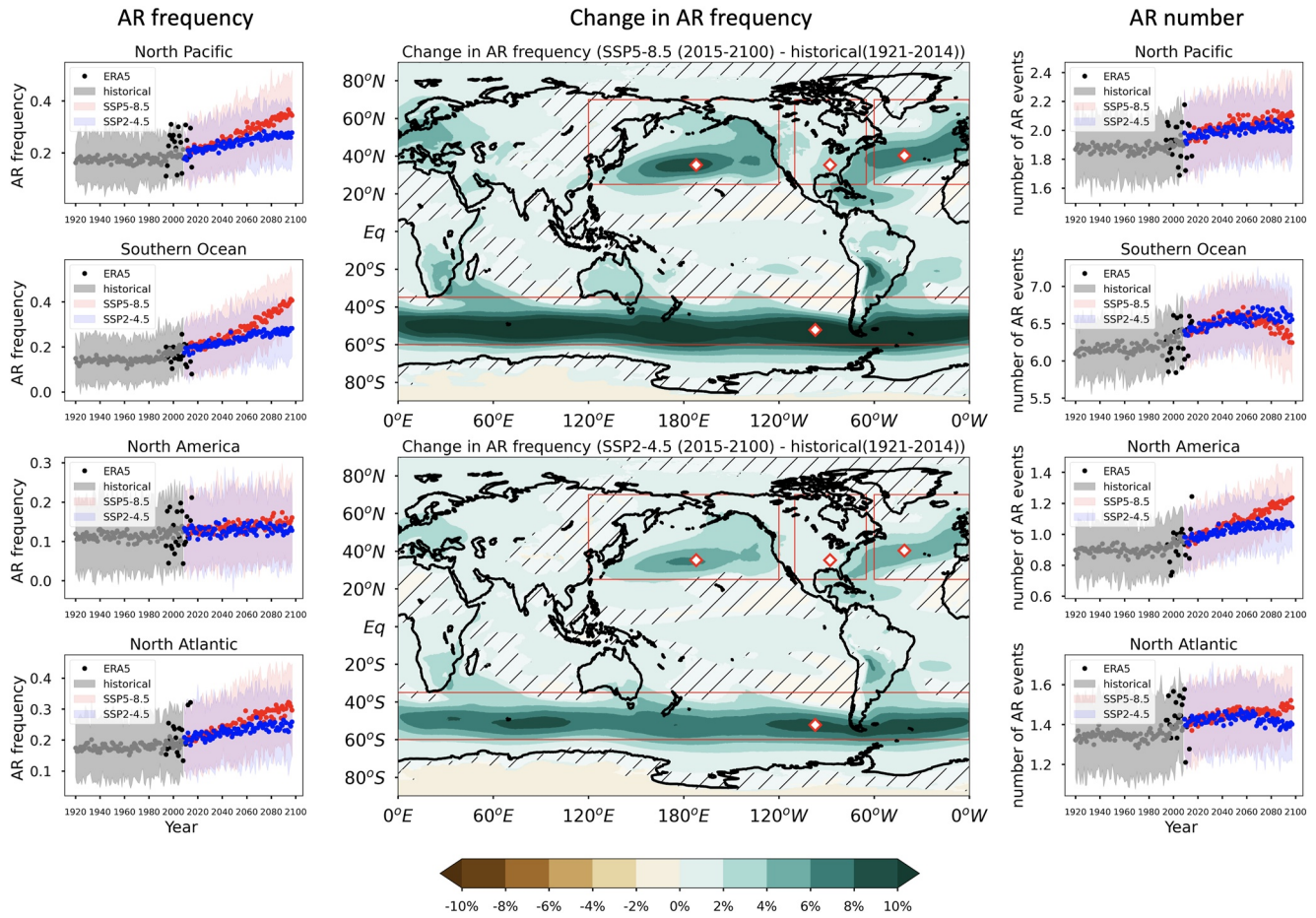
Figure 1. Left Column: Time series of the December–February (DJF) atmospheric river (AR)-day frequency (AR days/day) in the three **S**eamless system for **P**rediction and **EA**rth system **R**esearch simulations and observations from four chosen locations (diamond marks in the middle column). The color scatters are the ensemble mean from the historical (gray), SSP5-8.5 (red), and SSP2-4.5 (blue) simulations and from ERA5 (black). The shading shows the range of 2 standard deviations of the ensemble spread from historical (gray), SSP5-8.5 (red) and SSP2-4.5 (blue). **Right Column:** The same as left column except for the DJF AR numbers in the four specified domains (red boxes in the middle column). **Middle Column:** The epoch difference of AR-day frequency between the warming (top: SSP5-8.5, bottom: SSP2-4.5) and the historical simulations. The warming simulation is averaged over the periods of 2015–2100 and the historical simulations is averaged over the periods of 1921–2014. AR definition uses the the 94% integrated water vapor transport threshold. Regions without hatching indicate the change in AR-day frequency is statistically significant at 5% level based on a *t*-test.

According to the left column of Figure 1, the increase in AR-day frequency ranges from near 0 (e.g., North America) to ~300% (AR-day frequency rises from around 12% in 1921 to 36% in 2100 over the Southern Ocean) for SSP5-8.5. This increase is similar in spatial pattern but about half the amplitude for SSP2-4.5. For global-mean values, the frequency increases by ~80% for SSP5-8.5 and ~40% for SSP2-4.5. This result is qualitatively consistent with previous studies (Espinoza et al., 2018), which showed the increase in AR-day frequency due to the increase of atmospheric moisture approximately following C-C scaling. However, the change in amplitude is slightly different from the values suggested by Espinoza et al. (2018) (~50%). This difference may result from the different periods for the analysis as well as the threshold in detection algorithms. For example, their historical period is defined from 1979 to 2002 and the warming period is defined from 2079 to 2096. If we use the same period for analysis, the global mean AR-day frequency increases by ~73% for SSP5-8.5 and ~39% for SSP2-4.5, which is closer to the value given by Espinoza et al. (2018). We also show the result based on 85% IVT threshold in the appendix (Figure S2 in Supporting Information S1). In general, the change in AR-day frequency is similar to but smaller than that based on the 94% IVT threshold. This also suggests some sensitivity of change in AR-day frequency due to the AR detection threshold.

Regarding the total number of ARs, most regions show an increasing trend (right column of Figure 1) while the change in the Southern Ocean is not monotonic, reaching its maximum around 2070 for both SSP2-4.5 and

SSP5-8.5. This result contrasts most previous studies that suggest a slight decline in projected global AR numbers (Zhao, 2020). There are two possible explanations. First, Shields and Kiehl (2016) found the number of strong and long-lived ARs tends to increase in a warmer climate, while the short-lived ARs show an opposite trend. The 94% IVT intensity criterion is around the higher end among all detection algorithms, suggesting the detected ARs in our study are more likely to be stronger ones. Second, Shields and Kiehl (2016) and Zhao (2020) also note that the ARs become larger as the specific humidity increases in the atmosphere due to global warming, given that the IVT criterion is sensitive to changes in background specific humidity. This phenomenon further leads to the decline in AR number because bigger ARs tend to aggregate into a non-filament structure. The non-filament structure of AR is less likely to pass the geometry test in detection algorithms, which also leads to a decline of AR numbers (see Figure 10 in Zhao (2020)). These explanations are supported by Figure S2 in Supporting Information S1 where we use a lower IVT threshold (85%) in the detection algorithm. In the right column of Figure S2 in Supporting Information S1, except for North America where historical AR occurrence is lower, other regions all show a decline in AR numbers as the climate warms, indicating AR numbers have reached the maximum in these regions. We also find similar non-filament AR structures as identified by Zhao (2020) (figure not shown) suggesting that the AR numbers over different climate states are sensitive to the IVT threshold in AR detection algorithms and might not be a well-defined variable when studying climate change. Thus, we will focus on AR-day frequency in the following analysis.

