Geophysical Research Letters

RESEARCH LETTER
10.1029/2023GL105360

Subseasonal Prediction of Impactful California Winter Weather in a Hybrid Dynamical-Statistical Framework

Kristen Guirguis[1], Alexander Gershunov[1], Benjamin J. Hatchett[2], Michael J. DeFlorio[1], Aneesh C. Subramanian[3], Rachel Clemesha[1], Luca Delle Monache[1], and F. Martin Ralph[1]

[1]Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA, [2]Desert Research Institute, Reno, NV, USA, [3]University of Colorado Boulder, Boulder, CO, USA

Key Points:

- A hybrid dynamical-statistical model is described for 1–4-week forecasts of impactful California winter weather using circulation regimes
- This hybrid framework reduces the number of forecasts produced, but the ones issued can be interpreted with higher confidence
- This new methodology provides skillful subseasonal forecasts with potential to improve early warnings for impactful weather events

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

K. Guirguis,
kguirguis@ucsd.edu

Citation:

Guirguis, K., Gershunov, A., Hatchett, B. J., DeFlorio, M. J., Subramanian, A. C., Clemesha, R., et al. (2023). Subseasonal prediction of impactful California winter weather in a hybrid dynamical-statistical framework. *Geophysical Research Letters*, *50*, e2023GL105360. https://doi.org/10.1029/2023GL105360

Received 7 JUL 2023
Accepted 6 NOV 2023

Author Contributions:

Conceptualization: Alexander Gershunov, Benjamin J. Hatchett, Michael J. DeFlorio, Aneesh C. Subramanian, Rachel Clemesha, Luca Delle Monache, F. Martin Ralph
Data curation: Aneesh C. Subramanian
Funding acquisition: Alexander Gershunov, F. Martin Ralph

Abstract Atmospheric rivers (ARs) and Santa Ana winds (SAWs) are impactful weather events for California communities. Emergency planning efforts and resource management would benefit from extending lead times of skillful prediction for these and other types of extreme weather patterns. Here we describe a methodology for subseasonal prediction of impactful winter weather in California, including ARs, SAWs and heat extremes. The hybrid approach combines dynamical model and historical information to forecast probabilities of impactful weather outcomes at weeks 1–4 lead. This methodology uses dynamical model information considered most reliable, that is, planetary/synoptic-scale atmospheric circulation, filters for dynamical model error/uncertainty at longer lead times and increases the sample of likely outcomes by utilizing the full historical record instead of a more limited suite of dynamical forecast model ensemble members. We demonstrate skill above climatology at subseasonal timescales, highlighting potential for use in water, health, land, and fire management decision support.

Plain Language Summary California winter weather can alternate between very wet conditions from atmospheric rivers making landfall along the Pacific coast to hot, dry, and windy conditions brought by Santa Ana winds blowing in from the Southwest interior. Atmospheric rivers are important for water resources while also causing flooding, whereas Santa Ana winds are often associated with wildfire, especially following prolonged dry periods. Preparing for these types of weather events is important for managing resources and protecting life and property, yet reliable forecasts beyond about 7–10 days remain a challenge. We have developed a new prediction system that combines information about approaching atmospheric weather patterns from weather forecast models along with historical information relating those patterns to impacts over California to predict the likelihood of impactful weather at 1–4 weeks lead time. By extending the window of opportunity to take management action, this new approach should aid in resource and emergency planning in water, land, and fire sectors as well as protecting residents through improved warning systems.

1. Introduction

Extremes of California's winter weather variability can include heavy multiday precipitation from Pacific storms associated with atmospheric rivers (ARs) or dry offshore downslope winds blowing from the elevated continental interior. ARs cause most of the region's floods (Corringham et al., 2019; Dettinger, 2013; Ralph et al., 2006, 2011) while downslope winds are often associated with coastal heat waves as well as wildfire and smoke impacts (Abatzoglou, 2013; Aguilera et al., 2021; Cayan et al., 2022; Gershunov et al., 2021; Guzman-Morales et al., 2016; Hughes & Hall, 2010). Winter heat waves and dry spells accelerate mountain snowmelt (Hatchett et al., 2023), exacerbate drought and endanger human health, particularly along the densely populated coast (Gershunov et al., 2021; Schwarz et al., 2020). Improved prediction of these impactful weather events is critical for emergency preparedness and planning to mitigate impacts to society (DeFlorio et al., 2021). Climate change is increasing the likelihood and intensity of extreme weather in California (e.g., Corringham et al., 2022; Gershunov et al., 2019, 2021; Huang & Swain, 2022; Michaelis et al., 2022), highlighting the need for improved forecasts across a range of lead times to aid planning and mitigate negative outcomes (e.g., DeFlorio et al., 2021; Oakley et al., 2023).

