The role of air–sea interactions in atmospheric river events: Case studies using the SKRIPS regional coupled model

Rui Sun¹, Aneesh C. Subramanian², Bruce D. Cornuelle¹, Matthew R. Mazloff³, Arthur J. Miller¹, Hyodae Seo³, Ibrahim Hoteit⁴

¹Scripps Institution of Oceanography, La Jolla, California, USA
²Department of Atmospheric and Oceanic Sciences, University of Colorado Boulder, Colorado, USA
³Physical Oceanography Department, Woods Hole Oceanographic Institution, Massachusetts, USA
⁴Physical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

Key Points:
- The SKRIPS regional coupled model is used to hindcast a series of AR events.
- The coupled model better reproduces ARs than the uncoupled model with persistent SST, especially when strong SST cooling is observed.
- The coupled model has more skill below 850 hPa in modeling the atmospheric state (e.g., IWV and IVT).

Corresponding author: Rui Sun, rus043@ucsd.edu
Abstract
Atmospheric rivers (ARs) play a key role in California’s water supply and are responsible for most of the extreme precipitation events along the west coast of North America. Given the high societal impact, it is critical to improve our understanding and forecasting ability of ARs. In this study we use a regional coupled ocean–atmosphere modeling system to make hindcasts of several ARs up to 14 days. We investigate the role of air–sea interactions in AR events by comparing coupled model hindcasts to hindcasts made using persistent sea surface temperature (SST). Two groups of ARs are highlighted in the comparison: (1) ARs occurring during times with strong SST cooling and (2) ARs occurring during times with weak SST cooling. During the events with strong SST cooling, the coupled model simulates strong upward air–sea heat fluxes associated with ARs; on the other hand, when the SST cooling is weak, the coupled model simulates downward air–sea heat fluxes in the AR region. Validation data shows that the coupled model is able to skillfully reproduce the evolving SST, as well as the turbulent heat transfers between the ocean and atmosphere. To evaluate the influence of the ocean on ARs, we analyze the vertically integrated water vapor and water vapor transport. During strong SST cooling AR events the simulated IWV is improved by about 12% in the coupled run at lead times greater than one week. For IVT, which is about 1.8 times more variable, the improvement in the coupled run is about 5%.

Plain Language Summary
Atmospheric rivers (ARs) play a key role in extreme precipitation along the west coast of North America. Because of their important societal impact, an improved understanding of ARs is critical. In the present work, we use a coupled ocean–atmosphere modeling system to investigate the role of air–sea interactions in simulating ARs. We highlight two groups in our simulations for which the ocean’s response to ARs differs. One group is associated with strong ocean cooling, where the ocean cools everywhere. The other group is associated with weak ocean cooling, where the ARs can warm part of the ocean. We investigate the AR water vapor content and transport to evaluate the ocean impact on ARs. We find that the coupled model better simulates the air–sea exchanges and AR water vapor content. The improvements are more significant during the AR events associated with strong ocean cooling.

1 Introduction
Atmospheric rivers (ARs) are narrow, elongated plumes of enhanced water vapor transport over the oceans that can extend from the tropics and subtropics into the extratropics (F. M. Ralph et al., 2004, 2005; Bao et al., 2006; Jankov et al., 2009). Many studies over the past three decades have helped explain the atmospheric processes governing AR dynamics and thermodynamics (e.g., Newell et al., 1992; Zhu & Newell, 1998; F. M. Ralph et al., 2004; F. Ralph et al., 2010; Gimeno et al., 2014). ARs produce 25-50% of the annual precipitation in key areas of the western United States and are responsible for most of the extreme precipitation and flooding events in California (F. M. Ralph et al., 2004; Neiman et al., 2008; Leung & Qian, 2009; M. D. Dettinger et al., 2011; M. Dettinger & Cayan, 2014; Gershunov et al., 2019). ARs can have both beneficial (e.g., replenishing water reservoirs) and detrimental (e.g., causing destructive floods and landslides) impacts on regional economies and public safety (DeFlorio et al., 2018). Since they play such important societal roles, improved understanding and accurate forecasting of ARs and AR-induced precipitation are critical (F. Ralph et al., 2010; Martin et al., 2018).

To better understand ARs, numerical weather prediction models have been used to simulate and forecast ARs over the last several decades (Wick et al., 2013; Nayak et al., 2014; Lavers et al., 2016; DeFlorio et al., 2018; Martin et al., 2018). For example, Wick et al. (2013) assessed the ability of five global operational ensemble forecast sys-
tems in forecasting AR events, focusing on integrated water vapor (IWV). The models were skillful in predicting the overall occurrence of ARs out to 10 days, with the forecast skill degrading with increasing lead time. They also investigated the influence of model spatial resolution on forecasting ARs and found that the error in AR width is greater in coarser-resolution models. Lavers et al. (2016) investigated the global ensemble reforecasts of integrated vapor transport (IVT) and precipitation across 31 winters. Their results showed that IVT has higher predictability than precipitation, suggesting that IVT may be used to provide early awareness of extreme AR events. They also found large interannual variability in predicting IVT and precipitation. Martin et al. (2018) compared the forecasts of global and regional models against the observations. They demonstrated that improving the water vapor transport accuracy can significantly reduce precipitation error in the regional model, while this was not observed in the global model.

To extend the predictability of ARs by numerical weather prediction models, recent studies focused on the connection between ARs and synoptic-scale atmosphere features. This is because AR location, intensity, and frequency are strongly modulated by lower frequency variabilities, such as the El Niño Southern Oscillation (ENSO), the Madden–Julian Oscillation (MJO), the Pacific Decadal Oscillation (PDO), and the Pacific North America (PNA) teleconnection patterns (Guan et al., 2013; Mundhenk et al., 2016; Payne & Magnusdottir, 2014; Gershunov et al., 2017; Baggett et al., 2017; Zhou & Kim, 2018). These studies suggest that AR predictability can be potentially extended through the knowledge of these lower frequency signals. In addition, DeFlorio et al. (2018) studied the combined effect of lower frequency signals on AR predictability. They showed that AR predictability has increased over the north Pacific/western United States at a 10-day lead during El Niño + positive PNA conditions and over the north Atlantic/United Kingdom at a 7-day lead during La Niña + negative PNA conditions.

