

Global Assessment of Atmospheric River Prediction Skill[✉]

MICHAEL J. DEFLORIO

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California

DUANE E. WALISER AND BIN GUAN

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, and Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, Los Angeles, California

DAVID A. LAVERS

European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom

F. MARTIN RALPH

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

FRÉDÉRIC VITART

European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom

(Manuscript received 14 July 2017, in final form 5 December 2017)

ABSTRACT

Atmospheric rivers (ARs) are global phenomena that transport water vapor horizontally and are associated with hydrological extremes. In this study, the Atmospheric River Skill (ATRISK) algorithm is introduced, which quantifies AR prediction skill in an object-based framework using Subseasonal to Seasonal (S2S) Project global hindcast data from the European Centre for Medium-Range Weather Forecasts (ECMWF) model. The dependence of AR forecast skill is globally characterized by season, lead time, and distance between observed and forecasted ARs. Mean values of daily AR prediction skill saturate around 7–10 days, and seasonal variations are highest over the Northern Hemispheric ocean basins, where AR prediction skill increases by 15%–20% at a 7-day lead during boreal winter relative to boreal summer. AR hit and false alarm rates are explicitly considered using relative operating characteristic (ROC) curves. This analysis reveals that AR forecast utility increases at 10-day lead over the North Pacific/western U.S. region during positive El Niño–Southern Oscillation (ENSO) conditions and at 7- and 10-day leads over the North Atlantic/U.K. region during negative Arctic Oscillation (AO) conditions and decreases at a 10-day lead over the North Pacific/western U.S. region during negative Pacific–North America (PNA) teleconnection conditions. Exceptionally large increases in AR forecast utility are found 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. These results represent the first global assessment of AR prediction skill and highlight climate variability conditions that modulate regional AR forecast skill.

1. Introduction

Atmospheric rivers (ARs) are narrow plumes of strong horizontal water vapor transport that are typically found in the midlatitudes ahead of the cold front of an extratropical cyclone (Zhu and Newell 1998; Ralph et al. 2004; Neiman et al. 2008; Cordeira et al. 2013; American Meteorological Society 2017). They can intensify downstream precipitation

[✉] Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JHM-D-17-0135.s1>.

Corresponding author: Michael J. DeFlorio, michael.deflorio@jpl.nasa.gov

and influence flooding, snowpack, and water availability (e.g., [Ralph et al. 2004, 2005, 2006](#); [Neiman et al. 2008](#); [Dettinger et al. 2011](#); [Lavers and Villarini 2015](#)). The abundance or deficit of ARs strongly influences the availability or lack of freshwater in a given water year ([Guan et al. 2010](#); [Dettinger et al. 2011](#); [Dettinger 2013](#)). In addition, heavy precipitation events, flooding conditions, and hazardous winds are strongly correlated to the occurrence of ARs in many regions around the globe ([Ralph et al. 2006](#); [Ralph and Dettinger 2011](#); [Ralph and Dettinger 2012](#); [Sodemann and Stohl 2013](#); [Lavers and Villarini 2015](#); [Nayak et al. 2016](#); [Ralph et al. 2016](#); [Waliser and Guan 2017](#)).

ARs can have both beneficial (e.g., replenishing water reservoirs) and detrimental (e.g., causing damaging floods and landslides) impacts on regional economics and public safety. Consequently, there is significant incentive at the intersection of the research, resource management, hazard preparedness and response, and policymaking communities to understand the linkage between AR landfall and subsequent impacts. This includes the need to quantify the skill of predicting AR occurrence, especially in terms of landfall, and to identify large-scale climate variability conditions during which forecasts of ARs are more or less skillful.

Given the importance of forecasting AR activity, recent studies have assessed prediction skill in ensemble forecasts of ARs. [Wick et al. \(2013\)](#) assessed AR landfall prediction skill of five operational ensemble forecast systems [National Centers for Environmental Prediction (NCEP), European Centre for Medium-Range Weather Forecasts (ECMWF), the Met Office (UKMO), the Canadian Meteorological Centre (CMC), and the Japan Meteorological Agency (JMA)] over the North Pacific/western U.S. region from 2008 to 2011 and found that AR occurrence is relatively well forecast out to a 10-day lead time (probability of detection > 84% at all lead times, and false alarms < 12% at all lead times), but that AR landfall position is subject to significant errors of over 800 km at a 10-day lead time. [Nayak et al. \(2014\)](#) used five operational control forecasts [four of the five operational models that were used in [Wick et al. \(2013\)](#)] from 2007 to 2013 and concluded that the models are not skillful in forecasting AR activity beyond a 7-day lead time over the central United States due to errors in both AR occurrence and AR landfall location. Note that for each of these studies, only the control forecast was used. In an ensemble forecast framework, which is used in this paper, particular ensemble members may have greater ability than demonstrated by the control forecast.

Considerable progress has also been made recently in understanding limits of prediction skill of AR-related quantities such as integrated vapor transport (IVT) and

precipitation (as opposed to explicitly detecting or defining AR objects). [Sukovich et al. \(2014\)](#) evaluated extreme quantitative precipitation forecasts (QPFs) and found higher skill of seasonal extreme QPFs during boreal winter over the contiguous United States. [Moore et al. \(2015\)](#) verified precipitation forecasts from the NCEP Global Ensemble Forecast System from 1985 to 2014 and demonstrated that strong-IVT (likely AR) extreme precipitation events over North America are associated with higher forecast skill than weak-IVT (likely not AR) events out to a 6-day forecast lead time. Other studies have focused on predictability of these AR-related meteorological parameters. [Lavers et al. \(2014\)](#) and [Lavers et al. \(2016a\)](#) distinguished between forecasts of IVT and precipitation and found higher predictability in IVT out to a 9-day lead time over both Europe and the western United States.

The study presented here extends these studies of AR (and AR related) prediction skill and predictability by focusing on a global (rather than regional) domain and by examining the ability of numerical weather prediction models to predict ARs out to a 14-day lead time, with an added focus on forecasts that are initialized during certain conditions of large-scale climate variability.

2. Data and methods

a. ECMWF ensemble hindcasts: A subset of the S2S database

To assess the prediction skill of ARs, global hindcast data from the newly created Subseasonal to Seasonal (S2S) Project database (Table 1 of [Vitart et al. 2017](#)) are used. The analysis is limited to one of the S2S models, the ECMWF model, over the period from 1996 to 2013. The model data are retrieved at a $1.5^\circ \times 1.5^\circ$ grid, which is the default archive grid for ECMWF hindcasts in the S2S database. Hindcasts are made twice a week, and there are 10 perturbed forecast members available. Ocean coupling is included in the ECMWF hindcasts, but sea ice coupling is not. Hindcasts are made “on the fly,” which means that ECMWF updates the version of their model at various times during a given year. The ECMWF hindcasts used in the present study were produced with model versions between May 2015 and May 2016 and therefore used mostly Integrated Forecast System (IFS) model versions 41r1 (41r2 after March 2016).

The ECMWF interim reanalysis (ERA-I) dataset ([Dee et al. 2011a](#)) is used as a proxy observation. The reanalysis data are available on a $0.75^\circ \times 0.75^\circ$ grid and are bilinearly interpolated to match the coarser hindcast grid.

Ten perturbed ECMWF hindcast ensemble members are analyzed (the control hindcast is not included because it was not available for download at the time of data retrieval), as is ERA-I data from 1996 to 2013 at 0000 UTC using specific humidity q and zonal u and meridional v winds at 300-, 500-, 700-, 850-, and 1000-hPa vertical levels. The zonal and meridional components of the IVT [defined as the vertical integral of the product of specific humidity and zonal/meridional wind; see [Zhu and Newell 1998](#), their Eqs. (7a) and (7b)] are calculated first. Subsequently, the [Guan and Waliser \(2015, hereafter GW2015\)](#) AR detection algorithm, which identifies ARs objectively as contiguous regions (objects) where IVT at each grid cell exceeds the 85th percentile specific to the region and season and with the length of the object > 2000 km and length-to-width ratio > 2 , is applied on these IVT fields. This algorithm detects ARs globally in the manner described above as a function of ensemble member and lead time (1–14 days at a 1-day step). Focus is placed on the November–March (NDJFM) and May–September (MJJAS) seasons.

b. Object-based detection for forecast verification

Object-based forecast verification involves evaluating a model’s ability to predict sporadic, episodic atmospheric phenomena that are heterogeneous in space and time ([Ebert and McBride 2000](#); [Davis et al. 2006](#)). Recent studies have explained how traditional methods of verifying numerical weather forecasts (e.g., root-mean-square error) of relatively smooth fields (e.g., upper-tropospheric winds) may not be suitable for

verification of episodic, localized features (e.g., atmospheric rivers) that have considerable spatiotemporal variability ([Ebert 2008](#); [Mittermaier et al. 2015](#)).

The [GW2015](#) algorithm employs an object-based approach and detects ARs based on synoptic features that represent a rough consensus of AR characteristics (IVT, geometry, etc.). The algorithm introduced in this study takes these AR objects as input, identifies the latitude and longitude coordinates of their IVT-weighted centroid, and compares the distance between any grid cell and all AR object centroids to a user-defined distance threshold. This process is described in more detail in [sections 2c](#) and [3b\(1\)](#).

c. ATRISK algorithm description

The Atmospheric River Skill (ATRISK) algorithm helps us achieve two primary goals in this paper. The first goal is to compute the percentage of ECMWF ensemble members that can skillfully predict an observed AR (i.e., AR hit percentage) as a function of season, location, lead time, and an acceptable (great circle) distance threshold between the centroid of the observed and forecasted AR (1000, 500, and 250 km tested in this study). The great circle distance D between the centroid of a given grid cell and the centroid of an AR object is calculated as

$$D = R \times 2 \times \arctan\left(\frac{\sqrt{a}, \sqrt{1-a}}{1}\right),$$

where R = radius of Earth in kilometers and

$$a = \sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos(\varphi_1) \times \cos(\varphi_2) \times \sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)},$$

for latitude φ and longitude λ .

The second goal is to compute relative operating characteristic (ROC) curves ([Hanley and McNeil 1982](#)) for ARs, which account for hits (observed AR and forecasted AR), misses (observed AR, but no forecasted AR), false alarms (no observed AR, but forecasted AR), and correct rejections (no observed AR and no forecasted AR). ROC curves have been previously used in applied weather and climate studies that seek to discern the utility of operational forecasts of atmospheric and/or oceanic fields (e.g., [Mason and Graham 2002](#)) and have recently been used to evaluate how well IVT can discriminate extreme precipitation events ([Lavers et al. 2016b](#)). Hit rates are calculated as [number of hits/(number of hits + number of misses)] and are equivalent to probability of detection (POD). False alarm rates are

calculated as [number of false alarms/(number of false alarms + number of correct rejections)] over a given time period of interest. This quantity is distinct from false alarm ratio (FAR), which is often implemented as an operational forecast verification tool and is calculated as [number of false alarms/(number of false alarms + number of hits)] (see [Barnes et al. 2009](#)). Area under ROC curve values (ROC scores), which are introduced in [section 3b\(2\)](#), are proportional to the critical success index (CSI), which is defined as [number of hits/(number of hits + number of misses + number of false alarms)]. The ROC sensitivity to season, location, lead time, and states of various modes of climate variability is further examined.