Figure 1 provides a general picture of how the ensemble mean AR-day frequency changes in a warmer climate– that is the "signal" part of warming response. It is also intriguing to understand how the "noise" (i.e., internal climate variability) changes with time, which is relevant both for likelihood of more extreme changes and for determining the AR ToE. Figure 2 shows the ensemble spread of AR-day frequency as a function of time (left column) and the histograms of ensemble spread shown on the left column (right column). The maps in the middle column show the difference in median values of the histograms between warming and historical simulations. In Figure 2, a few interesting features can be observed. First, the ensemble spread increases as the climate warms and the increasing amplitudes are comparable between SSP5-8.5 and SSP2-4.5 (left column). To support this, we evaluate if the model spread significantly increases from the historical simulation to the two warming simulations and if the difference between two warming simulations is significant. Here we use a bootstrapping analysis to determine the distribution that could arise from sampling variability. Specifically, we randomly draw two values, either from the historical simulation or from the SSP5-8.5 simulation (50% chance), and calculate the difference. By repeating this process 1,000 times, we can approximate the distribution of model spread difference based on random sampling and determine if the median values of model spread significantly increase from the historical simulation (gray histograms in Figure 2) to the SSP5-8.5 simulation (red histograms in Figure 2). The result in Figure 2 is based on a two-tailed test at the 5% significance level. We also apply the same analysis to the two warming simulations and test if the ensemble spread from the two warming simulations is significantly different from each other.

In the middle column of Figure 2, the color shading indicates the following two criteria are both satisfied: (a) the median values of ensemble spread from SSP5-8.5/SSP2-4.5 are significantly higher than that of the historical simulation and (b) the median values of ensemble spread from the two warming simulations are *not* significantly different from each other. Over the four targeted regions, the increase in ensemble spread in the warming simulations is indeed significant, but the differences in spread between the SSP5-8.5 and SSP2-4.5 simulations are not. These findings support the visual impression of the middle column of Figure 2, which shows widespread regions around the globe with increased internal AR climate variability, especially the mid-latitudes, implying the increased chance of experiencing AR extremes. Similar features are observed when the 85% IVT threshold is used (Figure S3 in Supporting Information S1) except that the AR spread increase is lower in amplitude and the difference is only significant in a more confined area over the globe.

To briefly summarize the findings illustrated in Figures 1 and 2, both the ensemble mean and ensemble spread of AR-day frequency in SPEAR increase as the climate warms. These features are qualitatively insensitive to SSPs or AR detection criteria, but the increases in the modest SSPs (SSP2-4.5) and for the weaker IVT threshold for AR detection (85% IVT threshold) are lower in magnitude. In the next section, we will apply the approach proposed by Thompson et al. (2015) to AR-day frequency trends and explore how the competing factors of stronger signal and noise with time influence the estimation of AR ToE.
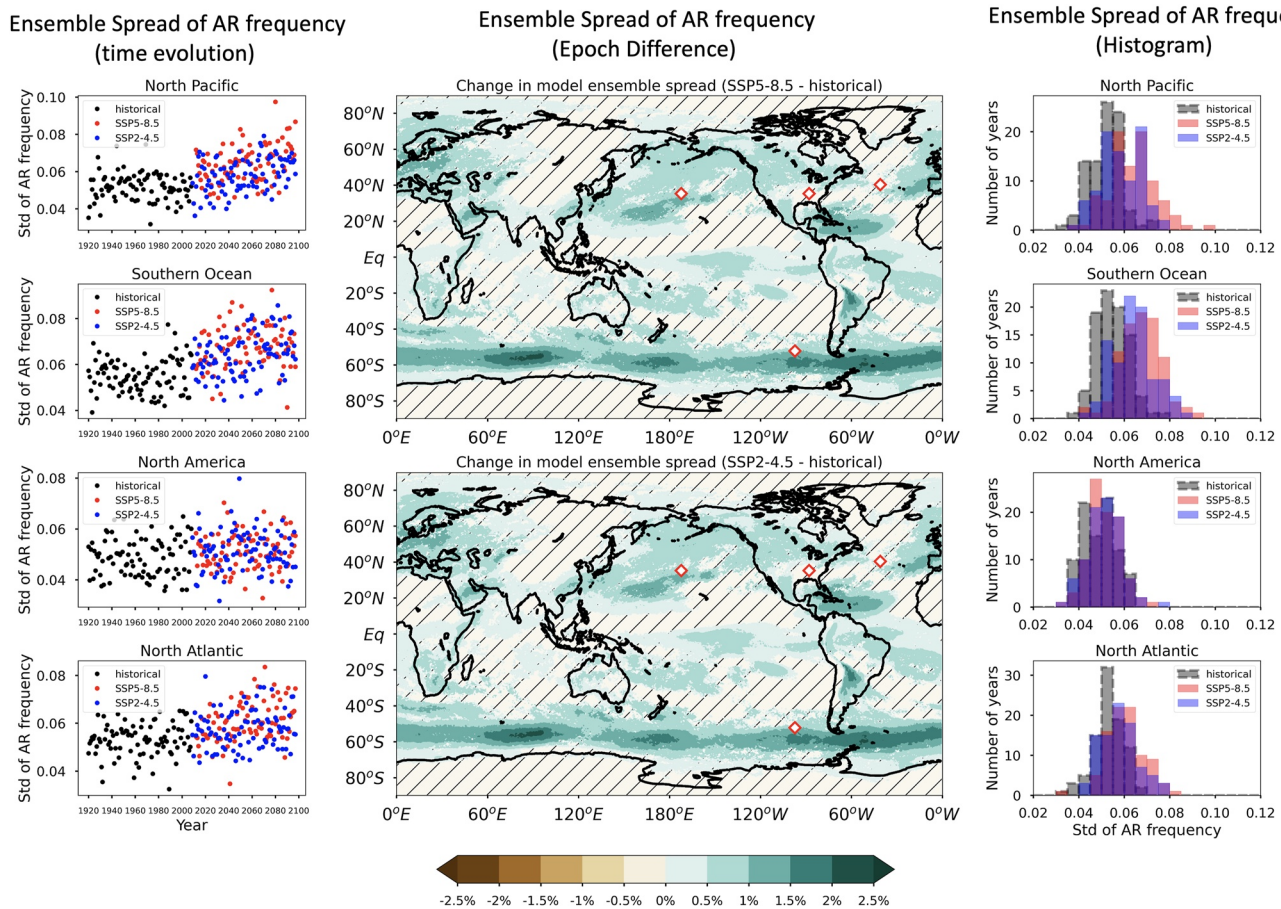
Figure 2. Left Column: The time series of December–February (DJF) ensemble standard deviations of atmospheric river (AR)-day frequency for the 30 **S**eamless system for **P**rediction and **EA**rth system **R**esearch ensemble members from the historical simulation (black), SSP5-8.5 (red) and SSP2-4.5 (blue). The four chosen locations are shown as diamond marks in the middle column. **Right Column:** The histograms of the DJF AR ensemble standard deviations in the left column from the historical simulation (black), SSP5-8.5 (red) and SSP2-4.5 (blue). **Middle Column:** Difference in median values of histograms in the right column between the warming and the historical simulations. Regions without hatching indicate the following two criteria are both satisfied: (a) the median values of ensemble spread from SSP5-8.5/SSP2-4.5 are significantly higher than the historical simulations and at the 5% level. (b) the median values of ensemble spread from two warming simulations are **not** significantly different from each other.