Air–sea interactions can also impact ARs and their predictability. Recent studies emphasized on the importance of AR-induced strong winds (Waliser & Guan, 2017; Shinoda et al., 2019). These winds are often associated with large pressure gradients between the extratropical cyclone and anticyclone located on the southeast and northwest sides of ARs (e.g., Newell et al., 1992; Newman et al., 2012; Shinoda et al., 2019). Large air–sea fluxes of momentum, heat, and moisture then result from the strong winds, generating a substantial ocean responses. Neiman et al. (2013) investigated a few landfalling AR events and showed that the upward latent heat flux can be 200 W/m² in the AR region, and even higher on the northwest side of AR at 550 W/m². The recent study of Shinoda et al. (2019) showed a dipole-like structure that cooler/warmer SST is observed on the northeast/southwest side of the AR center due to strong surface winds and air–sea heat fluxes. The AR-induced sea surface temperature (SST) variations and air–sea fluxes could feedback on the ARs and play a critical role in their evolution. However, although there are many studies on AR dynamics and thermodynamics (e.g., F. M. Ralph et al., 2004; F. Ralph et al., 2010; Martin et al., 2018; Shinoda et al., 2019), very little is known about the influence of air–sea interactions on modeling and forecasting ARs. There are still fundamental questions to be addressed:

1. How do ARs impact the ocean?
2. How does the ocean impact ARs?
3. Can a coupled ocean–atmosphere model better simulate AR events?

The goal of this work is to investigate the influence of air–sea interactions on AR events. To this end, we perform a series of coupled and uncoupled numerical simulations in the northeastern Pacific region, where ARs have been well-studied. We first present the SST variations and the ocean surface heat fluxes in a series of AR events, aiming to show how ARs impact the ocean. Then, by comparing the coupled and uncoupled runs, we are able to isolate the effect of SST variations to investigate how the ocean impacts
ARs. Finally, we use observational and reanalysis data to quantify the difference in skill between the coupled and uncoupled simulations.

The rest of this paper is organized as follows. The coupled model, the design of the experiments, and the data used in this work are introduced in Section 2. An overview of the AR events is presented in Section 3. Section 4 details the impact of air–sea interactions on modeling AR events. Section 5 discusses IWV and IVT skill, and assesses sources of errors. The last section concludes the paper.

2 Methodology

2.1 Coupled Model

In this case study, the Scripps–KAUST Regional Integrated Prediction System (SKRIPS, version 1.0) is used (Sun et al., 2019). The SKRIPS is a regional coupled ocean–atmosphere model: the oceanic model component is the MIT general circulation model (MITgcm) (Marshall et al., 1997) and the atmospheric model component is the Weather Research and Forecasting (WRF) model (Skamarock et al., 2019). The Earth System Modeling Framework (ESMF) (Hill et al., 2004) is used as the coupler to drive the coupled simulation. The National United Operational Prediction Capability (NUOPC) layer in the ESMF is also used to simplify the implementations of component synchronization, execution, and other common tasks in the coupling (Hill et al., 2004; Sitz et al., 2017). The schematic description of the coupled model is shown in Fig. 1. In the coupling process, the MITgcm sends SST and ocean surface velocity to ESMF, and ESMF sends them to WRF as the bottom boundary conditions. WRF sends surface fields to ESMF, including (1) surface longwave and shortwave radiative flux, (2) surface latent and sensible turbulent heat flux, (3) 10-m wind speed, (4) precipitation, and (5) evaporation. The MITgcm uses these variables to prescribe surface forcing, including (1) total net surface heat flux, (2) surface wind stress, and (3) freshwater flux. The total net surface heat flux is computed by adding latent heat flux, sensible heat flux, shortwave radiation flux, and longwave radiation flux. The surface wind stress is computed by using the 10-m wind speed (Large & Yeager, 2004). The freshwater flux is the difference between precipitation and evaporation. The latent and sensible heat fluxes are computed using the COARE 3.0 bulk algorithm in WRF (Fairall et al., 2003).

2.2 Experimental Design

The AR events in the northeastern Pacific region are investigated. We perform 93 pairs of hindcast simulations initialized on each day in three Januaries from 2016 to 2018 (3 years × 31 days/year). We select these events because they capture different thermodynamic characteristics of ARs, which will be detailed in Section 4. Each simulation aims to examine the model skill up to 14 days and the ensemble of the runs allow us to examine the mean and spread of the hindcasts. In each simulation, a few ARs (about 5 AR events) can be observed throughout the domain, and the duration of ARs can be a few days.

The model domain extends from 18.16°N to 54°N and from 116°W to 180°. To generate the grids, we choose latitude–longitude (cylindrical equidistant) map projection for both MITgcm and WRF. The horizontal grid resolution has 448×800 (lat×long) points and grid spacing is 0.08° in both directions. We use identical grids for both MITgcm and WRF to eliminate the issue of regridding winds near steep orography and complex coastlines (Seo et al., 2016). There are 40 sigma layers in the atmosphere model and 60 z-layers in the ocean model. The top of the atmosphere is at the 50 hPa pressure level.

The bathymetry of the ocean model is extracted from the 2-minute Gridded Global Relief Data (ETOPO2) (National Geophysical Data Center, 2006). The time step of the
Figure 1. The schematic description of the regional coupled ocean–atmosphere model. The yellow block is the ESMF/NUOPC coupler; the white blocks are the ocean and atmosphere components; the red blocks are the implemented MITgcm–ESMF and WRF–ESMF interfaces.

The ocean model is 120 seconds. The horizontal sub-grid mixing is parameterized using non-linear Smagorinsky viscosities, and the K-profile parameterization (KPP) (Large et al., 1994) is used for vertical mixing processes. The time step for the atmospheric simulation is 30 seconds. The Morrison 2-moment scheme (Morrison et al., 2009) is used to resolve the microphysics. The updated version of the Kain–Fritsch convection scheme (Kain, 2004) is used. The Mellor–Yamada–Nakanishi–Niino (MYNN) 2.5-order closure scheme (Nakanishi & Niino, 2004, 2009) is used for the planetary boundary layer (PBL), and the Rapid Radiation Transfer Model for GCMs (RRTMG; Iacono et al. (2008)) is used for longwave and shortwave radiation transfer through the atmosphere. The Noah land surface model is used for the land surface processes (Tewari et al., 2004).