ARs are low-tropospheric jets of water vapor that produce up to 50% of California's annual precipitation (Dettinger et al., 2011; Gershunov et al., 2017). They can be beneficial and hazardous (Ralph et al., 2019); accumulating snowpack (Shulgina et al., 2023) and replenishing water supplies, while also causing the most damag-

Methodology: Alexander Gershunov, Benjamin J. Hatchett, Michael J. DeFlorio, Aneesh C. Subramanian, Rachel Clemesha, Luca Delle Monache, F. Martin Ralph
Supervision: Alexander Gershunov
Visualization: Benjamin J. Hatchett
Writing – original draft: Alexander Gershunov, Benjamin J. Hatchett, Michael J. DeFlorio
Writing – review & editing: Alexander Gershunov, Benjamin J. Hatchett, Michael J. DeFlorio, Aneesh C. Subramanian, Rachel Clemesha, Luca Delle Monache, F. Martin Ralph

ing California floods (Corringham et al., 2019; Guirguis et al. 2020a, 2022). Santa Ana winds (SAWs)—the downslope winds of Southern California—are characterized by strong, dry, gusty northeasterly-easterly winds that warm by adiabatic compression as they flow over and through the Transverse and Peninsular Ranges down to sea level. SAWs can bring anomalously hot or cold temperatures, but the hot variety are associated with Southern California's wildfires (Gershunov et al., 2021). The hot SAWs are increasing in frequency (Guirguis et al., 2022) raising concerns about future wildfire hazard and health impacts.

Numerical weather prediction has notably advanced in recent years. Ensemble probabilistic forecasts provide improvements over deterministic forecasts because they account for uncertainty arising from observational error, model limitations, and the chaotic nature of the earth-atmosphere system (Bauer et al., 2015; Palmer, 2017). This improvement, along with computational and satellite advances, has extended forecast skill and lead time. However, the time limit of predictability for high-impact weather events remains limited to about 1–2 weeks (Bauer et al., 2015) and warnings of heat waves or fire weather are typically issued on the order of a week or less.

Skillful prediction of large-scale weather patterns and regime transitions has been demonstrated at leads of a month or longer (Bauer et al., 2015; Gibson et al., 2020b; Robertson et al., 2020). This has motivated work to extend forecast lead time by focusing on atmospheric circulation patterns and then inferring associated impacts for a region leading to new operational forecast products (e.g., DeFlorio et al., 2021; Ferranti et al., 2015). These studies show dynamical models demonstrate some skill at longer lead times in forecasting certain large-scale circulation features, but this skill is not consistent from forecast to forecast. Progress could be made by developing ways of recognizing when a subseasonal forecast is likely to be skillful or less reliable. In the absence of dynamical model skill, a forecast could be supplemented or replaced by a statistical forecast. Here, we describe and evaluate a dynamical-statistical hybrid prediction system that uses dynamical model forecasts to predict four key modes of atmospheric variability on subseasonal timescales (1–4 weeks lead), filters for uncertainty and error, and draws on known relationships between these modes and high-impact West Coast weather to predict the likelihood of impactful weather events.

Winter weather variability in California is largely modulated by four modes of atmospheric variability over the North Pacific Ocean (called the "NP4 modes," Guirguis et al., 2018, 2020a, 2022) named as the Baja-Pacific (BP), Alaskan-Pacific (AP), Canadian-Pacific (CP) and Offshore-California (OC) modes (Figure 1a). These modes (identified by rotated empirical orthogonal function analysis, Guirguis et al., 2018, hereinafter GGR'18) influence California weather on daily to seasonal timescales (Guirguis et al. 2020a; Guirguis et al., 2022, hereinafter GGR'20 and GGR'22, respectively). While the NP4 modes are not the most dominant modes of variability in the domain (GGR'18), they collectively explain most of the variance (up to 89% in some locations) in mid-tropospheric circulation over a vast region over the North Pacific Ocean and West Coast (e.g., Figure S1 in Supporting Information S1, using multiple linear regression, GGR'20). Daily interactions between these four modes have been linked to reoccurring weather patterns responsible for much of California's daily weather variability and extremes, including wildfires, heat waves, and damaging floods (GGR'20, GGR'22). These modes are also influential for California precipitation on seasonal timescales due to their tendency to persist in one phase or another during a season (GGR'20).