3.2. Estimating ToE in the SPEAR Large Ensemble

In this section, we estimate the AR ToE in the SPEAR large ensemble and examine the approach proposed by Thompson et al. (2015). According to Section 2.4, the CI in Equation 3 is determined by projecting x_{n_t} from each ensemble member onto the forced response, $f(n_t)$, which is the ensemble mean (see Supporting Information S1). The assumption here is that the ensemble size is big enough to sample different phases of internal climate variability and to remove ξ_{n_t} in Equation 3. In addition, we also examine the approach proposed by Thompson et al. (2015) and analytically derive the CIs based on Equations 4–7. To compute the ToE based on the method of Thompson et al. (2015), we need to decide the functional form of $f(n_t)$ first. According to Li et al. (2017), there are numerous ways to determine $f(n_t)$ in Equation 3, such as using linear regression or 2nd order polynomial fitting. A proper selected function will closely follow the ensemble mean. In addition, to connect the CIs of the forced response with t-statistics, the residual should be well modeled by an AR-1 process (i.e., Gaussian process) and the $f(n_t)$ from each ensemble member is bounded within the desired CIs based on predetermined t-statistics according to Thompson et al. (2015).

Figure 3 illustrates the CIs based on large ensemble simulations and the method of Thompson et al. (2015) with the use of a 2nd order polynomial fitting for $f(n_t)$ at the 4 chosen locations specified in Figure 1. The left column corresponds with SSP5-8.5 and the right column with SSP2-4.5. The dashed lines are the ensemble average of $f(n_t)$ from each member and the shading shows the analytically derived CIs (i.e., e_{n_t}). One should notice that we
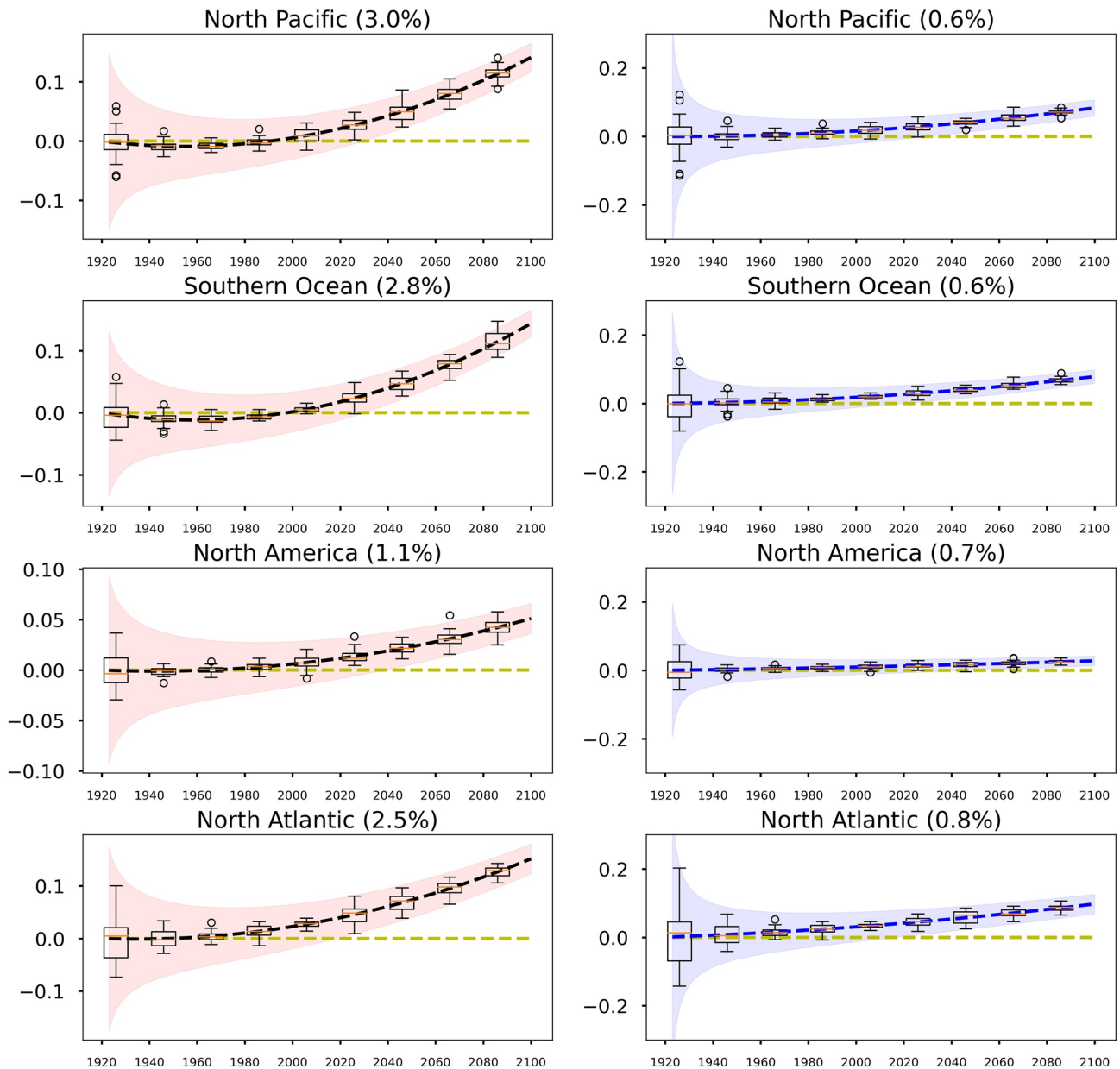
Figure 3. The forced trend of December–February atmospheric river-day frequency (i.e., $f(n_t)$) from the four chosen locations indicated in the middle column of Figure 1. The dashed curves are ensemble mean and the shading is the analytically derived confidence intervals (CIs) (Equation 3). Here we choose two-tailed 95% CIs ($t_s \sim 2.26$ for 30 ensemble members). The box plots are the forced response derived from each ensemble member (see Supporting Information S1) at the given year. The number in the title shows the percentage of ensemble member that fall outside the analytically derived CIs. In this figure, we use 2nd order polynomial fitting for $f(n_t)$. The left column is from SSP5-8.5 and the right column is from SSP2-4.5.

assume the ensemble average of $f(n_t)$ is the best approximation of the forced response while it can be replaced with $f(n_t)$ from any single member with a small change in the derived CIs (results not shown). The box plots show the ensemble distribution of $f(n_t)$ at each 10-year interval calculated directly from the large ensemble data (see Supporting Information S1 for details). The number in the title of each subplot shows the percentage of ensemble members falling outside the analytically derived CI. These values are generally around and less than the expected 5%, indicating that the uncertainty caused by internal climate variability is well modeled by an AR-1 process. The other evidence is that the range of $f(n_t)$ derived from each ensemble member (i.e., the error bars for the box plots) ties closely to the analytically derived CI and tapers off from 1921 to 2100. The conclusion holds for the SSP2-4.5 simulation as well. In Figure S4 in Supporting Information S1, we also include the results based on

a linear trend. With a linear fit, more members fall outside the CIs than in Figure 3, indicating that a 2nd order polynomial is a more appropriate model. The result suggests that as long as the parametric fitting for $f(n_t)$ is properly selected, we can model internal AR climate variability as an AR-1 process and estimate the AR ToE with a single ensemble member.