In the present study, we perform the following simulations:

2. Run ATM.STA: stand-alone atmosphere (WRF) simulations with the initial SST kept persistent. This run serves as a benchmark to highlight the difference between the coupled and uncoupled runs. It allows assessing the atmospheric model behavior with realistic, but persistent SST.

Both CPL and ATM.STA are initialized using global analysis data. The initial conditions, boundary conditions, and forcing terms of the simulations are summarized in Table 1. In CPL, the ocean model uses the assimilated HYCOM/NCODA 1/12° global analysis data (http://hycom.org/data-server/glb-analysis) as initial and boundary conditions for ocean temperature, salinity, and horizontal velocities. The boundary conditions for the ocean are updated on a daily basis and are linearly interpolated between two prescribed records A restoring layer with a thickness of 13 grid cells is applied at the lateral boundaries. The inner and outer boundary relaxation timescales are 10 and 0.5 days, respectively. The atmosphere is initialized using the NCEP FNL (Final) Operational Global Analysis data. The same data also provide the boundary conditions for air temperature, wind speed, and air humidity every 6 hours. The atmosphere boundary conditions are also linearly interpolated between two prescribed records. The `spec-
ified’ zone in WRF prescribes the lateral boundary values, and the ‘relaxation’ zone is
used to nudge the solution from the domain toward the boundary condition value. Here
we use the default width of one point for the specific zone and four points for the relax-
ation zone.

Importantly, both CPL and ATM.STA runs derive skill from boundary conditions
(i.e. they are dynamically downscaled hindcasts). This better allows us to focus on high-
lighting the extent to which air–sea interactions impact ARs. In the CPL run HYCOM/NCODA
data is used for the oceanic initial and lateral boundary conditions. Thus in the ATM.STA
runs, HYCOM/NCODA SST is used as the initial condition and is persistent through-
out the run. The atmospheric initial and lateral boundary conditions in ATM.STA are
the same as in the coupled run. The coupling interval used for the coupled run (CPL)
is 20 minutes to allow capturing the diurnal cycle of air–sea fluxes (Seo et al., 2014). In
this study, we do not compare the coupled run with atmosphere-only model driven by
daily SST from HYCOM or other SST datasets. This is because (1) we aim to show the
difference in IWV and IVT due to the coupling and (2) daily SST may not be available
in a real-time forecast.

Table 1. The initial condition, boundary condition and forcing terms used in present simula-
tions.

<table>
<thead>
<tr>
<th>run</th>
<th>CPL</th>
<th>ATM.STA</th>
</tr>
</thead>
<tbody>
<tr>
<td>initial and</td>
<td>NCEP FNL (atmosphere)</td>
<td>NCEP FNL</td>
</tr>
<tr>
<td>boundary</td>
<td>HYCOM/NCODA (ocean)</td>
<td></td>
</tr>
<tr>
<td>conditions</td>
<td>from MITgcm</td>
<td>HYCOM/NCODA (persistent)</td>
</tr>
<tr>
<td>ocean surface</td>
<td></td>
<td></td>
</tr>
<tr>
<td>conditions</td>
<td>from WRF</td>
<td>not necessary</td>
</tr>
<tr>
<td>(for ocean</td>
<td></td>
<td></td>
</tr>
<tr>
<td>model only)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.3 Validation of the results

To evaluate the performance of CPL and ATM.STA runs, the model outputs are
compared with validation data. For water vapor during the AR events we compare IWV
and IVT with ERA5 reanalysis data (ECMWF, 2017). IWV is calculated from specific
humidity $q$ (kg kg$^{-1}$) in the atmosphere:

$$\text{IWV} = \frac{1}{g} \int_{p_{\text{surface}}}^{300 \text{ hPa}} q dp,$$

where $g$ is the gravitational acceleration (equal to 9.81 m s$^{-2}$); and $p$ is pressure (Pa).

IVT is calculated from specific humidity and wind speed:

$$\text{IVT} = \frac{1}{g} \sqrt{ \left( \int_{p_{\text{surface}}}^{300 \text{ hPa}} qu dp \right)^2 + \left( \int_{p_{\text{surface}}}^{300 \text{ hPa}} qv dp \right)^2 },$$

where $u$ and $v$ are the zonal and meridional wind speeds (m s$^{-1}$), respectively. Note that
we integrate IWV and IVT from the surface pressure $p_{\text{surface}}$ to 300 hPa (Lavers et al.,
2016). To better illustrate the difference between model outputs and validation data, IWV
and IVT are averaged on a daily basis (ending at 0000 UTC) using hourly instantaneous
diagnostics (Lavers et al., 2015; Hecht & Cordeira, 2017). We use ERA5 data to vali-
date the IWV and IVT in AR events (Demirdjian et al., 2020). The simulated turbu-
lent heat fluxes (THFs) are validated against the OAFlux data (Yu et al., 2008). The
simulated SST fields are validated against the HYCOM/NCODA global analysis data. Note that we interpolate the validation data onto the model grid to achieve a uniform spatial scale. The validation data are summarized in Table 2. The differences of IWV/IVT, THFs, and SST with validation data are also quantified outside the AR regions, and thus we present these quantities for the entire simulation.

The Brier skill score (BSS) is used to examine the skill difference between CPL and ATM.STA runs (Von Storch & Zwiers, 2001). Here, we use the modified version that simplifies the comparability of positive and negative scores (Winterfeldt et al., 2011):

\[
BSS = \begin{cases} 
1 - \frac{\sigma_F^2 \sigma_R^2}{\sigma_F^2}, & \text{if } \sigma_F^2 \leq \sigma_R^2, \\
\frac{\sigma_R^2}{\sigma_F^2 - 1}, & \text{if } \sigma_F^2 > \sigma_R^2,
\end{cases}
\]

where \( \sigma_F^2 \) and \( \sigma_R^2 \) are the mean squared error (MSE) of the “forecast” and “reference”, respectively. According to Eq. (3), positive BSS means the forecast is more skillful than the reference, whereas negative BSS means the forecast is less skillful than the reference.

Table 2. The dataset used to validate the simulation results.