The first goal diagnoses the ability of ECMWF to forecast an observed AR 1–14 days in advance, while the

second objective assesses AR forecast utility in ECMWF by explicitly considering all contingency table values (hits, misses, false alarms, correct rejections). The ROC is introduced to achieve the second objective as a way to expand the scope of this study and more thoroughly assess model performance. Note that if one only uses the hit percentage from the first part of this study and does not include ROC analysis (or a similar metric), a model that forecasts ARs everywhere at all times will appear to be perfect. It is only by considering false alarms, misses, and correct rejections (in addition to hits) that this can be adequately accounted for. Both approaches together allow us to quantify sensitivities of AR prediction skill to region, lead time, season, and background phases of climate variability at hindcast initiation.

ATRISK follows the steps below to achieve the first goal. Additional steps are given in [section 3b\(1\)](#) that allow for accomplishing the second goal.

- 1) Retrieve observed (ERA-I) and modeled (ECMWF) AR centroid longitude and latitude values. For ECMWF, this is done for all lead times (1–14 days) and ensemble members 1–10.
- 2) For each day, assign the value of 1 to all grid cells whose distance is within D kilometers of the centroid of an observed AR object present at that given time step. Other grid cells are assigned the value of 0.
- 3) For each modeled hindcast (at a given lead time and for a given ensemble member), assign the value of 1 to all grid cells across the globe whose centroid is within D kilometers of the nearest AR object at that given time step. Shaded grid cells are saved with values of 1.
- 4) Using the 10 ensemble member shaded fields from step 3, compute an ensemble-average field for each hindcast and lead time at each grid cell.

As an example, consider all 1996–2013 4 January hindcasts in the ECMWF dataset, and consider the 7-day lead time. In step 3, global maps are created with shaded values of 1 that correspond to grid cells that are within D kilometers of the nearest AR object centroid. This is done for each ensemble member, yielding 10 global maps with shaded orbs (which have values of 1), similar to the orbs shown in [Fig. 1b](#). Step 4 then creates an ensemble-averaged field from these 10 maps, with “ensemble-average orbs” ranging between 0 and 1.

A visual summary of the [GW2015](#) and ATRISK algorithms is shown in [Fig. 1](#). AR objects at 0000 UTC 2 March 1996, as detected by the algorithm introduced in [GW2015](#), are shown in [Fig. 1a](#) and are shaded in blue. ATRISK objects within 1000 km of the AR objects are shown in [Fig. 1b](#) and are also shaded in

blue. The method of determining if a predicted AR is a “hit” or “miss” relative to an observed AR is shown in [Fig. 2](#). In this example, the prediction of AR₁ is considered skillful (a hit) since its centroid falls within the distance threshold of the observed AR, while the prediction of AR₂ is not considered skillful (a miss) since its centroid falls outside the distance threshold of the observed AR.

To achieve the first goal, all observed AR objects are identified, and their centroid longitude and latitude are compared to each ensemble member AR object (as a function of lead time). These results are described in [section 3a](#) and with [Figs. 2–4](#). Achieving the second goal is more complex, and the steps taken to create [Figs. 5–8](#) are described in [section 3b](#).

3. Results and discussion

a. Global climatology of AR hits and sensitivity to lead time, season, and distance threshold

The average 1996–2013 NDJFM percentage of ECMWF ensemble member AR hits at a 7-day lead time is shown globally in [Fig. 3](#) using a 1000- and 500-km distance threshold between observed and forecasted ARs. The blue polygons drawn in [Fig. 3a](#) denote the area average domains used throughout this study. This is the first approach used to assess global AR prediction skill (i.e., focusing only on hits, namely, the fraction of ensemble members that forecast an AR within the distance threshold).

For a given distance threshold, hit percentages are generally higher over ocean basins than over land regions, and hit percentage decreases by 25%–35% in most locations for the 500-km threshold relative to the 1000-km threshold. Because ARs are sporadic phenomena, the global hit percentage map for a single hindcast would contain some grid cells with missing hit percentage values, since the presence of observed and forecasted ARs is less likely in a smaller sample of data. Using all hindcasts from a single season (as in [Figs. 3a,b](#)) has the effect of smoothing the statistics in space, since more observed and forecasted AR pairs are captured in a larger sample. There are still areas in [Figs. 3a and 3b](#) where NA (missing) values are present due to a lack of observed and forecasted ARs. This might occur, for example, in regions of low climatological AR frequency (e.g., eastern equatorial Pacific; see [GW2015](#), their [Fig. 7a](#)), or because the twice-a-week hindcast frequency omits some observed ARs. During boreal summer, hit percentages are highest in the Southern Hemisphere and decrease by 20%–30% in most North Hemispheric regions for both 1000- and 500-km distance thresholds (not shown).

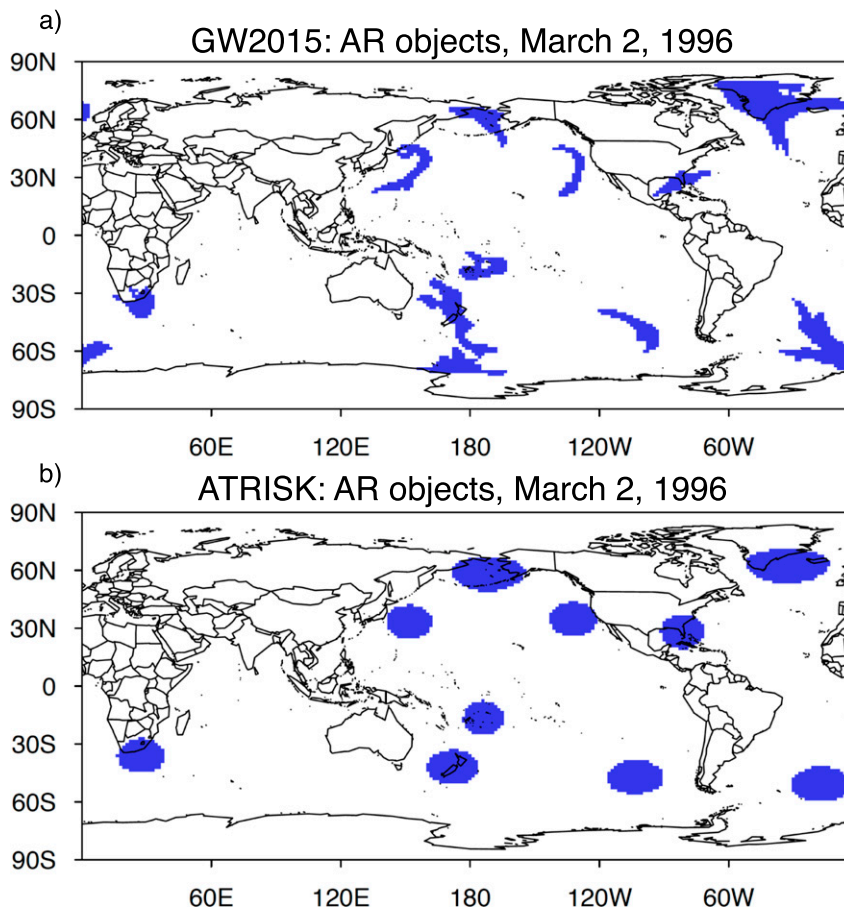


FIG. 1. GW2015 and ATRISK algorithm schematics. (a) AR objects at 0000 UTC 2 Mar 1996 as detected by the algorithm introduced in GW2015. (b) Shaded grid cells from ATRISK within 1000 km of the AR objects shown in (a).

Area-averaged AR hit percentages are shown as a function of lead time, season (colors), and distance threshold (1000 km, solid lines; 500 km, dashed lines; 250 km, dotted lines) in Figs. 4a–d for the North Pacific/western United States (35° – 45° N, 150° – 125° W), North Atlantic/United Kingdom (45° – 60° N, 30° W– 0°), South Pacific/Australia (30° – 40° S, 145° – 170° E), and South Pacific/Chile (30° – 40° S, 110° – 70° W) regions during NDJFM and MJJAS. These regions are chosen because they coincide with midlatitude storm tracks, are upstream of heavily populated regions, and include coastlines where AR landfalls are frequent (GW2015, their Fig. 8a). GW2015 identified NDJFM as a climatologically active AR season for the Northern Hemispheric regions and MJJAS as a climatologically active AR season for the Southern Hemispheric regions (see their Fig. 9b). Mundhenk et al. (2016) and Eiras-Barca et al. (2016) identified boreal winter (DJF) as a climatologically active season for ARs relative to boreal summer over the North Pacific/western U.S. and North Atlantic/U.K.

regions, respectively, used in this study. In addition, for all regions, AR landfalls can have major socioeconomic consequences in association with extreme precipitation and winds (e.g., Waliser and Guan 2017). NDJFM and MJJAS are used instead of DJF and JJA in order to increase the sample size of AR hits for a given season, thus increasing the signal-to-noise ratio. In addition, for the west coast of North America and northwestern Europe, ARs are quite typical during this extended winter season.

Horizontal solid lines represent the observed ERA-I 1996–2013 reference climatologies of AR frequency for the given region, season, and distance threshold (DT). These lines represent the long-term chance of a grid cell (or in this case, a regional average of grid cells) being within DT kilometers of an observed AR (with DT = 1000, 500, or 250 km), or put another way, the skill based on using climatology as the forecast model. Therefore, the hit percentages shown in Fig. 4 can be thought of as skillful (or useful) relative to the observed climatological

Method of determining if a predicted atmospheric river (AR) is a “hit” or a “miss” relative to an observed AR

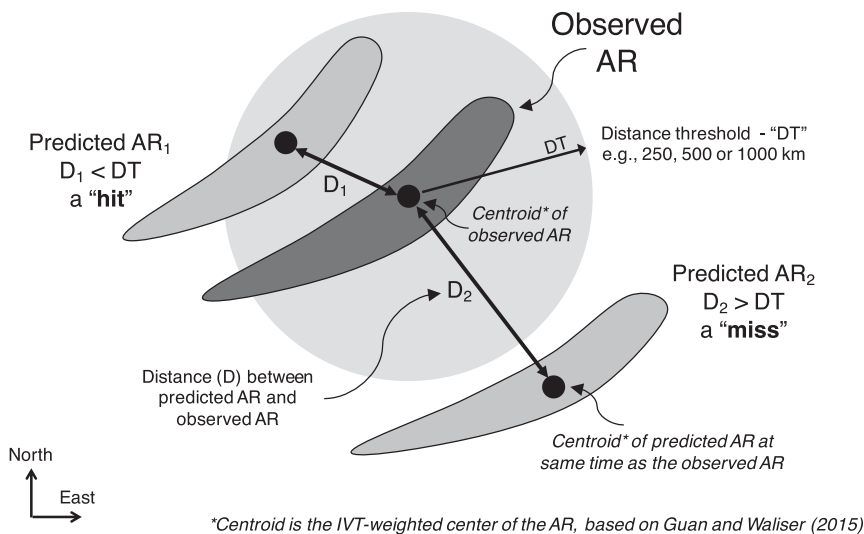


FIG. 2. Method of determining if a predicted AR is a hit or a miss relative to an observed AR. Predicted and observed ARs are shown as shaded light and dark shaded ovals, respectively. Their IVT-weighted centroids are shown as black dots, and the distances D_1 and D_2 between each predicted AR and the observed AR are shown as black arrows. DT , which indicates the acceptable horizontal distance between an observed and predicted AR for a prediction to be considered skillful, is shown as a black arrow. In this example, the prediction of AR_1 is considered skillful (a hit) since its centroid falls within the distance threshold of the observed AR, while the prediction of AR_2 is not considered skillful (a miss) since its centroid falls outside the distance threshold of the observed AR.