While the above analysis is based on the assumption that s in Equation 5 is roughly stationary in time, one might wonder if the conclusion is still valid when we consider s as a function of time. In Figure S5 in Supporting Information S1, we illustrate the result when we replace a constant s with 10-year forward running averaged $s(n_t)$. For example, to calculate the e_{n_t} in the year of 1931, we only use the detrended x_{n_t} from 1931 to 1940 to calculate s. For the year later than 2091, we use the rest of the data. In Figure S5 in Supporting Information S1, it is evident that the conclusion generally holds except with fewer ensemble members falling outside the CI. The reason that Figures 3 and S5 in Supporting Information S1 exhibit only small differences is because the increase of ensemble spread (left column of Figure 2) is much smaller than the increase of the trend (left column of Figure 1). Thus, even though the AR internal climate variability increases with time, its contribution to uncertainty of the forced response (i.e., e_{n_t}) is still negligible.

3.3. AR Time of Emergence

The results in the previous section suggest that for the purpose of calculating the uncertainty of AR trends and ToE, the role of internal variability in a large ensemble of warming simulations can be quantified to a high degree of accuracy from the statistics of unforced variability. In this section, we follow the calculation in Figure 3 by using the 2nd order polynomial for $f(n_t)$ (average over all members) and CIs based on Equation 3. We further explore the robustness of AR ToE estimates by evaluating the sensitivity of AR ToE to SSPs and AR detection criteria. Figures 4a–4d show the maps of AR ToE from the two different warming simulations and two IVT criteria used in the AR detection algorithms (left column), where the ToE is defined as the n_t when the null hypothesis H_o: $f(n_t) = 0$ is rejected at the 5% level. We also demonstrate the difference between these maps in the right column of Figure 4. Here we choose 2nd-order polynomial fitting for $f(n_t)$ while the result based on ensemble mean is nearly identical, as suggested in the previous section (figure not shown). Figure 4 reveals a few interesting features. First, in most extratropical oceans, the AR-day frequency changes emerge from internal variability before 2060 regardless of the SSPs or the IVT criteria, while some regions, such as the Southern Ocean, show signal emergence as early as 2020 (dark red shading). This indicates that if we were to have reliable AR observations extending back to 1921, we would expect that the AR-day frequency changes over much of the Southern Ocean would stand out clearly above the noise of internal climate variability. The second feature is that the AR ToE sensitivity to SSPs and IVT threshold is relatively small (generally ≤20 years), which indicates the estimation of AR ToE is quite stable, at least in this model. Only a few regions around the subtropical oceans have differences greater than 20 years (black contours in Figures 4e–4g).

While the above analyses focus on DJF, we also investigate all other seasons and show the result in Figure 5. In general, the conclusion that the estimated ToE is not sensitive to SSP scenarios or the chosen IVT threshold is still valid. Another interesting feature is that the AR-day frequency has a more significant increase and earlier ToE in the summer than in the winter. This feature is especially clear on the poleward flank of the storm track. One possible reason is that the poleward shift of AR geometry center is more evident during the summer (figure not shown). While this might be the case, it does not necessarily indicate that the center of storm track (i.e., regions with maximum eddy kinetic energy) significantly shifts poleward. Instead, it might reflect the fact that a more significant increase in summer IVT over higher latitudes. This statement can be justified in the later part of this section where we repeat the same analysis for IVT without applying any AR detection algorithm. The reason for more robust increase in summer AR is not clear at this moment and the mechanisms responsible for this feature will be explored in the future study.

The reasons for a stable estimation of AR ToE between two warming simulations and between two IVT thresholds are quite different. The low sensitivity to SSPs results from much of the AR signal emergence occurring within several decades of the historical period (generally before 2060), which provides a limited time for the different SSPs to diverge prior to ToE. Consistently, the areas with the largest ToE sensitivity to SSPs, such as the subtropical regions (black contours in Figure 4e), have the latest ToE. The limited sensitivity to IVT threshold for
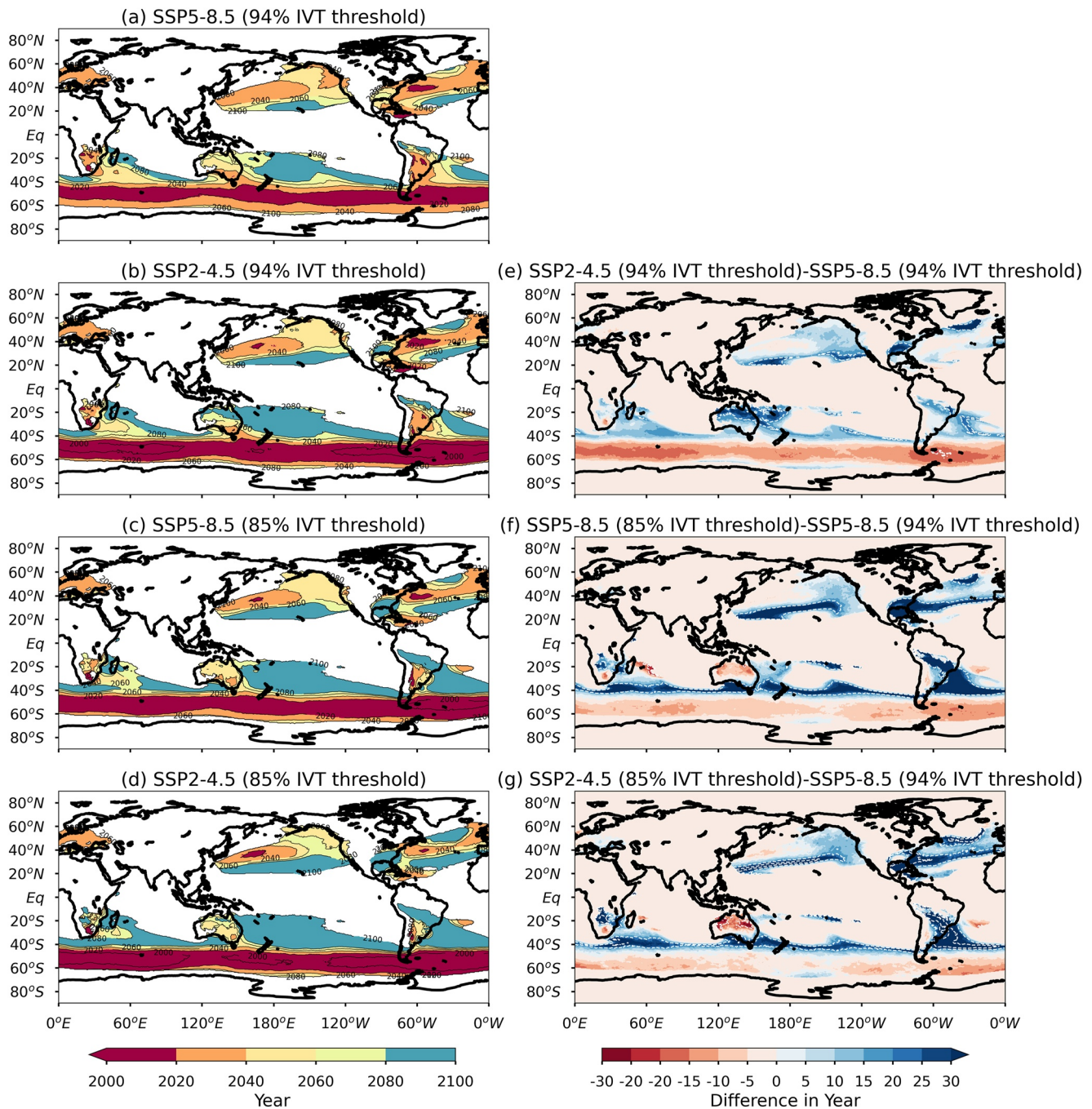
Figure 4. **Left Column:** the time of emergence (ToE) of atmospheric river (AR)-day frequency changes from (a) SSP5-8.5/94% integrated water vapor transport (IVT) threshold (b) SSP2-4.5/94% IVT threshold (c) SSP5-8.5/85% IVT threshold and (d) SSP2-4.5/85% IVT threshold. The ToE is defined as the n_t when $H_o: f(n_t) = 0$ is rejected at the 5% significance level. We only show the regions with December–February AR-day frequency $\geq 2\%$ (AR days/days) in the last 10 years of simulations (2090–2100). **Right Column:** The difference corresponding to each panel on the left side: (e) SSP2-4.5/94% IVT threshold minus SSP5-8.5/94% IVT threshold (f) SSP5-8.5/85% IVT threshold minus SSP5-8.5/94% IVT threshold (g) SSP5-4.5/85% IVT threshold minus SSP5-8.5/94% IVT threshold. White-dashed contours are regions with difference greater than 20 years.

the AR definition results from similar behavior in the AR-day frequency changes under increasing greenhouse gases relative to the baseline climatology, despite large sensitivity of the AR climatology to IVT threshold.