<table>
<thead>
<tr>
<th>variable</th>
<th>validation data</th>
</tr>
</thead>
<tbody>
<tr>
<td>interpolated water vapor (IWV)</td>
<td>ERA5</td>
</tr>
<tr>
<td>interpolated vapor transport (IVT)</td>
<td>ERA5</td>
</tr>
<tr>
<td>latent and sensible heat</td>
<td>OAFlux</td>
</tr>
<tr>
<td>sea surface temperature (SST)</td>
<td>HYCOM/NCODA</td>
</tr>
</tbody>
</table>

3 Overview of the AR events

A series of AR events with different thermodynamic interactions are observed in the simulations. To illustrate the different characteristics of ARs, the results obtained in two representative coupled simulations are shown: CASE1 initialized at 0000 UTC Jan 09 2018; and CASE2 at 0000 UTC Jan 25 2018. The evolution of the ARs is shown in Fig. 2 by plotting the daily-averaged IVT fields 4, 6, 8, and 10 days after initiation. Here, we use IVT > 250 kg m\(^{-1}\)s\(^{-1}\) to define the AR region (Moore et al., 2012; Rutz et al., 2014). It can be seen in Fig. 2 that ARs are observed in the selected domain throughout the simulations. Figure 2(a) shows several west–east oriented ARs in CASE1, with a maximum IVT of about 1250 kg m\(^{-1}\)s\(^{-1}\), whereas Fig. 2(b) shows CASE2 has several ARs with a more south–north orientation and with a maximum IVT of about 900 kg m\(^{-1}\)s\(^{-1}\).

Figure 3 displays the 14-day averaged IVT and the number of days under AR conditions. The mean IVT in CASE1 is higher than that of CASE2, but the AR events cover similar regions in both cases.

To demonstrate different AR thermodynamic interactions in CASE1 and CASE2, the 14-day averaged THFs, the 14-day time-integrated \( Q_{net} \) (net surface heat flux), and the SST difference between day 14 and day 1 (dSST\(_{14}\)) are plotted in Fig. 4. The THFs in Fig. 4(a) indicate the ocean is losing energy from turbulent heat transfer in both cases. The energy loss in CASE1 (mean THFs: -130 W m\(^{-2}\)) is more significant than that in CASE2 (mean THFs: -103 W m\(^{-2}\)). In the extratropical region, the maximum daily-mean upward latent heat flux (LHF) in CASE1 (466 W m\(^{-2}\), observed in Jan 13) is smaller than that of extreme extratropical hurricanes (713 W m\(^{-2}\) in Olabarrieta et al. (2012)), but it is comparable to the mean LHF within the 250-km radius from the storm center (about 500 W m\(^{-2}\) in Li and Pu (2008) and Olabarrieta et al. (2012)). Figure 4(b) shows the time-integrated \( Q_{net} \) in the representative cases. In both cases, the mean \( Q_{net} \) in the domain is negative, indicating the ocean loses energy. However, in CASE2 the ocean gains
**Figure 2.** The daily-averaged IVT in two representative coupled simulations. The snapshots show the IVT after 4, 6, 8, and 10 days from the simulation initial time. The black contours denote the AR region where IVT $> 250$ kg m$^{-1}$ s$^{-1}$. CASE1 in panel (a) is initialized at 0000 UTC, Jan 07, 2018; CASE2 in panel (b) is initialized at 0000 UTC, Jan 25, 2018.

**Figure 3.** The averaged IVT and the number of “AR days” in CPL runs. Panel (a) plots the 14-day averaged IVT in CASE1 and CASE2, and the contours indicate the regions where 14-day averaged IVT $> 250$ kg m$^{-1}$ s$^{-1}$. Panel (b) shows the number of “AR days” with daily averaged IVT $> 250$ kg m$^{-1}$ s$^{-1}$, and the contours highlight the regions that are under AR condition for more than 6 days.
energy in the AR region of about $0.4 \times 10^8 \text{ J m}^{-2}$ between $160^\circ-140^\circ \text{ W and } 18^\circ-42^\circ \text{ N}$. Compared with CASE1, the total surface energy loss in CASE2 is only about half of that in CASE1 (CASE1: $2.34 \times 10^{21} \text{ J};$ CASE2: $1.14 \times 10^{21} \text{ J}$). Compared with tropical cyclones (Liu et al., 2011; Andersen et al., 2013; Trenberth et al., 2018), the ARs impact larger regions and thus the ocean loses more energy. Figure 4(c) shows the SST difference between the start and the end of the simulations. In CASE1, because the ocean loses heat, SST cooling is observed in the AR region, whereas SST warming is observed in CASE2, especially in the AR region where $Q_{net}$ is positive (between $160^\circ-140^\circ \text{ W and } 18^\circ-42^\circ \text{ N}$). Despite the SST warming in the AR region for CASE2, the domain mean SST differences (dSST$_{14}$) are negative for both cases (CASE1: -0.49 $^\circ\text{C};$ CASE2: -0.07 $^\circ\text{C}$).

4 Case Study

The comparison of the two representative cases in Section 3 demonstrates the different ARs thermodynamic interactions. In CASE1, SST cools about 1 $^\circ\text{C}$ in the AR region, whereas in CASE2, SST warming is observed in parts of the AR region. Here, we first examine the SST evolution in all coupled simulations and use the statistics (e.g., mean, standard deviation, ensemble spread) to demonstrate how ARs impact the ocean. We then investigate the ocean impact on ARs by comparing coupled and uncoupled simulation results. We focus on difference statistics rather than individual simulations due to the chaotic nature of atmosphere. However snapshots of individual AR simulations are shown in the Appendix.

4.1 Sea Surface Temperature

The evolution of SST in all 93 simulations is summarized in Fig. 5. It can be seen in Fig. 5(a) that coupled simulations generally reproduce the evolution of domain-averaged SST in consistency with HYCOM (mean error < 0.2 $^\circ\text{C}$; root-mean-square error < 0.6 $^\circ\text{C}$). The figure also shows variance in SST trends: in some cases the SST cooling can be as significant as in CASE1; in some other cases the SST variations are less significant. Because of different SST trends in all simulations, we highlight two groups of simulations that each have 31 members (1/3 of all simulations) in Fig. 5(b): (1) strong cooling ARs and (2) weak cooling ARs. The strong cooling ARs include 31 events that have more significant SST cooling (mean dSST$_{14}$: $-0.50 \ ^\circ\text{C}$); the weak cooling ARs include 31 events that have less significant SST cooling (mean dSST$_{14}$: $-0.22 \ ^\circ\text{C}$). Note that in weak cooling AR events, the SST may increase in parts of the AR region (example shown in Fig. 4), but the domain-averaged SST is still cooling. Here, we use the magnitude of SST cooling to separate the AR events because (1) SST changes are determined by the surface heat fluxes that are important in ocean–atmosphere coupling, and (2) SST is used as the boundary condition in the atmospheric model. The 31 intermediate cooling events are included in the ”all AR” statistics presented, but are not shown in isolation.