Our forecast system uses the circulation regime methodology of GGR'18, GGR'20 and GGR'22 applied to 20 years of ensemble hindcasts from the European Center for Medium-Range Weather Forecasts (ECMWF) model as well as real-time forecasts from water year 2022 (WY2022). We then apply a statistical model that relates these circulation regime-based forecasts to impactful weather over California. Using this dynamical-statistical hybrid approach, we demonstrate skillful probabilistic forecasts of ARs, SAWs, and heat extremes in California at subseasonal (1–4 weeks) lead times. By filtering out uncertain forecast periods, we improve the accuracy and reliability of the forecasts relative to dynamical model skill without filtering. In this novel approach, we use dynamical model information when it is likely to be reliable, attempt to filter error and uncertainty, and then combine the filtered dynamical model information with a statistical model to forecast an impactful weather event. At shorter lead times (~1 week) the forecast is largely determined by the dynamical model information, whereas at longer lead times (when the dynamical model skill degrades) the statistical information becomes more important. The aim of this work is to provide tools and reliable information for decision support to improve outcomes from extreme weather events.

10.1029/2023GL105360

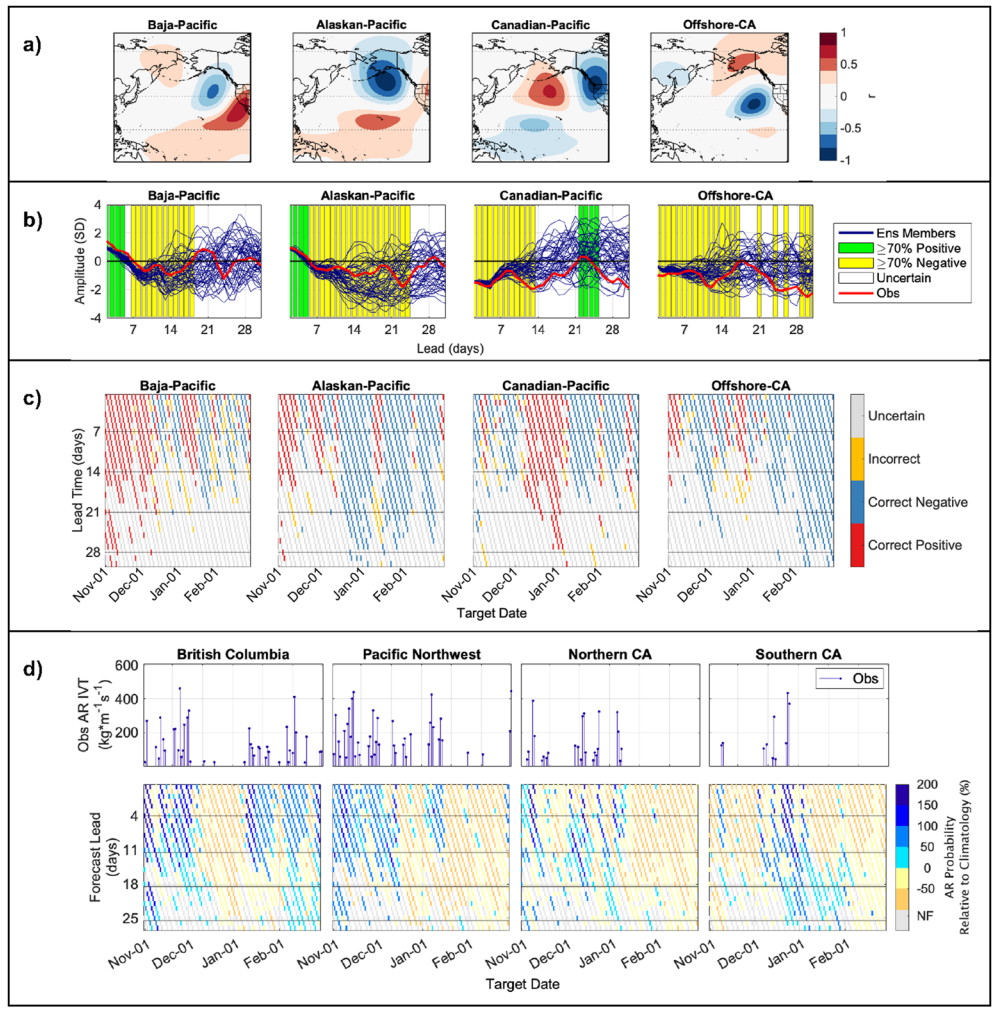

Figure 1. (a) Positive phase of the NP4 modes shown in units of correlation (r) between each mode and observed 500 mb geopotential height anomalies. (b) Example showing the forecast mode amplitude for 50 ensemble members out to 30 days lead, where green (yellow) shading indicates when the 70% consensus is reached for the positive (negative) phase, white indicates uncertainty, and the red line gives the observed mode amplitude. (c) Validated WY2022 forecasts of the NP4 mode phase after filtering, where each individual forecast is displayed on a diagonal line, the x-axis gives the target date from 1 November to 28 February, the y-axis gives the lead time from 1 to 30 days, red (blue) shading indicates that a mode was correctly forecast to be in the positive (negative) phase, yellow indicates error, and gray represents forecasts classified as uncertain. (d) Observed and forecast AR behavior during WY2022 for different coastal regions, where the top four panels (blue stem plots) show observed daily AR IVT, and the bottom panels show AR forecasts for each region at different lead times. Each forecast is shown on a diagonal line, blue indicates above normal AR probability forecasts (wet), yellow/orange shows low probability forecasts (dry), gray indicates uncertainty (no forecast, NF).