Given the robust emergence of forced response in ARs by the late 21st century over most regions, one might wonder if similar ToE features can be found in IVT since the most extreme IVT over the extratropics is usually
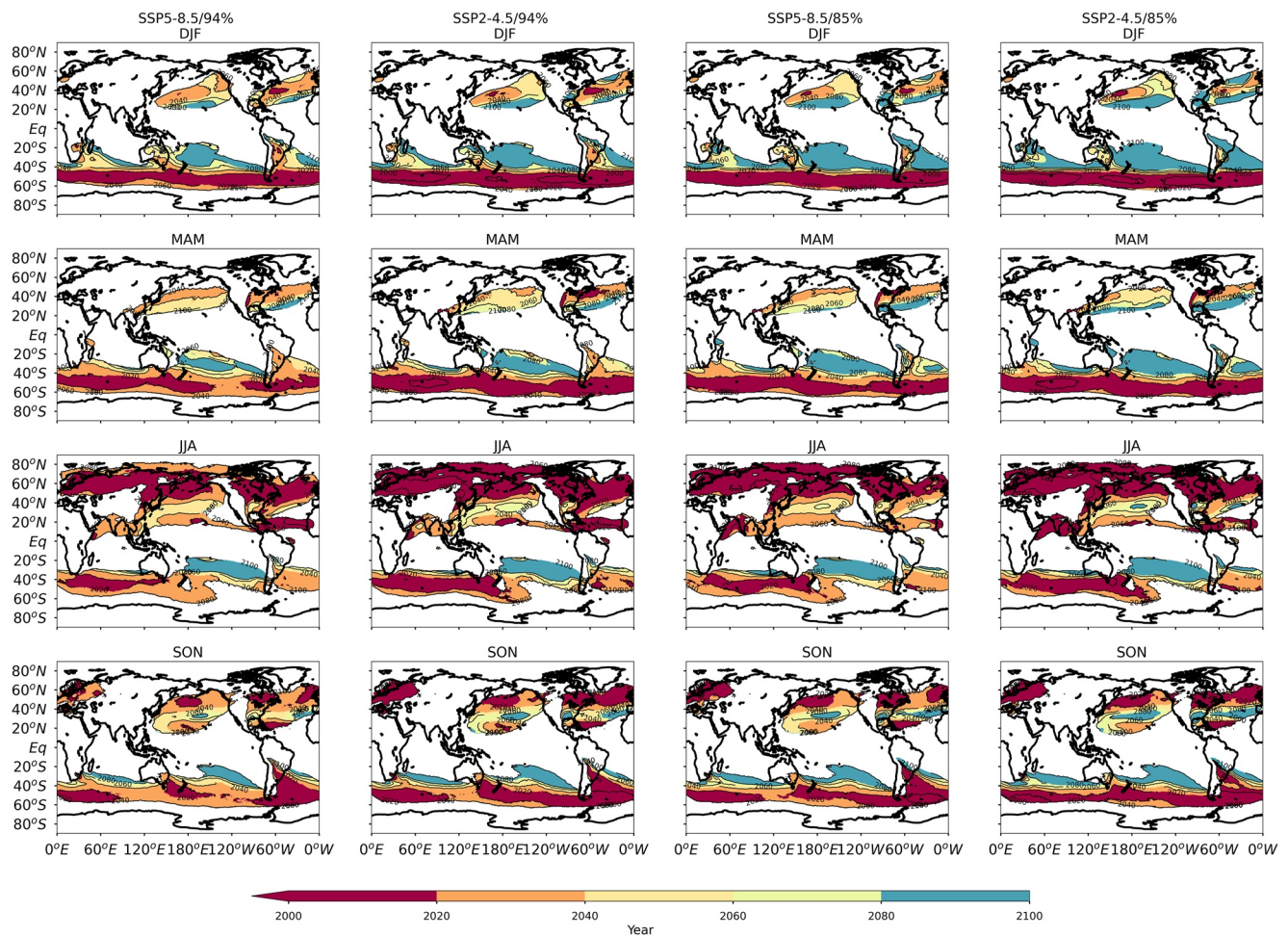
Figure 5. The atmospheric river-day frequency time of emergence for four different seasons, two warming scenario simulations and two detection thresholds. From top to bottom are December–February (DJF), March-May (MAM), June-August (JJA) and September-November (SON) respectively. From left to right are SSP5-8.5/94%, SSP2-4.5/94%, SSP5-8.5/85% and SSP2-4.5/85%. The first row is identical to the left column of Figure 4.

associated with ARs. Here we repeat the same analysis but for extreme IVT, defined as the 95th percentile of seasonal IVT. Specifically, we aggregate the daily IVT from all ensemble members over a given season and find the 95th percentile. Each grid point is scrutinized independently, and the time series of extreme IVT is analyzed in the same way that we analyze the AR-day frequency. Since no AR detection algorithm is applied, the spatial coherence (i.e., filament structure) of high IVT regions is not required. This analysis enables us to explore the intensity changes of the most extreme IVT in a warmer climate without the geometry constraints that characterize ARs. Figure 6 illustrates the ToE of extreme daily IVT. In general, the top 5% IVT shows similar ToE features as those shown in Figures 4 and 5 where the storm track regions have the earliest ToE and the summer hemisphere emerges earlier than the winter hemisphere. This result supports that the change in AR-day frequency reflects the change in the amplitude of the most extreme IVT.