The SST simulated with the coupled model is now compared with the validation data to demonstrate the skill improvement over assuming a persistent SST (Fig. 6). In Fig. 6(a) we plot the root-mean-square errors (RMSEs) of SST obtained in CPL as a function of lead time in red. The upper (lower) whiskers represent maximum (minimum) values; the upper (lower) box bounds represent upper (lower) quartile $Q_1$ ($Q_3$); the box center lines represent median RMSEs for the 93 simulations. The interquartile range is $IQR = Q_3 - Q_1$, and the values above the upper (lower) fence $Q_1 + 1.5IQR$ ($Q_3 - 1.5IQR$) are outliers. In comparison, the RMSEs of persistent SST are also plotted in gray. It can be seen that the median, the upper/lower quartiles, and the maximum/minimum RMSE$_{SST}$ in CPL are all smaller than persistence from day 1 to day 14. Because the persistent SST is used in ATM.STA, this demonstrates that the SST in CPL agrees better with the validation data than ATM.STA. In addition, we plot BSS$_{SST}$ to quantify the improved skill in CPL (Fig. 6b). Here, $\sigma_{SST}^2$ in Eq. (3) is calculated between HYCOM data and the sim-
Figure 4. The mean THF, the time-integrated $Q_{\text{net}}$, and the SST difference in two representative coupled simulations. In panel (a) and (b), the positive values denote downward heat fluxes that warm the ocean; the negative values denote upward heat fluxes that cool the ocean. In panel (c), the positive values indicate warming SST; the negative values indicate cooling SST. The left panels are showing CASE1 that is initialized at 0000 UTC, Jan 07, 2018; the right panels are showing CASE2 that is initialized at 0000 UTC, Jan 25, 2018. The black contours highlight the AR region where IVT > 250 kg m$^{-1}$ s$^{-1}$. 
Figure 5. The evolution of domain-averaged SST from all coupled simulations in comparison with HYCOM data. Panel (a) shows the SST evolution throughout all simulations in each year; panel (b) highlights the SST trend in strong and weak cooling AR events in all simulations. Here, we highlight two groups of simulations that each has 31 members (1/3 of all simulations). The strong cooling ARs include 31 events that have more significant SST cooling (mean $dSST_{14}$: $-0.50^\circ$C); the weak cooling ARs include 31 events that have less significant SST cooling (mean $dSST_{14}$: $-0.22^\circ$C). The dashed line in Panel (A) is the daily climatology SST (Banzon et al., 2014).
ulated SST obtained in CPL; $\sigma^2_R$ is calculated between HYCOM data and the persistent SST used in ATM.STA. It can be seen that the median BSSs for all AR events are about 0.20 from day 1 to day 14. The skill differences between strong and weak cooling AR events are less than 0.10 in the first week (Table 3). However, in the second week, the BSSs of strong cooling ARs are higher than those of weaker cooling events by about 0.45, resulting from the combined effect of the coupled model being able to skillfully simulate the stronger SST changes as well as persistence being less skillful during the strong cooling events.

![Figure 6.](image)

**Figure 6.** Evaluation of coupled model skill in simulating SST. The SST obtained in CPL is compared with persistent SST used in ATM.STA. The SST data are validated against HYCOM/NCODA data. Panel (a) shows the RMSEs plotted as functions of hindcast lead time; panel (b) shows the evolution of BSSs. Note that each marker in the background represents the raw data in each simulation.

<table>
<thead>
<tr>
<th></th>
<th>week 1</th>
<th>week 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>all ARs</td>
<td>all ARs (strong cooling/weak cooling)</td>
<td>all ARs (strong cooling/weak cooling)</td>
</tr>
<tr>
<td>SST</td>
<td>21.2% (26.8%/18.7%)</td>
<td>17.8% (44.6%/6.4%)</td>
</tr>
<tr>
<td>THF</td>
<td>10.3% (11.6%/10.2%)</td>
<td>15.0% (20.0%/11.2%)</td>
</tr>
<tr>
<td>IWV</td>
<td>2.7% (3.0%/1.9%)</td>
<td>6.9% (11.7%/4.0%)</td>
</tr>
<tr>
<td>IVT</td>
<td>0.2% (0.2%/0.6%)</td>
<td>1.6% (4.6%/-0.6%)</td>
</tr>
</tbody>
</table>

### 4.2 Turbulent Heat Fluxes

In the AR events associated with stronger SST cooling, stronger turbulent heat losses of the ocean can be observed; in the AR events associated with weaker SST cooling, there is much less turbulent heat transfer between ocean and atmosphere. This section aims to demonstrate how the coupled model better simulates the turbulent heat transfer during the AR events.

To demonstrate how the coupled model better simulates the THFs, the comparison between the simulations and validation data is shown in Fig. 7. In Fig. 7(a), the RM-
SEs of THFs are plotted as functions of lead time. It can be seen that the RMSEs of CPL are smaller than those of ATM.STA from day 1 to day 14. Note that the RMSEs do not increase significantly (less than 5 W m$^{-2}$) for longer lead simulations. To quantify the improvement of the coupled model, the BSSs are shown in Fig. 7(b). Here, $\sigma_F^2$ in Eq. (3) is calculated between the OAFlux data and the THFs obtained in CPL; $\sigma_R^2$ is calculated between the OAFlux data and the THFs obtained in ATM.STA. It can be seen in Fig. 7 that the median of BSSs are increasing from 0.06 to 0.17 with lead time. The difference of BSSs between strong and weak cooling AR events are about 0.01 in week one, outlined in Table 3. However, in week two, the BSSs in strong cooling events are much higher than those in weak cooling ARs (strong cooling ARs: 0.20; weak cooling ARs: 0.11). This indicates that the skill improvement of the coupled model is more significant for strong cooling AR events.

Figure 7. Comparison of the THF skill between CPL and ATM.STA runs. The THF data are validated against the OAFlux data. Panel (a) shows the RMSEs plotted as functions of lead times; panel (b) shows the evolution of BSSs. Note that each marker in the background represents the raw data obtained in each simulation.