2. Data

2.1. Atmospheric Circulation

Atmospheric circulation is represented by daily 500 mb geopotential height (Z500) anomalies from NCEP/NCAR 2.5° Global Reanalysis (R1, Kalnay et al., 1996) available from 1948, where daily anomalies were calculated by fitting and removing annual and semiannual cycles using least-squares regression (GGR'18). To reduce the complexity of the data while retaining the most important features, we use the NP4 mode circulation regimes (Figure 1a) identified in GGR'18 and GGR'20. This data set is available for 1948–2017 (Guirguis et al., 2020b) and was extended through 2022 for this study.

2.2. ECMWF Ensemble Hindcasts and Realtime Forecasts

We use global hindcasts of Z500 from the S2S Project database (Vitart et al., 2017) for 2001–2020. We selected one model, the ECMWF model, a state-of-the-art dynamical weather forecast model shown to outperform other models (e.g., DeFlorio et al., 2019; Gibson et al., 2020b). Ocean coupling is included in these hindcasts, but sea ice coupling is not. Data were produced with the Integrated Forecast System (IFS). Hindcasts are made twice weekly yielding 34 forecasts per year over the 20-year period (680 total), including one control and 10 perturbed ensemble members. We additionally use real-time forecasts of Z500 from the ECMWF during WY2022, produced twice weekly using 50 perturbed ensemble members. The forecasts are available out to 46 days, but we focus on days 1–30 for this study. The focus of this study is extended winter (November-February) spanning the hindcast period 2001–2020 and including WY2022.

2.3. Heat Extremes

Daily maximum temperatures (tmax) are from GridMET (Abatzoglou, 2013), available 1979-present at ∼4 km spatial resolution. Extreme heat is defined as temperatures above the historical 90th percentile after removing the seasonal cycle, where the historical period is defined as 1979–2018 (the period of overlap between temperature, SAW, and AR data sets). We focus on three regions (Figure S2 in Supporting Information S1): the Central Sierra Nevada (a) for their importance for snow accumulation and water resources, as well as coastal Southern California (b) and the San Francisco Bay area (c) where millions of people are exposed to hazards. It is worth noting that climatological probability of heat extremes is slightly higher for the 2001–2020 hindcast period (>10%) than for the historical period (=10% by definition).