AR frequency (i.e., the long-term chance of a grid cell being within DT kilometers of an observed AR) if their values at a given lead time are above the horizontal line (the observed AR frequency climatology). A calculation of a similar reference climatology is made for ECMWF at all lead times and shows a very close agreement with the observed reference climatology in all regions, indicating that there are no significant model biases in AR occurrence in these regions (even out to 28 days, not shown).

AR hit percentage is most variable over the North Pacific/western U.S. region, while both Southern Hemispheric regions show little seasonal variability. In all regions, hit percentages asymptote to a “noise floor” around 10-, 8-, and 6-day lead times for the 1000-, 500-, and 250-km thresholds, respectively. The noise floor for each distance threshold and seasonal pairing is very close to the observed reference climatology, which is consistent with the interpretation that the values of the asymptotes represent the odds of a random AR falling within the associated distance threshold of an observed AR.

The lead time at which AR hit percentage equals 50% as a function of distance threshold is shown in Fig. 5a for

the North Pacific/western U.S. and South Pacific/Chile regions during NDJFM (black) and MJJAS (blue). Such information is an important consideration if the landfall location and associated landfall distance error is relevant, for example, for reservoir management or hazard response. For example, if a decision-maker, impacted by North Pacific/western U.S. ARs, is able to make a decision based on at least 50% hit percentage but requires distance errors to be no larger than 250 km, then the forecast may not generally be reliable in this sense until 2–3 days before the landfall occurs. On the other hand, if the decision can be made from a consideration of 1000-km landfall error, then the forecast may be useful at a lead time of 7 days.

The AR hit percentage at a 7-day lead time as a function of distance threshold for the same regions and seasons is shown in Fig. 5b, along with reference climatology values. This information is relevant for medium-range weather forecasting or reservoir management. For example, if a water reservoir manager in California can improve his or her management by knowing that the climatological chance of a forecasted AR occurring within 500 km of an observed AR at their

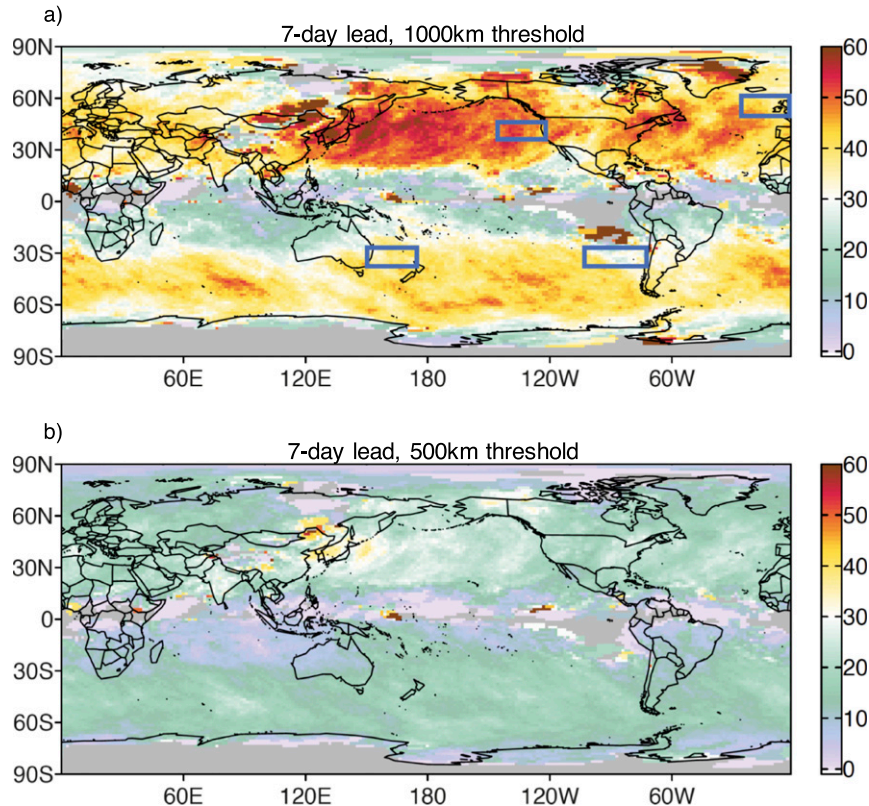


FIG. 3. The 1996–2013 ECMWF NDJFM average percentage of ensemble members that forecast AR hits at a 7-day lead time using (a) 1000- and (b) 500-km distance thresholds. The blue polygons in (a) denote the area average domains used throughout this study. Gray shading indicates regions where no ARs are detected.

reservoir based exceeds 20%, only boreal winter (NDJFM) forecasts would be reliable compared to boreal summer (MJJAS) forecasts.

Results from previous studies generally corroborate the findings presented here, though no studies to date have estimated prediction skill on a global scale at lead times extending out to 14 days of the AR objects themselves in the manner of Figs. 2–4. Wick et al. (2013) computed the correlation coefficient between three boreal winter seasons (2008–11) of observed integrated water vapor (IWV) fields [from Special Sensor Microwave Imager (SSM/I) satellite data] and five numerical weather prediction model control forecasts of AR-related IWV over the North Pacific at 1-, 3-, 5-, 7-, and 10-day lead times. In most models, the correlation coefficient reduces by approximately 30% between 7- and 10-day leads. Though not directly comparable due to differences in methodology, record length, and domain size, the results in Fig. 4a are generally consistent, as the North Pacific/western U.S. hit percentages during NDJFM decrease by approximately 30% of their initial values at around 7- to 8-day lead times. Lavers et al. (2016a) estimated predictability of IVT over the eastern

North Pacific/western U.S. region from the NCEP Global Ensemble Forecast System (GEFS) hindcast data from 1984 to 2015 and found indications of predictability out to a 9-day lead time (and out to a 7-day lead time for precipitation). Note that in this study, prediction skill is quantified rather than estimating predictability, and thus skill values are expected to be less than predictability values.

An additional calculation of AR prediction skill is now made using ROC curves, which explicitly account for false alarms in addition to hits. To illustrate the need for this consideration in simple terms, consider a forecast model that forecasts ARs everywhere all the time. Its hit percentage would be perfect all the time, yet it would actually be a poor forecast model as it would be overly fraught with false alarms.

b. Assessing AR prediction skill using ROC curves

The ROC methodology mentioned in section 2c is now employed as a second approach to assessing global AR prediction skill. To account for both hits and false alarms, ROC curves and areas are calculated for ARs using the ERA-I and ECMWF hindcast ensemble datasets.

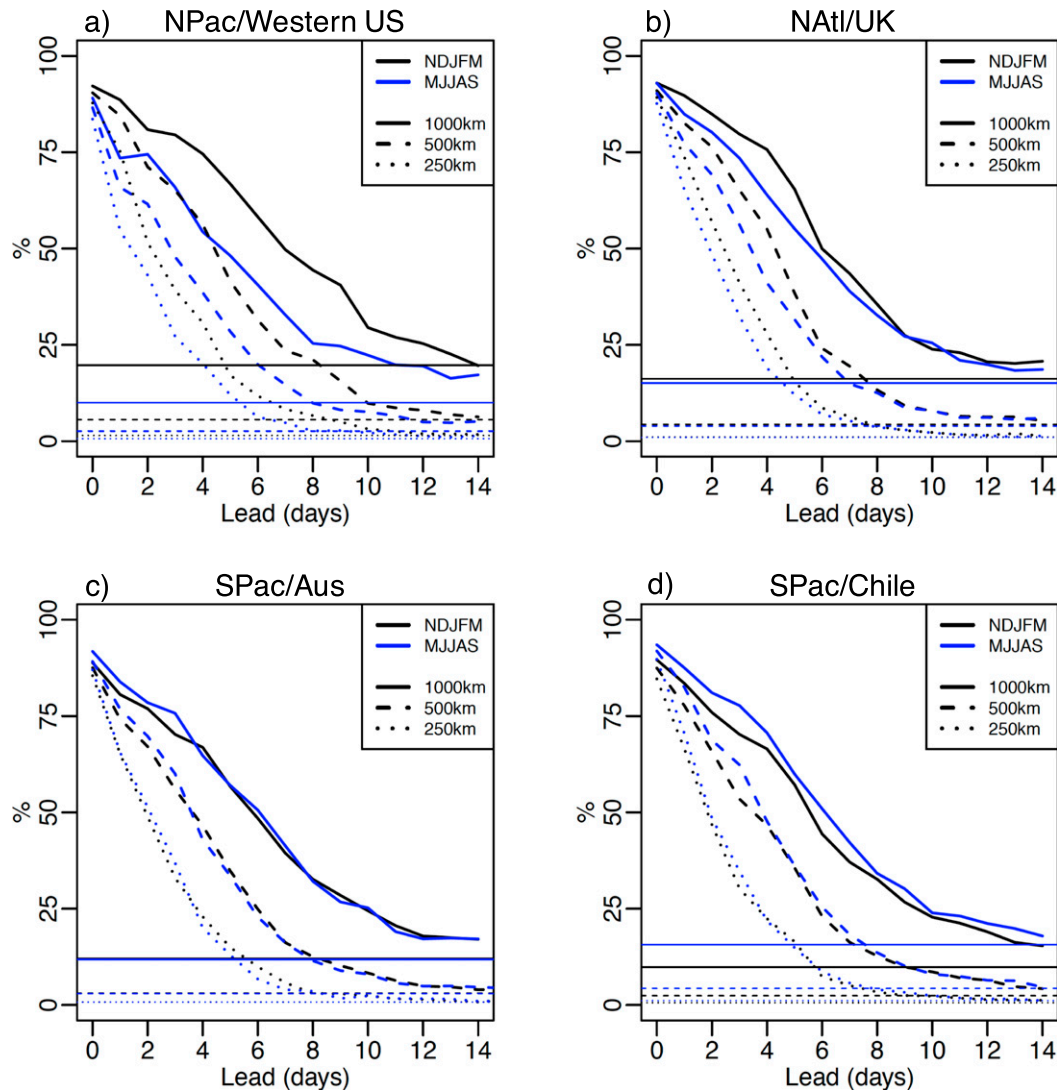


FIG. 4. The 1996–2013 ECMWF average percentage of ensemble member AR hits vs lead time over the (a) North Pacific/western U.S., (b) North Atlantic/U.K., (c) South Pacific/Australia, and (d) South Pacific/Chile regions during NDJFM (black) and MJJAS (blue) and using distance thresholds of 1000 (solid), 500 (dashed), and 250 km (dotted). Horizontal solid lines represent ERA-I reference climatology of AR frequency for the given region, season, and distance threshold.