4.3 Improved Skills in Simulating ARs

Because of the improved skill of the coupled model in simulating SST and THF, the question arises whether the coupled model can also better simulate the ARs. This section investigates how much skill is added by the coupled model in simulating ARs. The diagnosed IWV and IVT are used to demonstrate the influence of air–sea interactions on ARs.

The RMSEs of both IWV and IVT are shown in Fig. 8, along with the errors of persistent values. It can be seen that the RMSEs of CPL and ATM.STA runs are only 25% of those of persistent values, showing both coupled and uncoupled models have much better skills than persistent. In week one, the RMSE$_{IWV}$ and RMSE$_{IVT}$ of CPL do not differ much from those of ATM.STA. In week two, the RMSEs of CPL are more significant: the median RMSE$_{IWV}$ of CPL is smaller by about 0.1 mm and the median RMSE$_{IVT}$ of CPL is smaller by about 1 kg m$^{-1}$ s$^{-1}$. It is noted that there are a few simulations that have more than twice larger RMSEs than the median, but the model outputs are still better than the persistent values.

To demonstrate the relative skill improvement of the coupled model, the BSSs in IWV and IVT are shown in Fig. 9. Here, $\sigma_F^2$ is calculated between the ERA5 and the results obtained in CPL; $\sigma_R^2$ is calculated between the ERA5 and the results obtained in ATM.STA. The BSSs are plotted as functions of lead time. In Fig. 9, the mean RMSE$_{IWV}$
and RMSE_{IVT} are shown; the standard error of the mean are also plotted as error bars\(^1\); the median, the upper/lower quartiles, and the maximum/minimum RMSEs are shown in the inset figures. It can be seen that the mean BSS_{IWV} and BSS_{IVT} are all positive. The coupled model is even better at simulating strong cooling AR events for both IWV (about 12% in week two) and IVT (about 5% in week two), shown in Table 3. However, the skill improvement is much less in weak cooling AR events, where the air-sea heat exchanges are smaller. The skill improvement of IWV is higher than that of IVT, because IVT is far more variable than IWV. This difference will be discussed further in Section 5.

Figure 8. Comparison of RMSE_{IWV} and RMSE_{IVT} obtained in CPL and ATM.STA runs. The simulation results are validated using ERA5. Panel (a) and (b) shows the statistics of RMSE_{IWV} and RMSE_{IVT}, respectively. The inset figures shows the differences between the simulations results and persistent IWV/IVT.

Figure 9. Comparison of BSS_{IWV} and BSS_{IVT} between CPL and ATM.STA runs. The simulation results are validated using ERA5. Panel (a) and (b) shows BSS_{IWV} and BSS_{IVT}, respectively. The markers are the mean BSSs and the error bars are the standard errors of the mean. The inset figures are the box plots of the BSSs that shows the median, the upper/lower quartiles, and the maximum/minimum RMSEs.

\(^1\) Since the errors of persistent IWV and IVT become saturated after 5 days in Fig. 8, we divide the number of simulations by 5 to get sample size \(n\) when calculating the standard error in \(\sigma/\sqrt{n}\). For all AR events, \(n = 18\); for strong/weak cooling ARs, \(n = 6\).
Since the SST can be a source of predictability on sub-seasonal time scales, it is of interest to investigate the improvement of model skill versus the SST variation. To this end, the BSSs are plotted as functions of SST changes in Fig. 10. The SST differences are binned at by 0.06°C intervals, and the BSSs in each bin are averaged. It can be seen that both IWV and IVT skills in CPL increase when SST cooling is stronger in the simulations. The predictions of IWV and IVT of CPL are similar to those of ATM.STA when the SST cooling is less than 0.2 °C. On the other hand, when the SST cooling is stronger than 0.5 °C, the mean BSSs of IWV and IVT are 0.18 and 0.16, respectively. The BSSs are also plotted as functions of the integrated surface heat flux in Fig. 11. It can be seen that the skill of the coupled model increases with increasing $Q_{\text{net}}$. The BSSs are insignificant when the mean surface energy loss is smaller than $0.6 \times 10^8 \text{ J m}^{-2}$, but the mean BSSs of IWV and IVT are 0.18 and 0.15 when the mean surface energy loss is more than $0.6 \times 10^8 \text{ J m}^{-2}$.

Figure 10. The BSSs of IWV and IVT plotted as functions of mean SST difference. Panel (a) and (b) show $\text{BSS}_{\text{IWV}}$ and $\text{BSS}_{\text{IVT}}$, respectively. The markers in the background are the daily-averaged BSS of all simulations (14 days $\times$ 93 simulations).

Figure 11. The BSSs of IWV and IVT plotted as functions of surface heat flux integrated starting from the simulation initial time. Panel (a) and (b) show $\text{BSS}_{\text{IWV}}$ and $\text{BSS}_{\text{IVT}}$, respectively. The markers in the background are the daily-averaged BSS of all simulations (14 days $\times$ 93 simulations).
4.4 BSSs at different atmospheric levels

Although the changing SST influences the ARs in the simulations, its impact is height dependent. In this section we analyze two representative levels: the lower level is from the surface to 850 hPa; the upper level is from 850 hPa to 300 hPa. These levels are selected because each contains about 50% of the water vapor transport.

The comparison of the relative skill at lower and upper levels is shown in Fig. 12. At the lower level BSS$_{IWV}$ and BSS$_{IVT}$ are all positive from day 1 to day 14, suggesting the coupled model better captures the water vapor in this level. In the second week, the improvement in IWV and IVT is about 12% and 4% respectively (Table 4). However, the BSSs in the upper level are almost neutral (medians between -0.02 to +0.02) for both IVT and IWV, indicating the average impact of the SST on forecast skill is insignificant for the upper level. However improved forecast skill is apparent when splitting the strong and weak cooling AR events. As shown in Fig. 13, the BSSs in strong cooling events are higher than those in the weak cooling events, and relative skill improvement of IWV and IVT in week two is 19% and 6% for the lower level and 10% and 3% for the upper layer (Table 4).

Figure 12. The relative skill improvements (BSS$_{IWV}$ and BSS$_{IVT}$) at lower and upper atmosphere levels plotted as a function of lead time. Panel (a) and (b) show BSS$_{IWV}$ and BSS$_{IVT}$, respectively. The markers in the background are the daily-averaged BSS of all simulations.