The above analysis suggests that some populated mid-latitude regions, such as the West Coast of North America and Western Europe, might detect significant changes in AR statistics within the next few decades. These regions are located near or within the climatological storm tracks, and so the robust AR response likely reflects the thermodynamic effect of increased moisture within the saturated moisture plumes that characterize ARs and that preferentially follow the storm tracks (O'Gorman and Schneider (2009)). This result might especially be of great interest to regions like California, where most of the annual precipitation is determined by AR occurrences in a narrow window of the year, and the difference between wet and dry years is typically due to the occurrence or absence of a few big winter storms (Dettinger, 2013). A previous study (Gershunov et al., 2019) has documented a
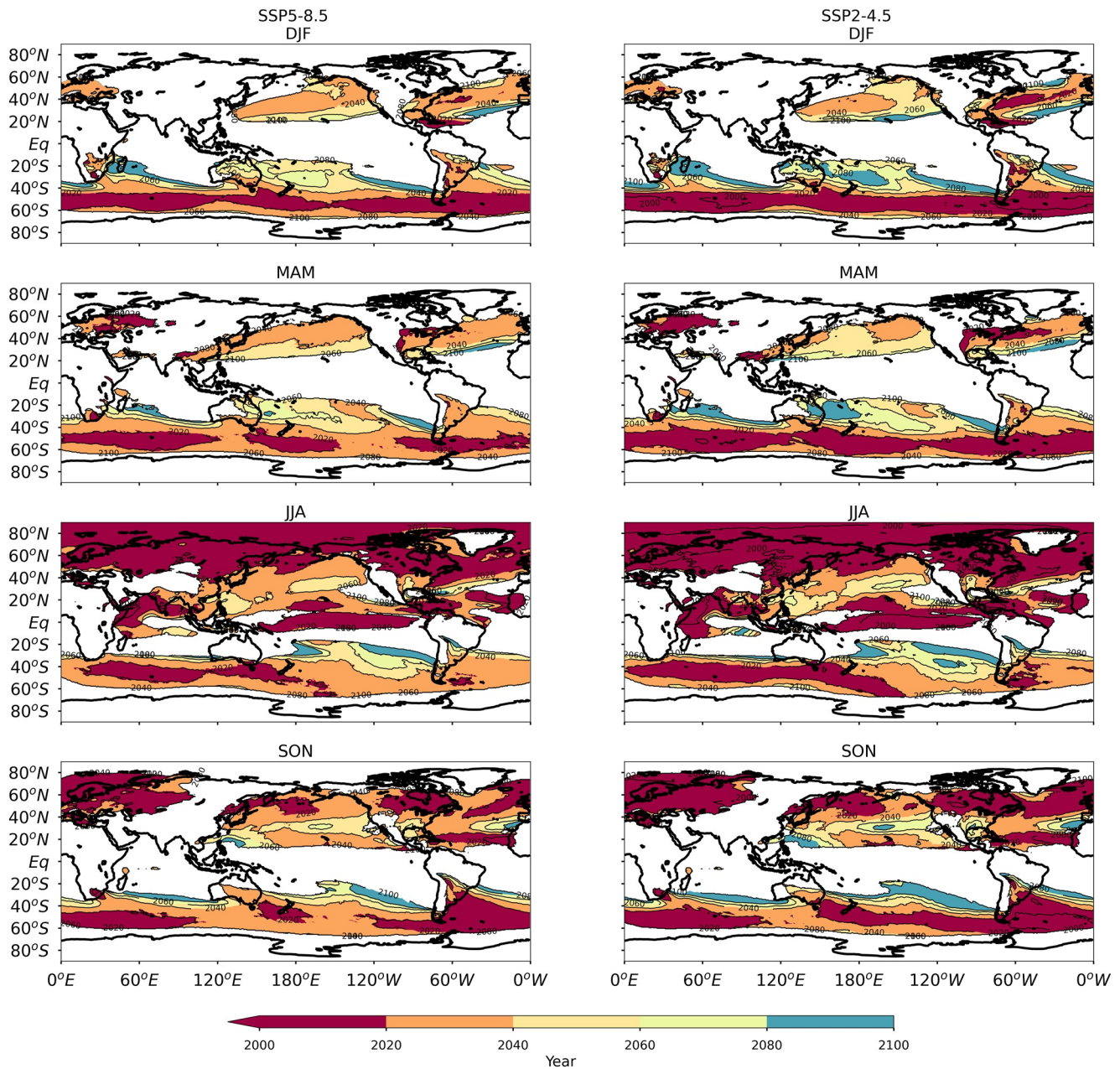
Figure 6. The time of emergence of extreme daily integrated water vapor transport (IVT), defined as the 95th percentile of daily IVT from all ensemble members in each individual season.

projected increase in both year-to-year annual precipitation variability and extreme precipitation associated with ARs. These results are consistent with our finding, while our ToE analysis further provides a timeline that may aid mitigation and adaptation measures.

3.4. Sensitivity of AR ToE to Time-Dependent Thresholds

While the preceding analysis focuses on the AR detection method of (Mundhenk et al., 2016), which uses stationary IVT thresholds based on the historical record, we now consider alternative methods for AR detection. Previous studies have documented the AR statistics detected by TECA-BARD and Tempest are quite different from the statistics detected by other algorithms according to AR Tracking Method Intercomparison Project (O'Brien
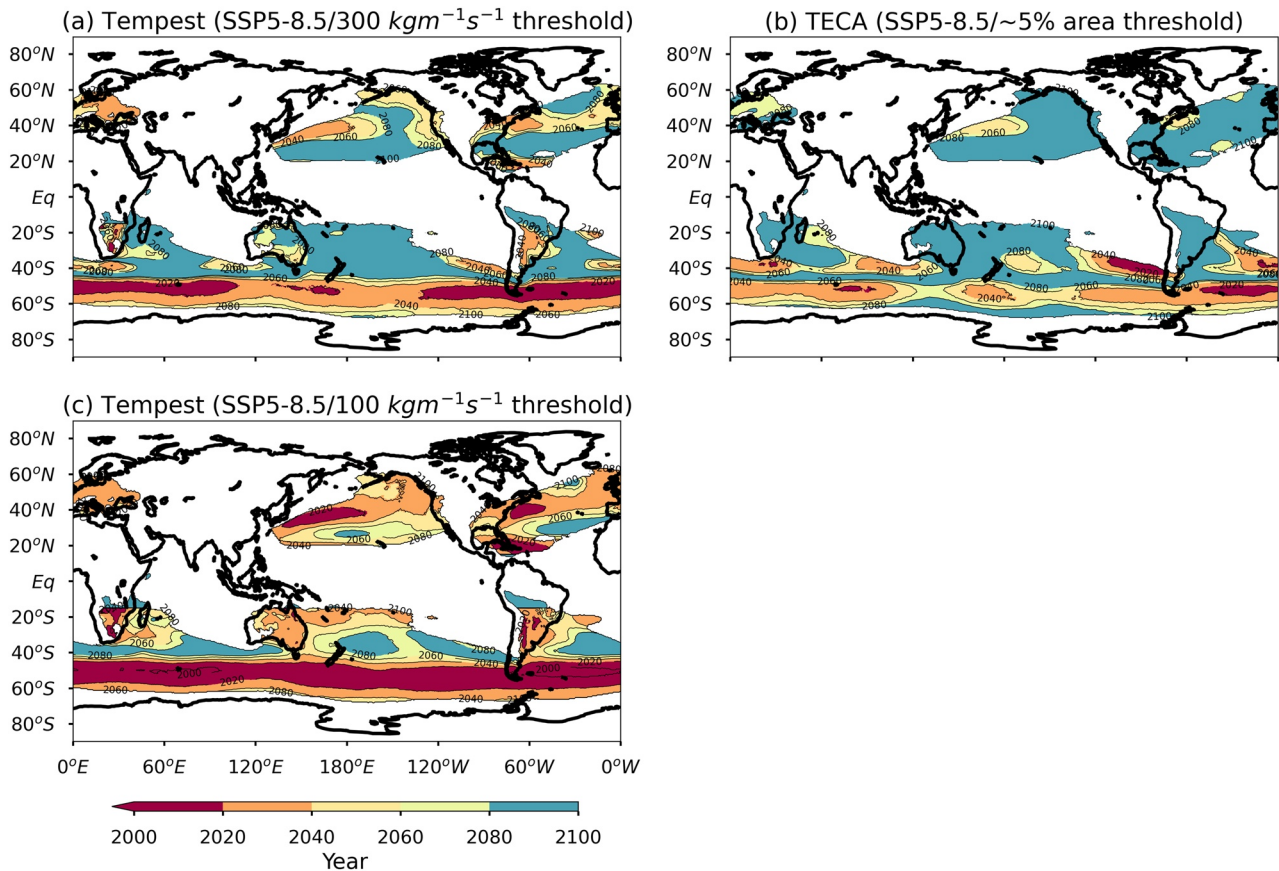
Figure 7. The December–February atmospheric river (AR)-day frequency time of emergence from SSP5-8.5 estimated by (a) Tempest with ∼300 kg·m⁻¹s⁻¹ threshold in integrated water vapor transport (IVT) difference between the core of an AR and its environment, (b) Toolkit for Extreme Climate Analysis, Bayesian AR Detector with the constraint that ARs occupy a maximum planetary area of ∼5%, (c) as in (a) but with a ∼100 kg·m⁻¹s⁻¹ threshold in IVT difference.