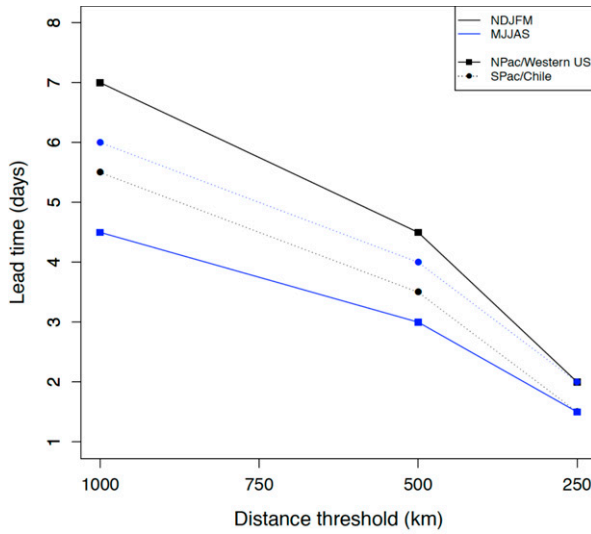
1) STEPS CONTAINED IN ATRISK TO CALCULATE ROC CURVES AND SCORES IN FIGS. 5–8

To calculate ROC curves for a given great circle distance threshold of D kilometers, the following steps are used:

- 1) Follow steps 1–4 outlined in section 2c.
- 2) Use a range of trigger thresholds above which decision-makers might issue warnings or enact contingency plans for an impending impactful event. Here, the trigger thresholds are 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0. They represent the fraction of ECMWF ensemble mem-

- bers that agree on a forecast of an AR. Then, compare each trigger threshold to the fields created in steps 1–4 from section 2c. If the average forecast probability shaded field is greater than the trigger threshold value, store a value of “1.” If not, store a value of “0.”
- 3) Compute contingency table fields as follows, as a function of time and trigger threshold, at each grid cell:
 - (i) Hit field: store value of 1 if the observed field = 1 (observed AR present within D kilometers), and the model field = 1 (modeled AR present within D kilometers at the given threshold). Otherwise, store a 0.

a) Lead time at which AR hit % = 50



b) AR hit % at 7-day lead

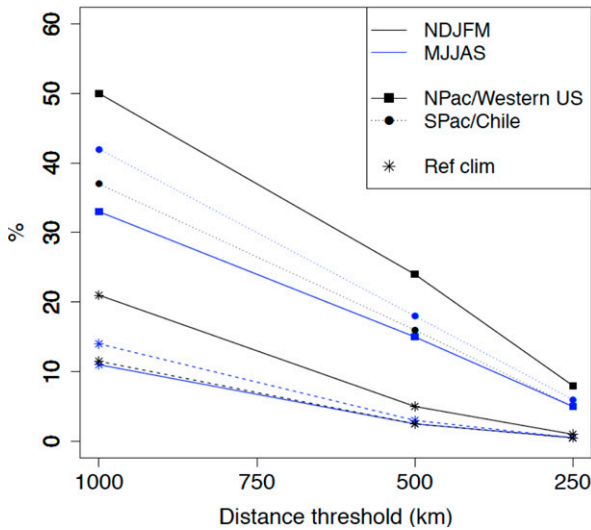


FIG. 5. (a) Lead time at which 1996–2013 ECMWF average AR hit percentage = 50% vs distance threshold and (b) AR hit percentage at a 7-day lead time vs distance threshold during NDJFM (black) and MJJAS (blue). Reference climatology values in (b) at each distance threshold are denoted with an asterisk.

- (ii) Miss field: store value of 1 if the observed field = 1 (observed AR present within D kilometers), and the model field = 0 (no modeled AR present within D kilometers at the given threshold). Otherwise, store a 0.
- (iii) False alarm: store value of 1 if the observed field = 0 (no observed AR present within D kilometers), and the model field = 1 (modeled AR present within D kilometers at the given threshold). Otherwise, store a 0.

- (iv) Correct rejection: store value of 1 if the observed field = 0 (no observed AR present within D kilometers), and the model field = 0 (no modeled AR present within D kilometers at the given threshold). Otherwise, store a 0.

4) Compute hit rates and false alarm rates and create ROC curves and global maps of the area under the ROC curves.

2) LONG-TERM MEAN ROC CURVES

The 1996–2013 average ROC curve for NDJFM over the North Pacific/western U.S. region at various lead times is shown in Fig. 6a. Triangles denote hit rate and false alarm rate values at a given trigger threshold. The threshold labels below the x axis correspond to the triangles on the plot. At short (3 and 5 days) lead times, the ROC curves are closer to the top-left corner of the graph, indicating that forecasts of ARs at those lead times are useful (i.e., high hit rates and small false alarm rates). At longer (14 days) lead times, the ROC curves are closer to the solid black line, which means that a forecast of an AR at those lead times is only marginally more useful than a random guess (i.e., equal probability of hit and false alarm). This representation of AR prediction skill improves the understanding of the extent to which an AR forecast is more skillful than a random guess and distinguishes actual skill from stochastic chance. We will introduce a more formal statistical significance test in section 3b(5) to quantify the extent to which ROC curves composited on various modes of climate are significantly different from long-term average ROC curves. This comparison more explicitly distinguishes a forecast ROC curve from a random guess. The 1996–2013 average area under the ROC curve calculated at each grid cell globally at a 7-day lead time is shown in Figs. 6b and 6c using a 1000- and 500-km distance threshold. Area under ROC curve values (or “ROC scores”) for the 1000-km threshold uniformly range between 0.7 and 0.9 at most places, while values at the 500-km threshold are smaller by about 0.1–0.2 and noisier.

Area-averaged ROC scores are shown in Fig. 7 as a function of lead time, season (colors), and distance threshold (as in Fig. 4). The bold horizontal line at 0.5 represents the baseline for a useful forecast, since ROC scores that equal 0.5 indicate an equal probability of hits and false alarms. ROC scores, which explicitly account for both AR hits and AR false alarms, are higher during boreal winter in the North Pacific/western U.S. and North Atlantic/U.K. regions, especially at 1000- and 500-km distance thresholds.

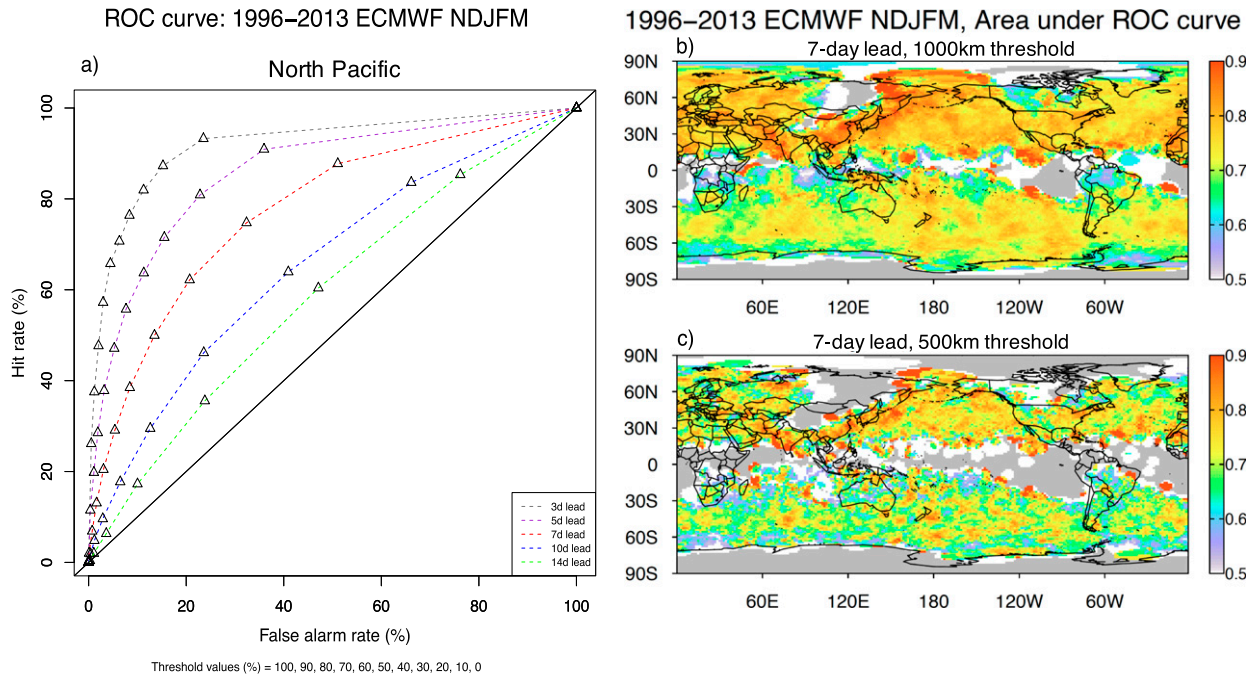


FIG. 6. (a) ROC curve of North Pacific ARs for 1996–2013 NDJFM average using 1000-km distance threshold at 3-, 5-, 7-, 10-, and 14-day lead times. Triangles denote hit rate and false alarm rate values at a given trigger threshold. Area under 1996–2013 NDJFM ROC curve (i.e., ROC score) at a 7-day lead time for a (b) 1000- and (c) 500-km threshold.

3) MOTIVATION FOR ASSESSING AR PREDICTION SKILL CONDITIONED ON VARIOUS STATES OF THE CLIMATE

A conditional verification is now undertaken using hindcasts that are initialized during positive and negative phases of several modes of climate variability. The purpose of this analysis is to test whether or not additional forecast skill and utility can be gained during certain phases of climate mode variability. The relationship between AR forecast utility and climate variations is examined here using the El Niño–Southern Oscillation (ENSO), the Arctic Oscillation (AO), and the Pacific–North America (PNA) teleconnection pattern. These climate variations are a focus because AR frequency is sensitive to the phase of each of these modes over particular regions around the globe (GW2015). In addition, ENSO phase can be a useful predictor not only for seasonal forecasts of precipitation (Pezzi and Cavalcanti 2001; Coelho et al. 2002; Taschetto et al. 2016), but for subseasonal and synoptic time-scale forecasts as well (e.g., Li and Robertson 2015). Black et al. (2017) also examined subseasonal prediction of the modes of climate variability themselves [namely, AO, PNA, and the North Atlantic Oscillation (NAO)]. Their results suggest that the AO and the PNA, along with the NAO, can be predicted with

statistically significant skill by a statistical model based on partial least squares regression up to lead times of 4–5 weeks relative to climatology and persistence.