Figure 13. The relative skill improvements (BSS$_{IWV}$ and BSS$_{IVT}$) in strong and weak cooling AR events at lower and upper atmosphere levels. The skill scores are plotted as functions of lead time. Only the median values are shown.
Table 4. Summary of relative skill improvements at lower and upper atmospheric levels. The average of the median BSSs in Figs. 12 and 13 are shown.

<table>
<thead>
<tr>
<th></th>
<th>week 1</th>
<th>week 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>all ARs (strong/weak cooling)</td>
<td>all ARs (strong/weak cooling)</td>
</tr>
<tr>
<td>IWV, lower level</td>
<td>5.6% (6.0%/0.4%)</td>
<td>12.6% (19.1%/3.7%)</td>
</tr>
<tr>
<td>IVT, lower level</td>
<td>1.1% (1.1%/-0.7%)</td>
<td>4.2% (6.2%/0.4%)</td>
</tr>
<tr>
<td>IWV, upper level</td>
<td>0.3% (5.6%/-0.3%)</td>
<td>1.6% (10.0%/-1.3%)</td>
</tr>
<tr>
<td>IVT, upper level</td>
<td>-0.5% (1.7%/0.2%)</td>
<td>-1.7% (3.4%/-4.2%)</td>
</tr>
</tbody>
</table>

5 Interpreting forecast skill for IWV and IVT

Section 4.3 demonstrated greater skill improvement by the coupled model in forecasting IWV than in forecasting IVT. To examine why $\text{BSS}_{\text{IWV}}$ is greater than $\text{BSS}_{\text{IVT}}$ we examine the components contributing to the total BSS. The BSS is computed by comparing the mean squared error (MSE) $\sigma^2$, which combines information of the “mean” and the “standard deviation”:

$$\sigma^2 = \text{BIAS}^2 + \text{STD}^2,$$

where BIAS and STD are the bias and the standard deviation between model outputs and validation data, respectively. Table 5 summarizes the MSEs, the biases, and the standard deviations as they result from the simulations. It can be seen that the mean IWV and IVT of CPL are smaller by about 1% than those of ATM.STA, suggesting the mean wind velocities are generally consistent in coupled and uncoupled simulations. It can also be seen that the biases of CPL are smaller than those of ATM.STA, but the standard deviations are almost the same. For both IWV and IVT, the STD² terms contribute to more than 90% of the MSEs, while the BIAS² terms contribute to less than 10%. However, the contributions of BIAS² in IWV are more than three times larger than those in IVT, making the difference in mean IWV more important when calculating the BSSs. In conclusion, the $\text{BSS}_{\text{IWV}}$ is higher than $\text{BSS}_{\text{IVT}}$ because the IVT is more variable and the BIAS² terms are less important.

Table 5. Summary of the IWV and IVT obtained in CPL and ATM.STA runs. Decomposition of the MSE $\sigma^2$ of SST, IWV, and IVT in the simulations

<table>
<thead>
<tr>
<th></th>
<th>mean IWV</th>
<th>$\sigma^2$</th>
<th>BIAS</th>
<th>STD</th>
<th>BIAS²/σ²</th>
<th>STD²/σ²</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPL</td>
<td>16.969</td>
<td>3.626</td>
<td>-0.413</td>
<td>1.859</td>
<td>4.7%</td>
<td>95.3%</td>
</tr>
<tr>
<td>ATM.STA</td>
<td>17.152 (+1.1%)</td>
<td>3.796 (+4.4%)</td>
<td>-0.596</td>
<td>1.855</td>
<td>9.4%</td>
<td>90.6%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>mean IVT</th>
<th>$\sigma^2$</th>
<th>BIAS</th>
<th>STD</th>
<th>BIAS²/σ²</th>
<th>STD²/σ²</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPL</td>
<td>208.84</td>
<td>1463.69</td>
<td>-3.76</td>
<td>38.07</td>
<td>1.0%</td>
<td>99.0%</td>
</tr>
<tr>
<td>ATM.STA</td>
<td>211.32 (+1.2%)</td>
<td>1483.75 (+1.4%)</td>
<td>-6.23</td>
<td>38.01</td>
<td>2.6%</td>
<td>97.4%</td>
</tr>
</tbody>
</table>

6 Summary and Conclusion

A series of atmospheric river events are simulated using a regional coupled ocean–atmosphere model (SKRIPS v1.0). The coupled simulation results are compared with those in uncoupled simulations to demonstrate the ocean and atmosphere interactions.
during AR events. We found that the SST cooling in different cases can be significantly
different, hence we highlighted two groups of simulations: (1) strong cooling ARs and
(2) weak cooling ARs. The strong cooling ARs group had the 31 AR events with the most
significant SST cooling and the weak cooling group had the 31 AR events with the weakest cooling. The 31 intermediate cooling events are analyzed as part of the "all AR" statistics, but not in isolation.

Two representative AR events are selected to analyze different thermal interactions
of strong and weak cooling ARs. CASE1 is west–east oriented with a maximum IVT of
about 1250 kg m$^{-1}$ s$^{-1}$; CASE2 is almost south–north oriented with a maximum IVT
of about 900 kg m$^{-1}$ s$^{-1}$. CASE1 exhibits much stronger SST cooling and surface energy loss, suggesting the influence of ARs on the ocean can differ significantly accord-
ing to the events and background ocean state. When performing coupled simulations,
the Brier skill score shows that simulated SST is about 20% more accurate than persist-
ent SST. The THFs resulting from the coupled simulations are about 10% more accurate.
The improvement of the coupled model is even more pronounced in strong cool-
ing AR events.

In addition, we investigated the skill improvement of the coupled model in simu-
lating ARs. Due to the chaotic nature of the atmospheric system, we compared the statistics of BSSs in all simulations instead of comparing the snapshots of each event. In the present case study, both coupled and uncoupled models realistically capture the general characteristics of the atmospheric vertical integrals. For the strong cooling AR events, the coupled model showed improved skill in predicting IWV and IVT by 12% and 5% respectively for lead times of longer than 7 days. The differences between coupled and uncoupled simulations in weak cooling AR events are less significant. The results presented here motivate further studies evaluating the effect of ocean–atmosphere coupling on AR events. Future work will involve exploring the response of SST to the atmosphere and ocean state (e.g., heat fluxes, wind stress, mixed layer deepening), the impact of the annual SST cycle, and the other characteristics of AR (e.g., AR intensity, orientation) on the coupling.