et al. (2021)). One key difference is the implementation of the time-dependent IVT threshold in Tempest and TECA-BARD. Specifically, Tempest uses a Laplacian operator to identify the local maximum of IVT while TECA-BARD constrains the total planetary area that ARs can cover. These thresholds are calculated independently at each time step. As the climate warms, the change in IVT difference between the core of an AR and its environment is not as big as the change in the background IVT. Thus, what Tempest detects is dominated by the meridional shift of storm tracks (i.e., the local maximum of eddy kinetic energy) instead of the increase of background moisture. This feature is evident in Figure S6 in Supporting Information S1, where we show the change in AR-day frequency as shown in Figure 1 except for Tempest. The implemented criterion is ∼300 kg m⁻¹s⁻¹ in IVT difference between the core of an AR and its surrounding environment. In Figure S6 in Supporting Information S1, we can see a clear poleward shift of the Southern Hemisphere storm track under increasing greenhouse gases. A similar feature is also found in Figure S7 in Supporting Information S1, where we apply TECA-BARD for AR detection. In Figure S7 in Supporting Information S1, the total area that ARs can cover is limited to ∼5% of the planetary area. The posterior probability of intensity threshold (i.e., pre-trained threshold based on experts' knowledge) is higher for the most extreme IVT (top 5%) due to stronger agreement by different experts (see Figure 5 in O'Brien, Payne, et al. (2020), O'Brien, Risser, et al. (2020)). Thus, only regions with the extreme IVT (e.g., top ∼5%) are more likely to be labeled as ARs. Along with the implemented criteria, the TECA-BARD elevates the minimum threshold of IVT as the climate warms due to the percentile calculated in space (in contrast to time). This time-dependent threshold limits the most salient AR changes to the regions with the strongest IVT and automatically removes the effect of increased background moisture.

The time-dependent thresholds also influence the estimation of AR ToE. In Figures 7a and 7b, we repeat the same ToE analysis, as in Figure 4, except for Tempest and TECA-BARD. We find that the AR-day frequency ToE is

delayed by ∼20 years in most regions relative to what is shown in Figures 4 and 5. This may not be a surprising result since the change in IVT is to first-order dominated by thermodynamics while the change in dynamics is relatively small (Gao et al., 2015). Implementation of time-dependent IVT thresholds substantially removes the contributions to changes from thermodynamics. To demonstrate how the time-dependent threshold influences the estimation of ToE, we conduct an additional test where we lower the time-dependent threshold in Tempest. The minimum difference between the core of an AR and its environment is lowered from ∼300 to ∼100 kg m^{-1}s^{-1}. By lowering the threshold, we allow ARs to grow in terms of area with the increase of background moisture. In Figure 7c, we can find the estimated ToE is nearly identical to the AR ToE based on Mundhenk et al. (2016) (Figure 4a) as well as the ToE of daily extreme IVT (Figure 6). This suggests that we can restore the thermodynamic components by lowering the time-dependent thresholds. Given the public interest in knowing the projected change in hydrological extremes, and the adaptation of infrastructure to the historical climate, considering the contributions from both dynamics and thermodynamics seems appropriate. With that said, Figure 7 demonstrates that detection algorithms with time-independent threshold still may fit this purpose and provide a reasonable timeline for mitigation and adaption when appropriate AR detection thresholds are chosen.

4. Conclusions and Remarks

Identifying the time when anthropogenically forced AR changes emerge from the noise of internal climate variability may benefit many sectors of society, including policy makers, water resource management, and disaster preparedness. However, different SSPs and user-defined thresholds in AR detection algorithms introduce additional uncertainty to the quantitative estimation of AR warming response. In this study, we leverage a large ensemble of warming simulations generated by a GFDL next-generation climate model, SPEAR, to explore the AR response to global warming and to determine when the forced response emerges from the background internal climate variability. We show that some highly populated extratropical regions, such as West Coast of North America and Western Europe may expect detectable changes in AR statistics relative to 1,920 levels before 2060 (1921 is the year with the lowest anthropogenic influence that we have available). This result is robust across different SSPs and time-independent thresholds.

Our analysis focuses on the AR detection method of Mundhenk et al. (2016), which uses time-independent thresholds based on the historical climate for AR detection, but we also analyze results from two other AR detection methods that use time-varying thresholds, which generally lead to a later ToE. The appropriate choice of AR detection method may depend on the application. If we assume that societal adaptation to the increase in background moisture and the associated increase in IVT is slow, then the time-independent thresholds of Mundhenk et al. (2016) may provide a more accurate characterization of the societal impacts of changing AR statistics. However, we also show that it is possible to reconcile differences among the detection algorithms by using lenient thresholds for the methods with time-varying thresholds. In these cases, the optimal thresholds for AR detection requires further study and may depend on the application.

We further examine a computationally efficient approach proposed by Thompson et al. (2015) and demonstrate the internal variability in a large ensemble of warming simulation can be quantified to a high degree of accuracy from the statistics of unforced components. One important application of the above conclusion is that we can estimate the ToE of AR changes with a single ensemble member or, in principle, with the single realization of nature.

The results of this study lead to a few interesting questions, which deserve further exploration in the future. First, since our results demonstrate that parametric fitting with a single ensemble member can capture the ensemble mean AR-day frequency changes with reasonable margins of error, can we reach similar conclusions, including a stable estimation of AR ToE, with different CMIP5/CMIP6 models? The sensitivity of AR warming response to model physics deserves exploration. Second, most previous research has shown that IVT response to global warming is dominated by thermodynamics (i.e., the change in specific humidity due to the change in temperature) while the dynamics plays a modest role. Deser et al. (2016) and Lehner et al. (2017) have used dynamical adjustment to isolate the dynamical components of projected climate changes. They demonstrate that internal climate variability over the extratropics is dominated by dynamical components while the anthropogenically forced response is closely tied to the thermodynamic component. Therefore, an open question is whether a dynamically adjusted AR data set can constrain uncertainty further and reaffirm the ToE estimates in our study. The result also can give us additional insights into the underlying mechanisms of the AR response to global warming.

Data Availability Statement

[Dataset] The data supporting the analysis in the main text is available at https://doi.org/10.5281/zenodo.6366550 (Tseng, 2022) and https://www.gfdl.noaa.gov/spear_large_ensembles/ (Delworth et al., 2020).