Emphasis is also placed on phase-locked ENSO–PNA conditions because there is some evidence for useful prediction skill of ENSO on monthly-to-seasonal time scales (Dayan et al. 2014; Barnston et al. 2012), and for a strong relationship between PNA circulation anomalies and El Niño (Abid et al. 2015). Muñoz et al. (2015) also suggest that skillful prediction of extreme precipitation can be increased during particular phase-locking configurations of ENSO, the Atlantic meridional mode (AMM), and other potential predictor modes of variability. In addition, it will be shown in Fig. 9 that AR frequency anomalies are magnified during these phase-locked conditions relative to the individual mode responses.

The conditional ROC curves shown in this section might be a useful reference, for example, in issuing a forecast over a given region where AR activity projects strongly onto precipitation. If a forecaster knows that AR frequency anomalies are positive or negative during a certain mode phase, it may be useful to know the relative utility of AR forecasts during those phases (as quantified by the conditional ROC curves) as a parameter to consider when making a forecast.

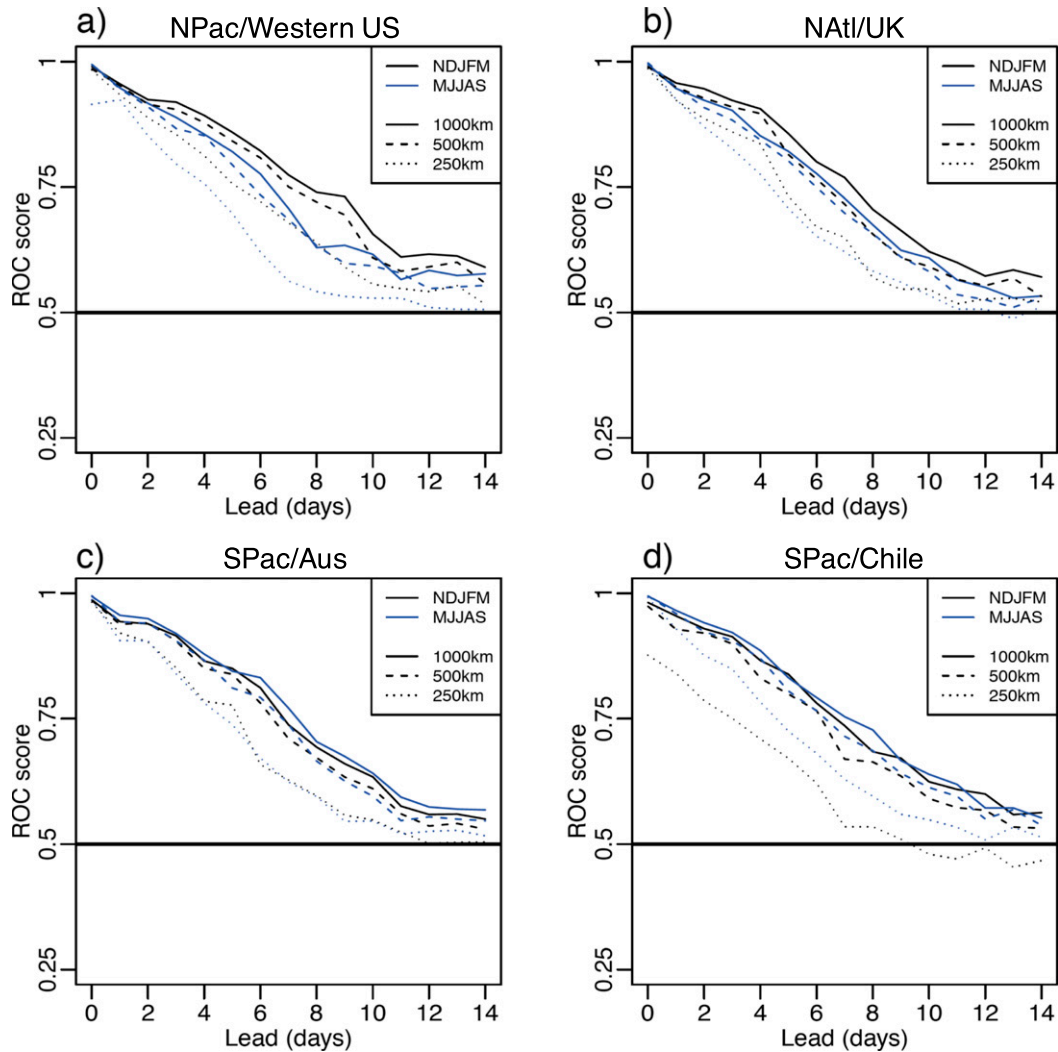


FIG. 7. The 1996–2013 ECMWF area under ROC curve (i.e., ROC score) vs lead time over the (a) North Pacific/western U.S., (b) North Atlantic/U.K., (c) South Pacific/Australia, and (d) South Pacific/Chile regions during NDJFM (black) and MJJAS (blue) and using distance thresholds of 1000 (solid), 500 (dashed), and 250 km (dotted).

A more detailed overview of atmospheric teleconnections related to these modes of variability is provided below.

4) OVERVIEW OF ATMOSPHERIC TELECONNECTIONS AND DESCRIPTION OF CLIMATE INDICES USED

Large-scale extratropical midlatitude dynamics can be strongly influenced by remote variations in tropical Pacific sea surface temperature (SST) associated with ENSO (Bjerknes 1969; Rasmusson and Wallace 1983; McCabe and Dettinger 1999; Meehl et al. 2007; DeFlorio et al. 2013). This connection between climate mode variability (e.g., ENSO) and atmospheric circulation can

have significant impacts on the orientation and frequency of ARs and resultant extreme precipitation (Cordeira et al. 2013; GW2015). In the North Pacific, variations in extratropical geopotential height anomalies that define the PNA are associated with distinct weather variations over the United States and have been shown to respond linearly to changes in tropical Pacific SST (Tsonis et al. 2008; Wallace and Gutzler 1981). Negative PNA conditions are associated with a weakening of the Aleutian low (Renwick and Wallace 1996) and more zonal flow over the North Pacific/western U.S. midlatitudes (Leathers et al. 1991, their Fig. 1).

The following climate indices, obtained from the National Oceanic and Atmospheric Administration’s

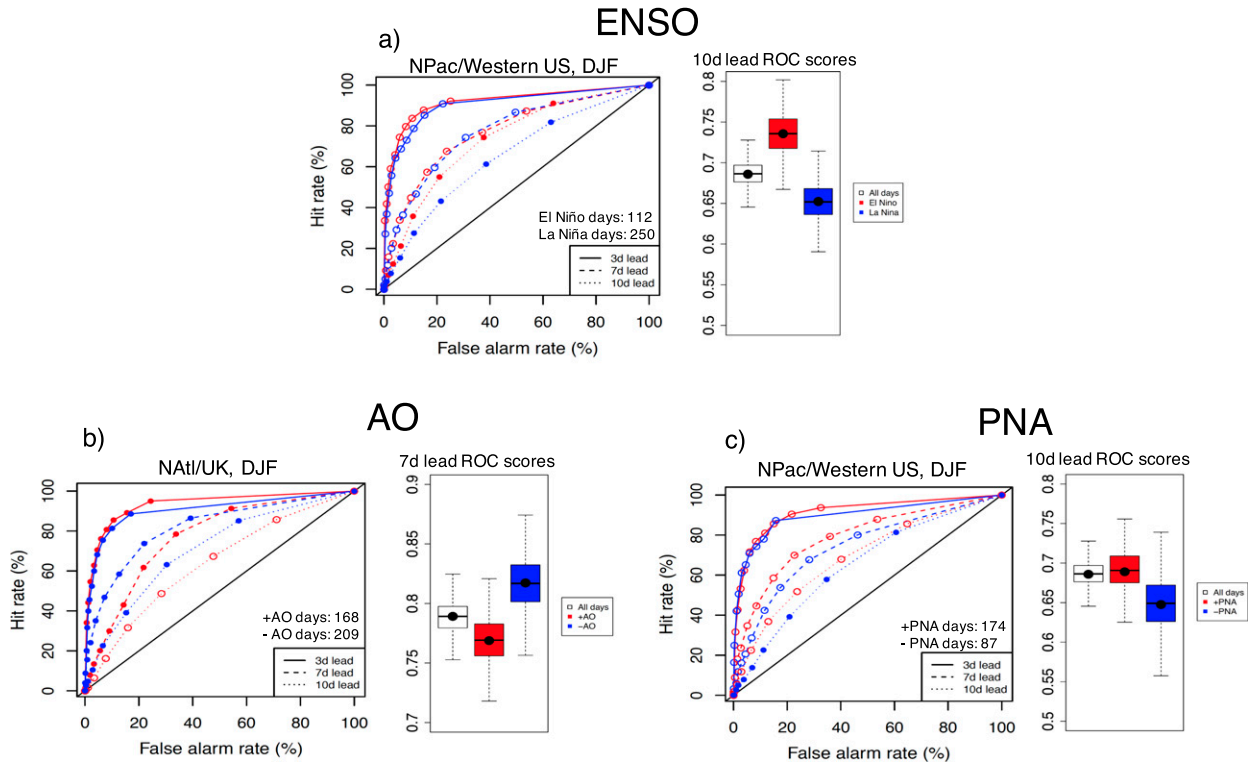


FIG. 8. ROC curves composited on positive (red) and negative (blue) phases of (a) ENSO in DJF over the North Pacific/western U.S. region, (b) AO in DJF over the North Atlantic/U.K. region, and (c) PNA teleconnection in DJF over the North Pacific/western U.S. region. The 1000-km distance threshold is used, and positive and negative phases are defined using ± 0.5 standardized values of the climate index for each mode. Lead times of 3 days (solid), 7 days (dashed), and 10 days (dotted) are shown. The number of positive and negative phase days for each mode phase is listed above the legend. Area under ROC curve distributions for each region/mode/lead time of relevance, calculated from a bootstrap process that was repeated 1000 times by using resampling of the composite positive and negative mode days (red and blue, respectively) and all days (white) distributions with replacement, is included beside the ROC curves.

Climate Prediction Center, are used in this study: 1) the Niño-3.4 index (mean SST anomalies in the central/eastern equatorial Pacific between 5°S and 5°N and between 190° and 240°E) and 2) daily values of the AO and PNA teleconnection indices, which are computed by projecting daily geopotential height anomalies at 1000 hPa (for AO) and 500 hPa (for PNA) onto their empirical orthogonal function loading patterns.