Acknowledgments
We gratefully acknowledge the research funding (grant number: OSR-2-16-RPP-3268.02)
from KAUST (King Abdullah University of Science and Technology). We also appre-
ciate the computational resources on supercomputer Shaheen II and the assistance pro-
vided by KAUST Supercomputer Laboratory. Additional funding from the NSF (OCE1419306)
and the National Oceanic and Atmospheric Administration (MAPP NA17OAR4310106)
is also greatly appreciated. We thank Caroline Papadopoulos for important technical sup-
port when installing software and using the Shaheen II cluster. We also thank the sci-
entists of the Center for Western Weather and Water Extremes at the Scripps Institu-
tion of Oceanography for helping us with simulating ARs and discussing the results.

Data Availability Statement
The source code of the coupled model is maintained on Github https://github.com/iurnus/scripps\_kaust\_model. The AR cases are also available in the same Github repository.

References
Andersen, T. K., Radcliffe, D. E., & Shepherd, J. M. (2013). Quantifying surface en-
ergy fluxes in the vicinity of inland-tracking tropical cyclones. *Journal of Ap-
plied Meteorology and Climatology, 52*(12), 2797–2808.
ing atmospheric river forecasts into subseasonal-to-seasonal time scales. *Geophysical Research Letters*, 44(14), 7528–7536.


National Geophysical Data Center. (2006). 2-minute Gridded Global Relief Data (ETOPO2) v2. National Geophysical Data Center, NOAA.


Neiman, P. J., Ralph, F. M., Moore, B. J., Hughes, M., Mahoney, K. M., Cordeira, J. M., & Dettinger, M. D. (2013). The landfall and inland penetration of a flood-producing atmospheric river in Arizona. Part I: Observed synoptic-scale, orographic, and hydrometeorological characteristics. Journal of Hydrometeorol-


Appendix A AR conditions of the simulations

The daily-averaged IWV and IVT are presented here to complement the results presented in Section 4. We show the difference between CPL and ATM.STA runs, the RMSEs of CPL and ATM.STA runs compared with ERA5, and the difference between the RMSEs. The aim is to demonstrate that the direct comparison of daily-averaged IWV and IVT suffers from the chaotic nature of the atmosphere in our cases.

The difference between the simulation results obtained from CPL and ATM.STA runs (CPL−ATM.STA) are shown in Fig. A1. We select the same representative simulations as Section 3 and show the daily-averaged IWV and IVT as obtained from these simulations. Generally, it can be seen that the IWV is smaller in CPL compared with ATM.STA, especially in CASE1. This is because the CPL captures the SST cooling and the reduction of evaporation, which is a source of the water vapor. The comparison of IVT in Fig. A1 shows that IVT is also smaller in CPL. The difference of IVT is associated with the difference of IWV. It can be also seen that a dipole pattern is observed in CASE2 after 8 and 10 days, which indicates a shift of the AR front between two simulations. The comparison between CPL and ERA5 (CPL−ERA5) is shown in Fig. A2. It can be seen that CPL−ERA5 is three times larger than CPL−ATM.STA. We did not show ATM.STA−ERA5 because it is similar to CPL−ERA5. In CASE1, the coupled model over-estimates the IWV in the warmer sector of AR, but under-estimates the IWV in the cooler sector of AR. However, in CASE2 the difference is not significant at warmer and cooler sectors.

The difference between the RMSEs of CPL and ATM.STA is presented in Fig. A3. Red color indicates the region where CPL has larger error or worse skill; blue color indicates the region where CPL has smaller error or better skill. For CASE1, the coupled
model has better skill with IWV, but the IVT skill improvement is significantly smaller
than that of IWV. For CASE2, the mean SST difference is very small and difference of
IWV/IVT are chaotic in Fig. A3. Thus, because of the chaotic nature of the atmosphere,
showing the difference in Fig. A4 cannot be helpful to quantitatively evaluate the skill
of the coupled model. Instead, we examined the statistics of the skill of coupled and un-
coupled models and detailed them in Section 4.

**Appendix B Comparison between early and late January cases**

In Fig. 5, we used the SST cooling to group the ARs in the simulations. It can be
seen that most strong/weak cooling ARs occurred in the simulations are initialized on
early/late January. Hence, we compared the cases initialized on the first 10 days and last
10 days (about 1/3 of all simulations).

The BSSs are plotted as functions of lead time in Fig. B1. Here, $\sigma_F^2$ is calculated
between the ERA5 and the results of CPL; $\sigma_R^2$ is calculated between the ERA5 and the
results of ATM.STA. The median, the upper/lower quartiles, and the maximum/minimum
RMSEs are plotted in the figure. It can be seen that the median BSS$_{IWV}$ in early Jan-
uary case are slightly better than late January cases, especially in the second week of
the simulations (early January cases: 9.7%; late January cases: 3.6%). On the other hand,
the median BSS$_{IVT}$ in early January cases is still better than late January cases, but the
improvement is much smaller (early January cases: 2.0%; late January cases: 0.5%). Com-
pared with Fig. 9, using the SST cooling can better show the differences in the AR events
in this case study.
Figure A1. The difference between IWV/IVT obtained in CPL and ATM.STA runs. The black contours highlight the AR region where IVT > 250 kg m$^{-1}$ s$^{-1}$. The results obtained in CASE1 and CASE2 in Section 3 are shown.
Figure A2. The difference between IWV/IVT obtained in CPL and ERA5 data. The black contours highlight the AR region where IVT > 250 kg m$^{-1}$ s$^{-1}$. The results obtained in CASE1 and CASE2 in Section 3 are shown.
Figure A3. The difference between IWV/IVT RMSEs in CPL and ATM.STA runs. Red color shows that CPL has larger error or worse skill; blue color shows that CPL has smaller error or better skill. The black contours highlight the AR region where IVT $> 250$ kg m$^{-1}$ s$^{-1}$. The results obtained in CASE1 and CASE2 in Section 3 are shown.

Figure B1. Comparison of BSS$_{IWV}$ and BSS$_{IVT}$ between CPL and ATM.STA runs. The simulation results are validated using ERA5. Panel (a) and (b) shows BSS$_{IWV}$ and BSS$_{IVT}$, respectively. The box plot shows the median, the upper/lower quartiles, and the maximum/minimum BSSs. The early January cases are initialized on the first 10 days; the late January cases are initialized on the last 10 days (about 1/3 of all simulations).