References

Corringham, T. W., Ralph, F. M., Gershunov, A., Cayan, D. R., & Talbot, C. A. (2019). Atmospheric rivers drive flood damages in the western United States. *Science Advances*, *5*(12), eaax4631. https://doi.org/10.1126/sciadv.aax4631

Delworth, T. L., Cooke, W. F., Adcroft, A., Bushuk, M., Chen, J.-H., Dunne, K. A., et al. (2020). SPEAR: The next generation GFDL modeling system for seasonal to multidecadal prediction and projection [Dataset]. Journal of Advances in Modeling Earth Systems, *12*(3), e2019MS001895. https://doi.org/10.1029/2019MS001895

Deser, C., Lehner, F., Rodgers, K. F., Ault, T., Delworth, T. L., DiNezio, P. N., et al. (2020). Insights from Earth system model initial-condition large ensembles and future prospects. *Nature Climate Change*, *10*(4), 277–286. https://doi.org/10.1038/s41558-020-0731-2

Deser, C., Terray, L., & Phillips, A. S. (2016). Forced and internal components of winter air temperature trends over North America during the past 50 years: Mechanisms and implications. *Journal of Climate*, *29*(6), 2237–2258. https://doi.org/10.1175/JCLI-D-15-0304.1

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

Espinoza, V., Waliser, D. E., Guan, B., Lavers, D. A., & Ralph, F. M. (2018). Global analysis of climate change projection effects on atmospheric rivers. *Geophysical Research Letters*, *45*(9), 4299–4308. https://doi.org/10.1029/2017GL076968

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

Gershunov, A., Shulgina, T., Clemesha, R. E. S., Guirguis, K., Pierce, D. W., Dettinger, M. D., et al. (2019). Precipitation regime change in western North America: The role of atmospheric rivers. *Scientific Reports*, *9*(1), 9944. https://doi.org/10.1038/s41598-019-46169-w

Hawkins, E., & Sutton, R. (2012). Time of emergence of climate signals. *Geophysical Research Letters*, *39*(1), L01702. https://doi.org/10.1029/2011GL050087

Held, I. M., Guo, H., Adcroft, A., Dunne, J. P., Horowitz, L. W., Krasting, J., et al. (2019). Structure and performance of GFDL's CM4.0 climate model. *Journal of Advances in Modeling Earth Systems*, *11*(11), 3691–3727. https://doi.org/10.1029/2019MS001829

Held, I. M., & Soden, B. J. (2006). Robust responses of the hydrological cycle to global warming. *Journal of Climate*, *19*(21), 5686–5699. https://doi.org/10.1175/JCLI3990.1. Retrieved from https://journals.ametsoc.org/view/journals/clim/19/21/jcli3990.1.xml

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

IPCC. (2013). Summary for policymakers [Book section]. In T.Stocker, et al. (Eds.), *Climate change 2013: The physical science basis. contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change* (pp. 1–30). Cambridge University Press. https://doi.org/10.1017/CBO9781107415324.004

Lehner, F., Deser, C., & Terray, L. (2017). Toward a new estimate of "time of emergence" of anthropogenic warming: Insights from dynamical adjustment and a large initial-condition model ensemble. *Journal of Climate*, *30*(19), 7739–7756. https://doi.org/10.1175/JCLI-D-16-0792.1

Li, J., Thompson, D. W. J., Barnes, E. A., & Solomon, S. (2017). Quantifying the lead time required for a linear trend to emerge from natural climate variability. *Journal of Climate*, *30*(24), 10179–10191. https://doi.org/10.1175/JCLI-D-16-0280.1

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

Nash, D., Waliser, D., Guan, B., Ye, H., & Ralph, F. M. (2018). The role of atmospheric rivers in extratropical and polar hydroclimate. *Journal of Geophysical Research: Atmospheres*, *123*(13), 6804–6821. https://doi.org/10.1029/2017JD028130

Newell, R. E., Newell, N. E., Zhu, Y., & Scott, C. (1992). Tropospheric rivers?—A pilot study. *Geophysical Research Letters*, *19*(24), 2401–2404. https://doi.org/10.1029/92GL02916

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

O'Brien, T. A., Risser, M. D., Loring, B., Elbashandy, A. A., Krishnan, H., Johnson, J., et al. (2020). Detection of atmospheric rivers with inline uncertainty quantification: TECA-BARD v1.0.1. *Geoscientific Model Development*, *13*(12), 6131–6148. https://doi.org/10.5194/gmd-13-6131-2020

O'Brien, T. A., Wehner, M. F., Payne, A. E., Shields, C. A., Rutz, J. J., Leung, L.-R., et al. (2021). Increases in future AR count and size: Overview of the ARTMIP Tier 2 CMIP5/6 experiment. *Journal of Geophysical Research: Atmospheres*, *127*(6), e2021JD036013. https://doi.org/10.1029/2021JD036013

O'Gorman, P. A., & Schneider, T. (2009). The physical basis for increases in precipitation extremes in simulations of 21st-century climate change. *Proceedings of the National Academy of Sciences*, *106*(35), 14773–14777. https://doi.org/10.1073/pnas.0907610106

Shaw, T., & Voigt, A. (2015). Tug of war on summertime circulation between radiative forcing and sea surface warming. *Nature Geoscience*, *8*(7), 560–566. https://doi.org/10.1038/ngeo2449

Shepherd, T. (2014). Atmospheric circulation as a source of uncertainty in climate change projections. *Nature Geoscience*, *7*(10), 703–708. https://doi.org/10.1038/ngeo2253

Shields, C. A., & Kiehl, J. T. (2016). Simulating the pineapple express in the half degree community climate system model, CCSM4. *Geophysical Research Letters*, *43*(14), 7767–7773. https://doi.org/10.1002/2016GL069476

Slinskey, E. A., Loikith, P. C., Waliser, D. E., Guan, B., & Martin, A. (2020). A climatology of atmospheric rivers and associated precipitation for the seven US national climate assessment regions. *Journal of Hydrometeorology*, *21*(11), 2439–24561. https://doi.org/10.1175/JHM-D-20-003910.1175/JHM-D-20-0039.1

Thompson, D. W. J., Barnes, E. A., Deser, C., Foust, W. E., & Phillips, A. S. (2015). Quantifying the role of internal climate variability in future climate trends. *Journal of Climate*, *28*(16), 6443–6456. https://doi.org/10.1175/JCLI-D-14-00830.1

Tseng, K.-C. (2022). Kuiper2000/AR_LE: Codes for "When will humanity notice its impacts on atmospheric rivers?" [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.6366550

Tseng, K.-C., Johnson, N. C., Kapnick, S. B., Delworth, T. L., Lu, F., Cooke, W., et al. (2021). Are multiseasonal forecasts of atmospheric rivers possible? *Geophysical Research Letters*, *48*(17), e2021GL094000. https://doi.org/10.1029/2021GL094000

Ullrich, P. A., & Zarzycki, C. M. (2017). TempestExtremes: A framework for scale-insensitive pointwise feature tracking on unstructured grids. *Geoscientific Model Development*, *10*(3), 1069–1090. https://doi.org/10.5194/gmd-10-1069-2017

White, A. B., Moore, B. J., Gottas, D. J., & Neiman, P. J. (2019). Winter storm conditions leading to excessive runoff above California's Oroville Dam during January and February 2017. *Bulletin of the American Meteorological Society*, *100*(1), 55–70. https://doi.org/10.1175/BAMS-D-18-0091.1

Wills, R. C., White, R. H., & Levine, X. J. (2019). Hemisphere stationary waves in a changing climate. *Current Climate Change Reports*, *5*(5), 372–389. https://doi.org/10.1007/s40641-019-00147-6

Zhang, Z., Ralph, F. M., & Zheng, M. (2019). The relationship between extratropical cyclone strength and atmospheric river intensity and position. *Geophysical Research Letters*, *46*(3), 1814–1823. https://doi.org/10.1029/2018GL079071

Zhao, M. (2020). Simulations of atmospheric rivers, their variability and response to global warming using GFDL's new high resolution general circulation model. *Journal of Climate*, *33*(23), 1–46. https://doi.org/10.1175/JCLI-D-20-0241

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