5) CLIMATE-CONDITIONED ROC CURVES

ROC curves for positive and negative phases of each mode (defined as 0.5 standard deviation above or below zero based on their respective indices) at hindcast initialization over the North Pacific/western U.S. region during DJF (for ENSO), the North Atlantic/U.K. region during DJF (for AO), and the North Pacific/western U.S. region during DJF (for PNA) at 3-day (solid line), 7-day (dashed line), and 10-day (dotted line) lead times are shown Figs. 8a–c. Focus is restricted to DJF in an attempt to capture periods where teleconnected climate variability is most prevalent. To estimate the statistical significance of the results, a bootstrap process is

repeated 1000 times with replacement to create ROC score distributions (shown to the right of each ROC curve plot) by using resampling of the positive mode days (red box plots), negative mode days (blue box plots), and “all days” distribution (white boxplots). If the difference between the actual ROC score of the all-days distribution and either positive or negative mode distribution lies outside of the 5th or 95th percentile of the distribution of bootstrap ROC score differences, then the ROC curve circles (representing each trigger threshold value) are filled. If the difference is insignificant at the 90% confidence level, then the ROC curve circles are unfilled. ROC curves that have large ROC score differences between positive and negative phases in the bootstrapped samples are shown in Fig. 8.

These results indicate that AR hindcasts from the ECMWF model are more useful (i.e., relatively higher ratio of hits to false alarms) during El Niño conditions over the North Pacific/western U.S. region at a 10-day lead time in boreal winter, and during negative AO conditions at a 7-day lead time over the North Atlantic/U.K. region boreal winter. Hindcasts are less useful (i.e.,

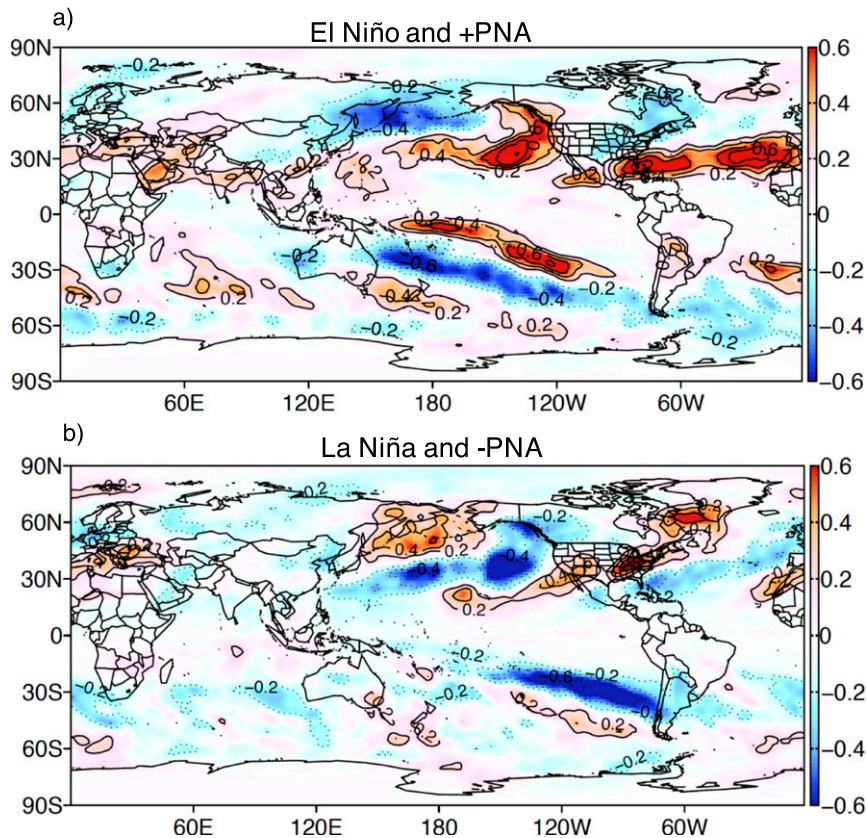


FIG. 9. The 1996–2013 ERA-I anomalies of AR occurrence on (a) El Niño and +PNA conditions and (b) La Niña and –PNA conditions. AR occurrence is defined as the number of ARs over a 14-day time window covering a grid cell. Anomalies are computed as the AR occurrence values on phase-locked ENSO/PNA dates minus the long-term mean AR occurrence values.

relatively lower ratio of hits to false alarms) during negative PNA conditions at a 10-day lead time over the North Pacific/western U.S. region in boreal winter. Observed AR frequency anomalies are relatively large for the region and mode pairings examined here (shown in GW2015), which is why these pairings are shown. AR hindcast utility is also higher during La Niña and negative PNA conditions over the North Atlantic/U.K. region at a 7-day lead time in boreal winter (Fig. S1 in the supplemental material).

Observed anomalies of AR occurrence (1996–2013, ERA-I) are shown in Fig. 9 for phase-locked conditions of ENSO and PNA (El Niño and +PNA in Fig. 9a; La Niña and –PNA in Fig. 9b). AR occurrence is defined as the number of ARs over a 14-day time window covering a grid cell. Anomalies are computed as the AR occurrence values on phase-locked ENSO/PNA days minus the long-term mean AR occurrence values. The spatial pattern of the individual mode AR occurrence anomalies is reminiscent of the individual mode AR frequency anomalies shown in GW2015. It is clear from

Fig. 9 that over the 1996–2013 observed period, phase-locked AR occurrence anomalies are enhanced relative to the individual mode magnitudes over certain regions shown in GW2015 (e.g., North Pacific/western United States).

ROC curves for positive and negative phases of the phase-locked mode conditions are shown in Fig. 10 at 3-day (solid line), 7-day (dashed line), and 10-day (dotted line) lead times during DJF over the North Pacific/western U.S. region and over the North Atlantic/U.K. region, along with the ROC scores from the bootstrapped distributions for each phase-locked mode conditions at 10- and 7-day lead times, respectively. Mean area under ROC curve scores for bootstrapped samples shown in Figs. 8 and 10 are displayed in Table 1. Over the North Pacific/western U.S. region at a 10-day lead, the difference between El Niño and +PNA and La Niña and –PNA mean ROC score shown in Fig. 10a is twice as large as the difference between either the El Niño and La Niña mean ROC score shown in Fig. 8a or the +PNA and –PNA mean ROC score shown in

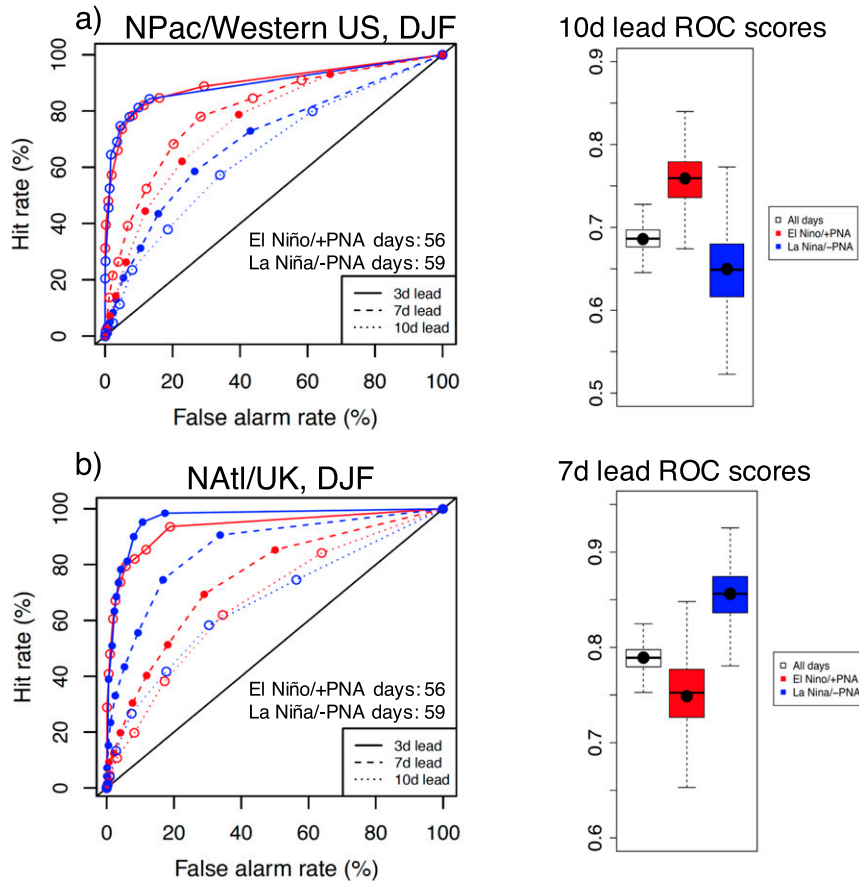


FIG. 10. As in Fig. 8, but for the combined ENSO and PNA teleconnection modes in DJF over (a) the North Pacific/western U.S. region and (b) the North Atlantic/U.K. region.

Fig. 8c. Similarly, over the North Atlantic/U.K. region, the difference shown in Fig. 10b is larger than the differences between the individual mode mean ROC scores shown in Fig. S1. These results indicate that the difference in AR forecast utility on positive and negative phase-locked conditions can be even larger than on positive and negative conditions of the individual modes themselves, suggesting the potential for “forecasts of opportunity” when a forecast is initialized during these phase-locked periods.

Although the influence of ENSO and PNA on North Pacific/western U.S. climate is well known, the linkage between these modes and circulation over the North Atlantic/U.K. region is less obvious. Brönnimann (2007) summarized the impact of ENSO on European climate and noted that in late winter, El Niño (La Niña) conditions generally resemble negative (positive) NAO conditions, but that the signal is different in early winter and spring. This review paper also notes that the mean ENSO signal over Europe is not strong, but can vary and is more important if the local atmosphere is preconditioned in a particular flow regime. Stan et al. (2017) note

that in general, a primary response of the extratropical circulation to organized tropical convection is modified Rossby wave propagation and midlatitude storm track amplitude and orientation. Song et al. (2009) found an anticorrelation between the PNA and NAO during boreal winter that they attributed to anomalous Rossby wave breaking events associated with the PNA pattern. Similarly, Drouard et al. (2015) found that geopotential height anomalies over the North Pacific/western U.S. region associated with ENSO and/or the PNA pattern modify the direction of downstream wave propagation and subsequent wave breaking over the North Atlantic region. It is therefore plausible that individual and phase-locked ENSO and PNA conditions could affect climate (and therefore prediction skill of ARs, which is intimately related to regional circulation) over the North Atlantic/U.K. region via anomalous upstream Rossby wave breaking events in the midlatitudes (ENSO projects onto midlatitude North Pacific climate by modulating the phase or amplitude of the subtropical jet, which also influences the PNA pattern, and these wave perturbations are then felt downstream over Europe days later).

TABLE 1. Mean area under ROC curve scores for bootstrapped samples shown in Figs. 8 and 10. The left number indicates the score for the positive phase of the mode, the middle number indicates the score for all hindcast days, the right number indicates the score for the negative phase of the mode, and the boldface indicates a statistically significant difference between a given mode phase and the all-days case at the 90% confidence level.

	North Pacific/western United States (DJF)	North Atlantic/United Kingdom (DJF)
ENSO		
3-day lead	0.92, 0.91, 0.91	0.93, 0.92, 0.92
7-day lead	0.77, 0.77, 0.77	0.73 , 0.79, 0.81
10-day lead	0.74 , 0.69, 0.65	0.66, 0.66, 0.66
AO		
3-day lead	0.91, 0.91, 0.90	0.93 , 0.92, 0.90
7-day lead	0.78, 0.77, 0.78	0.77 , 0.79, 0.82
10-day lead	0.66, 0.69, 0.69	0.64, 0.66, 0.71
PNA		
3-day lead	0.91, 0.91, 0.90	0.92, 0.92, 0.95
7-day lead	0.79, 0.77, 0.74	0.77, 0.79, 0.83
10-day lead	0.69, 0.69, 0.65	0.64, 0.66, 0.68
ENSO + PNA		
3-day lead	0.90, 0.91, 0.90	0.94, 0.92, 0.96
7-day lead	0.80, 0.77, 0.69	0.75 , 0.79, 0.86
10-day lead	0.76 , 0.69, 0.65	0.67, 0.66, 0.66

4. Conclusions

This study builds on recent advances in objective detection of atmospheric rivers that can be applied to both observations (GW2015) and global weather and climate models (Guan and Waliser 2017) and the availability of subseasonal hindcasts from operational models (Vitart et al. 2017) to quantify global prediction skill of ARs using two decades of ECMWF hindcast data at lead times ranging from 1 to 14 days. AR prediction skill is first quantified by diagnosing the ability of the ECMWF model to predict an observed AR as a function of lead time and distance threshold between the centroid of an observed and forecasted AR, averaged over all hindcasts. The average percentage of ECMWF ensemble members that skillfully predict an AR shows highest variability during boreal winter in the North Pacific/western United States (Fig. 4). Southern Hemispheric regions exhibit little seasonal dependence in AR forecast skill. Over the North Pacific/western U.S. region, the lead time at which AR hit percentage equals 50% increases by 2.5 days during NDJFM relative to MJJAS using a 1000-km threshold, and the AR hit percentage at a 7-day lead time increases by 15%–20% during NDJFM relative to MJJAS using a 1000-km distance threshold.

This study also examines 1996–2013 average ROC curves, which explicitly account for both hits (forecasted AR and observed AR) and false alarms (forecasted AR

but no observed AR). The long-term mean curves are shown in Fig. 6 and are composited on various phases of climate variability at hindcast initialization in Figs. 8 and 10. This alternative methodology quantifies the extent to which an AR forecast is more useful than a random guess. Evidence for statistical increases in AR forecast utility (i.e., relatively higher ratio of hits to false alarms) at a 10-day lead is demonstrated during El Niño conditions over the North Pacific/western U.S. region in boreal winter and at a 7-day lead during negative AO conditions over the North Atlantic/U.K. region in boreal winter. Evidence for statistically significant decreases in AR forecast utility (i.e., relatively lower ratio of hits to false alarms) is demonstrated at a 10-day lead during negative PNA conditions over the North Pacific/western U.S. region in boreal winter. It is also shown that the difference in AR forecast utility between positive and negative phase-locked conditions of ENSO and PNA is at least twice as large over the North Pacific/western U.S. region and North Atlantic/U.K. region compared to the positive and negative individual mode conditions, suggesting the potential for “forecasts of opportunity” when a forecast is initialized during these phase-locked periods of climate variability.

Future studies must address several aspects of quantifying AR prediction skill that are not explored in this study. Multimodel assessment of AR prediction skill is needed in order to test these results with respect to sensitivities in model configuration, hindcast frequency, ensemble size, and record period. Additional features of the ARs themselves (e.g., orientation with respect to coastline and orography) might also be considered, as they can have significant impacts on resultant extreme precipitation. More targeted metrics of AR skill (e.g., landfall location) might be considered in future studies to more directly target water resource management and decisions applications. The composite ROC results should be calculated using different time periods to account for potential nonstationarity of climate mode teleconnections (e.g., Diaz et al. 2001). Subsequent efforts might also quantify potential predictability limits (e.g., Waliser et al. 2003; Lavers et al. 2016a) of AR objects on subseasonal time scales, as well as AR frequency of occurrence, intensity, and landfall location. Finally, future studies might evaluate forecasts of greater than 7 days using time windows greater than 24 h, which could yield greater skill than the values reported here.

Acknowledgments. We gratefully acknowledge the availability of the S2S hindcast database that makes this work possible. S2S is a joint initiative of the World Weather Research Programme (WWRP) and the World Climate Research Program (WCRP). The original S2S

database is hosted at ECMWF as an extension of the “TIGGE database.” We would like to acknowledge support from the NASA Energy and Water Cycle Program and the California Department of Water Resources. We thank the two anonymous reviewers for greatly improving the quality of the manuscript. M.D.’s and D.W.’s contributions to this study were carried out on behalf of the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

REFERENCES

- Abid, M. A., I.-S. Kang, M. Almazroui, and F. Kucharski, 2015: Contribution of synoptic transients to the potential predictability of PNA circulation anomalies: El Niño vs La Niña. *J. Climate*, **28**, 8347–8362, <https://doi.org/10.1175/JCLI-D-14-00497.1>.
- American Meteorological Society, 2017: Atmospheric river. Glossary of Meteorology, http://glossary.ametsoc.org/wiki/Atmospheric_river.
- Barnes, L. R., D. M. Schultz, E. C. Grunfest, M. H. Hayden, and C. C. Benight, 2009: Corrigendum: False alarm rate or false alarm ratio? *Wea. Forecasting*, **24**, 1452–1454, <https://doi.org/10.1175/2009WAF2222300.1>.
- Barnston, A. G., M. K. Tippett, M. L. L’Heureux, S. Li, and D. G. DeWitt, 2012: Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing? *Bull. Amer. Meteor. Soc.*, **93**, 631–651, <https://doi.org/10.1175/BAMS-D-11-00111.1>.
- Bjerknes, J., 1969: Atmospheric teleconnections from the equatorial Pacific. *Mon. Wea. Rev.*, **97**, 163–172, [https://doi.org/10.1175/1520-0493\(1969\)097<0163:ATFTEP>2.3.CO;2](https://doi.org/10.1175/1520-0493(1969)097<0163:ATFTEP>2.3.CO;2).
- Black, J., N. C. Johnson, S. Baxter, S. B. Feldstein, D. S. Harnos, and M. L. L’Heureux, 2017: The predictors and forecast skill of Northern Hemisphere teleconnection patterns for lead times of 3–4 weeks. *Mon. Wea. Rev.*, **145**, 2855–2877, <https://doi.org/10.1175/MWR-D-16-0394.1>.
- Brönnimann, S., 2007: Impact of El Niño–Southern Oscillation on European climate. *Rev. Geophys.*, **45**, RG3003, <https://doi.org/10.1029/2006RG000199>.
- Coelho, C. A. S., C. B. Uvo, and T. Ambrizzi, 2002: Exploring the impacts of the tropical Pacific SST on the precipitation patterns over South America during ENSO periods. *Theor. Appl. Climatol.*, **71**, 185–197, <https://doi.org/10.1007/s007040200004>.
- Cordeira, J. M., F. M. Ralph, and B. J. Moore, 2013: The development and evolution of two atmospheric rivers in proximity to western North Pacific tropical cyclones in October 2010. *Mon. Wea. Rev.*, **141**, 4234–4255, <https://doi.org/10.1175/MWR-D-13-00019.1>.
- Davis, C. A., B. Brown, and R. Bullock, 2006: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784, <https://doi.org/10.1175/MWR3145.1>.
- Dayan, H., T. Izumo, J. Vialard, and M. Lengaigne, 2014: Does sea surface temperature outside the tropical Pacific contribute to enhanced ENSO predictability? *Climate Dyn.*, **43**, 1311–1325, <https://doi.org/10.1007/s00382-013-1946-y>.
- Dee, D., and Coauthors, 2011a: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- DeFlorio, M. J., D. W. Pierce, D. R. Cayan, and A. J. Miller, 2013: Western U.S. extreme precipitation events and their relation to ENSO and PDO in CCSM4. *J. Climate*, **26**, 4231–4243, <https://doi.org/10.1175/JCLI-D-12-00257.1>.
- Dettinger, M. D., 2013: Atmospheric rivers as drought busters on the U.S. West Coast. *J. Hydrometeorol.*, **14**, 1721–1732, <https://doi.org/10.1175/JHM-D-13-02.1>.
- , F. M. Ralph, T. Das, P. J. Neiman, and D. R. Cayan, 2011: Atmospheric rivers, floods and the water resources of California. *Water*, **3**, 445–478, <https://doi.org/10.3390/w3020445>.
- Diaz, H. F., M. P. Hoerling, and J. K. Eischeid, 2001: ENSO variability, teleconnections, and climate change. *Int. J. Climatol.*, **21**, 1845–1862, <https://doi.org/10.1002/joc.631>.
- Drouard, M., G. Rivière, and P. Arbogast, 2015: The link between the North Pacific climate variability and the North Atlantic Oscillation via downstream propagation of synoptic waves. *J. Climate*, **28**, 3957–3976, <https://doi.org/10.1175/JCLI-D-14-00552.1>.
- Ebert, E. E., 2008: Fuzzy verification of high resolution gridded forecasts: A review and proposed framework. *Meteor. Appl.*, **15**, 51–64, <https://doi.org/10.1002/met.25>.
- , and J. L. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrol.*, **239**, 179–202, [https://doi.org/10.1016/S0022-1694\(00\)00343-7](https://doi.org/10.1016/S0022-1694(00)00343-7).
- Eiras-Barca, J., S. Brands, and G. Miguez-Macho, 2016: Seasonal variations in North Atlantic atmospheric river activity and associations with anomalous precipitation over the Iberian Atlantic Margin. *J. Geophys. Res. Atmos.*, **121**, 931–948, <https://doi.org/10.1002/2015JD023379>.
- Guan, B., and D. E. Waliser, 2015: Detection of atmospheric rivers: Evaluation and application of an algorithm for global studies. *J. Geophys. Res. Atmos.*, **120**, 12 514–12 535, <https://doi.org/10.1002/2015JD024257>.
- , and —, 2017: Atmospheric rivers in 20 year weather and climate simulations: A multimodel, global evaluation. *J. Geophys. Res. Atmos.*, **122**, 5556–5581, <https://doi.org/10.1002/2016JD026174>.
- , N. P. Molotch, D. E. Waliser, E. J. Fetzer, and P. J. Neiman, 2010: Extreme snowfall events linked to atmospheric rivers and surface air temperature via satellite measurements. *Geophys. Res. Lett.*, **37**, L20401, <https://doi.org/10.1029/2010GL044696>.
- Hanley, J. A., and B. J. McNeil, 1982: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29–36, <https://doi.org/10.1148/radiology.143.1.7063747>.
- Lavers, D. A., and G. Villarini, 2015: The contribution of atmospheric rivers to precipitation in Europe and the United States. *J. Hydrol.*, **522**, 382–390, <https://doi.org/10.1016/j.jhydrol.2014.12.010>.
- , F. Pappenberger, and E. Zsoter, 2014: Extending medium-range predictability of extreme hydrological events in Europe. *Nat. Commun.*, **5**, 5382, <https://doi.org/10.1038/ncomms6382>.
- , D. E. Waliser, F. M. Ralph, and M. D. Dettinger, 2016a: Predictability of horizontal water vapor transport relative to precipitation: Enhancing situational awareness for forecasting western U.S. extreme precipitation and flooding. *Geophys. Res. Lett.*, **43**, 2275–2282, <https://doi.org/10.1002/2016GL067765>.
- , F. Pappenberger, D. S. Richardson, and E. Zsoter, 2016b: ECMWF Extreme Forecast Index for water vapor transport:

- A forecast tool for atmospheric rivers and extreme precipitation. *Geophys. Res. Lett.*, **43**, 11 852–11 858, <https://doi.org/10.1002/2016GL071320>.
- Leathers, D. J., B. Yarnal, and M. A. Palecki, 1991: The Pacific/North American teleconnection pattern and United States climate. Part I: Regional temperature and precipitation associations. *J. Climate*, **4**, 517–528, [https://doi.org/10.1175/1520-0442\(1991\)004<0517:TPATPA>2.0.CO;2](https://doi.org/10.1175/1520-0442(1991)004<0517:TPATPA>2.0.CO;2).
- Li, S., and A. W. Robertson, 2015: Evaluation of submonthly precipitation forecast skill from global ensemble prediction systems. *Mon. Wea. Rev.*, **143**, 2871–2889, <https://doi.org/10.1175/MWR-D-14-00277.1>.
- Mason, S. J., and N. E. Graham, 2002: Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Quart. J. Roy. Meteor. Soc.*, **128**, 2145–2166, <https://doi.org/10.1256/003590002320603584>.
- McCabe, G. J., and M. D. Dettinger, 1999: Decadal variations in the strength of ENSO teleconnections with precipitation in the western United States. *Int. J. Climatol.*, **19**, 1399–1410, [https://doi.org/10.1002/\(SICI\)1097-0088\(19991115\)19:13<1399::AID-JOC457>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1097-0088(19991115)19:13<1399::AID-JOC457>3.0.CO;2-A).
- Meehl, G. A., C. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. B. Mitchell, R. J. Stouffer, and K. E. Taylor, 2007: The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bull. Amer. Meteor. Soc.*, **88**, 1383–1394, <https://doi.org/10.1175/BAMS-88-9-1383>.
- Mittermaier, M. P., and D. B. Stephenson, 2015: Inherent bounds on accuracy due to observations uncertainty caused by temporal sampling. *Mon. Wea. Rev.*, **143**, 4236–4243, <https://doi.org/10.1175/MWR-D-15-0173.1>.
- Moore, B. J., K. M. Mahoney, E. M. Sukovich, R. Cifelli, and T. M. Hamill, 2015: Climatology and environmental characteristics of extreme precipitation events in the southeastern United States. *Mon. Wea. Rev.*, **143**, 718–741, <https://doi.org/10.1175/MWR-D-14-00065.1>.
- Mundhenk, B., E. A. Barnes, and E. Maloney, 2016: All-season climatology and variability of atmospheric river frequencies over the North Pacific. *J. Climate*, **29**, 4885–4903, <https://doi.org/10.1175/JCLI-D-15-0655.1>.
- Muñoz, Á. G., L. Goddard, A. W. Robertson, Y. Kushnir, and W. Baethgen, 2015: Cross-time scale interactions and rainfall extreme events in southeastern South America for the austral summer. Part I: Potential predictors. *J. Climate*, **28**, 7894–7913, <https://doi.org/10.1175/JCLI-D-14-00693.1>.
- Nayak, M. A., G. Villarini, and D. A. Lavers, 2014: On the skill of numerical weather prediction models to forecast atmospheric rivers over the central United States. *Geophys. Res. Lett.*, **41**, 4354–4362, <https://doi.org/10.1002/2014GL060299>.
- , —, and A. A. Bradley, 2016: Atmospheric rivers and rainfall during NASA's Iowa Flood Studies (IFloodS) campaign. *J. Hydrometeorol.*, **17**, 257–271, <https://doi.org/10.1175/JHM-D-14-0185.1>.
- Neiman, P. J., F. M. Ralph, A. B. White, J. Lundquist, and M. D. Dettinger, 2008: Meteorological characteristics and overland precipitation impacts of atmospheric rivers affecting the West Coast of North America based on eight years of SSM/I satellite observations. *J. Hydrometeorol.*, **9**, 22–47, <https://doi.org/10.1175/2007JHM855.1>.
- Pezzi, L. P., and I. F. A. Cavalcanti, 2001: The relative importance of ENSO and tropical Atlantic sea surface temperature anomalies for seasonal precipitation over South America: A numerical study. *Climate Dyn.*, **17**, 205–212, <https://doi.org/10.1007/s003820000104>.
- Ralph, F. M., and M. D. Dettinger, 2011: Storms, floods, and the science of atmospheric rivers. *Eos, Trans. Amer. Geophys. Union*, **92**, 265–266, <https://doi.org/10.1029/2011EO320001>.
- , and —, 2012: Historical and national perspectives on extreme west coast precipitation associated with atmospheric rivers during December 2012. *Bull. Amer. Meteor. Soc.*, **93**, 783–790, <https://doi.org/10.1175/BAMS-D-11-00188.1>.
- , P. J. Neiman, and G. A. Wick, 2004: Satellite and CALJET aircraft observations of atmospheric rivers over the eastern North Pacific Ocean during the El Niño winter of 1997/98. *Mon. Wea. Rev.*, **132**, 1721–1745, [https://doi.org/10.1175/1520-0493\(2004\)132<1721:SACAOO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1721:SACAOO>2.0.CO;2).
- , —, and R. Rotunno, 2005: Dropsonde observations in low-level jets over the northeastern Pacific Ocean from CALJET-1998 and PACJET-2001: Mean vertical-profile and atmospheric-river characteristics. *Mon. Wea. Rev.*, **133**, 889–910, <https://doi.org/10.1175/MWR2896.1>.
- , —, G. A. Wick, S. I. Gutman, M. D. Dettinger, D. R. Cayan, and A. B. White, 2006: Flooding on California's Russian River: Role of atmospheric rivers. *Geophys. Res. Lett.*, **33**, L13801, <https://doi.org/10.1029/2006GL026689>.
- , and Coauthors, 2016: CalWater field studies designed to quantify the roles of atmospheric rivers and aerosols in modulating U.S. West Coast precipitation in a changing climate. *Bull. Amer. Meteor. Soc.*, **97**, 1209–1228, <https://doi.org/10.1175/BAMS-D-14-00043.1>.
- Rasmusson, E. M., and J. M. Wallace, 1983: Meteorological aspects of the El Niño/Southern Oscillation. *Science*, **222**, 1195–1202, <https://doi.org/10.1126/science.222.4629.1195>.
- Renwick, J. A., and J. M. Wallace, 1996: Relationships between North Pacific wintertime blocking, El Niño, and the PNA pattern. *Mon. Wea. Rev.*, **124**, 2071–2076, [https://doi.org/10.1175/1520-0493\(1996\)124<2071:RBNPWB>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<2071:RBNPWB>2.0.CO;2).
- Sodemann, H., and A. Stohl, 2013: Moisture origin and meridional transport in atmospheric rivers and their association with multiple cyclones. *Mon. Wea. Rev.*, **141**, 2850–2868, <https://doi.org/10.1175/MWR-D-12-00256.1>.
- Song, J., C. Li, W. Zhou, and J. Pan, 2009: The linkage between the Pacific–North American teleconnection pattern and the North Atlantic Oscillation. *Adv. Atmos. Sci.*, **26**, 229–239, <https://doi.org/10.1007/s00376-009-0229-3>.
- Stan, C., D. M. Straus, J. S. Frederiksen, H. Lin, E. D. Maloney, and C. Schumacher, 2017: Review of tropical-extratropical teleconnections on intraseasonal time scales. *Rev. Geophys.*, **55**, 902–937, <https://doi.org/10.1002/2016RG000538>.
- Sukovich, E. M., F. M. Ralph, F. E. Barthold, D. W. Reynolds, and D. R. Novak, 2014: Extreme quantitative precipitation forecast performance at the weather prediction center from 2001 to 2011. *Wea. Forecasting*, **29**, 894–911, <https://doi.org/10.1175/WAF-D-13-00061.1>.
- Taschetto, A. S., R. R. Rodrigues, G. A. Meehl, S. McGregor, and M. H. England, 2016: How sensitive are the Pacific–tropical North Atlantic teleconnections to the position and intensity of El Niño-related warming? *Climate Dyn.*, **46**, 1841–1860, <https://doi.org/10.1007/s00382-015-2679-x>.
- Tsonis, A. A., K. L. Swanson, and G. Wang, 2008: On the role of atmospheric teleconnections in climate. *J. Climate*, **21**, 2990–3001, <https://doi.org/10.1175/2007JCLI1907.1>.
- Vitart, F., and Coauthors, 2017: The Subseasonal to Seasonal (S2S) Prediction project database. *Bull. Amer. Meteor.*

- Soc.*, **98**, 163–173, <https://doi.org/10.1175/BAMS-D-16-0017.1>.
- Waliser, D. E., and B. Guan, 2017: Extreme winds and precipitation during landfall of atmospheric rivers. *Nat. Geosci.*, **10**, 179–183, <https://doi.org/10.1038/ngeo2894>.
- , K. M. Lau, W. Stern, and C. Jones, 2003: Potential predictability of the Madden–Julian oscillation. *Bull. Amer. Meteor. Soc.*, **84**, 33–50, <https://doi.org/10.1175/BAMS-84-1-33>.
- Wallace, J., and D. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.*, **109**, 784–812, [https://doi.org/10.1175/1520-0493\(1981\)109<0784:TITGHF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2).
- Wick, G. A., P. J. Neiman, F. M. Ralph, and T. M. Hamill, 2013: Evaluation of forecasts of the water vapor signature of atmospheric rivers in operational numerical weather prediction models. *Wea. Forecasting*, **28**, 1337–1352, <https://doi.org/10.1175/WAF-D-13-00025.1>.
- Zhu, Y., and R. E. Newell, 1998: A proposed algorithm for moisture fluxes from atmospheric rivers. *Mon. Wea. Rev.*, **126**, 725–735, [https://doi.org/10.1175/1520-0493\(1998\)126<0725:APAFMF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<0725:APAFMF>2.0.CO;2).