2.4. Santa Ana Winds

Santa Ana winds are identified using the daily Santa Ana Winds Regional Index (SAWRI, Guzman-Morales et al., 2016, GM'2016). This record was developed from hourly surface winds spanning 65 years (1948–2012) from dynamically downscaled R1 using the California Regional Spectral Model (CaRD10, Kanamaru & Kanamitsu, 2007). SAWs impacting coastal Southern California were identified when northeasterly wind speeds exceeded the upper quartile (GM'16). Later updates converted the hourly index to a daily record and extended the data set through 2018 using statistical-dynamical downscaling of winds from R1, trained from high-resolution (10×10 km^2) CaRD10 and with results validated against observations (Guzman-Morales & Gershunov, 2019, GM'19). Here, the hindcast assessment of SAWs was limited to the overlapping period of the ECMWF hindcasts and the GM'19 SAW record (2000–2018).

2.5. Atmospheric Rivers

ARs landfalling the West Coast are identified using the SIO-R1 catalog of Gershunov et al. (2017), available 1948-present. The methodology uses data from R1 for vertically integrated horizontal water vapor transport (IVT) and integrated water vapor (IWV) to identify elongated plumes (>1,500 m) of concentrated moisture (IVT > 250 kg m^{-1} s^{-1} and IWV > 15 mm). The SIO-R1 catalog is provided at a 6-hourly temporal resolution, but for this study we define AR landfalls on a daily timescale such that an AR landfall is identified when a coastal location is within the AR footprint for at least one 6-hourly timestep in a day. The SIO-R1 AR detection methodology has been included in intercomparison studies and performs well against other methods while having the advantage of a 70+ year record (Ralph et al., 2019; Rutz et al., 2019).

3. Description of the Dynamical-Statistical Hybrid Model

The hybrid methodology uses information about evolving atmospheric circulation from the ECMWF model, filters for error and uncertainty, and applies a statistical model to predict the likelihood of an impactful weather event.

3.1. Dynamical Model Input and Postprocessing

The dynamical model input consists of ECMWF forecasts of Z500 fields over a domain spanning 20°S–80°N and 120–250°E for each ensemble member, where anomalies are calculated as in Section 2.1. Anomaly maps are

projected onto each of the four NP4 mode EOFs (Figure 1a) to calculate the forecast amplitude of the BP, AP, CP, and OC modes for each ensemble member and lead time. These amplitudes provide information about the magnitude and position of ridges and troughs over the North Pacific and along the West Coast. While the amplitude of these anomalies is extremely important for impacts along the coast, the models struggle to reach agreement about these details beyond short lead times (see Section 3.2). Even reaching a consensus about the phase of these anomalies (a lower bar) is challenging. In our current approach, we use the dynamical model to identify the mode phases, which we can relate to historical weather patterns and impacts based on earlier work (GGR'20, GGR'22).

3.2. Consensus Filtering

We use model confidence to filter for uncertainty using a consensus threshold of 70%. If ≥70% of ensemble members agree about the phase of a given mode, then we assume this information is reliable, whereas if this criterion is not met, then we consider the mode phase to be uncertain. Figure 1b shows an example forecast demonstrating that there is generally strong agreement about the mode phases at short lead times (on the order of 7–10 days), but uncertainty can become prominent at longer lead times. The green and yellow shading indicates where the 70% consensus criterion is met for each mode, with the remaining forecasts classified as uncertain. In physical terms, this means at least 70% of ensemble members agree that a ridge or trough will persist or develop over a certain location at a certain lead time. In this example, the Alaska-Pacific mode is forecast to become negative around day 7 and then persist in that phase for over 2 weeks. The negative phase of this mode is associated with a ridge over the Gulf of Alaska (c.f. Figure 1a). Knowledge about a developing persistent ridge over the Alaskan Gulf is useful information for West Coast weather prediction (Gibson et al., 2020a, GGR'20). This forecast also indicates long-lead confidence about the Baja-Pacific mode transitioning into the negative phase, and the Offshore-California mode remaining negative into week 3. The model is less confident about the phase of the Canadian-Pacific mode beyond day 12. The choice of 70% is based on exploratory analysis demonstrating that a lower threshold (50%–60%) leads to lower skill (Figure S3a in Supporting Information S1) and a higher threshold (80%) is too rarely met in weeks 3–4 (Figure S3b in Supporting Information S1).

3.3. Input to the Statistical Model

For a given forecast and lead time, each of the BP, AP, CP, and OC modes are filtered for uncertainty and classified as "positive," "negative," or "unknown." This filtered dynamical model information is then used as input into the statistical model, which is a conditional probability model based on historical outcomes. The model is represented as

$$P(X|BP, AP, CP, OC) = \frac{P(X, BP, AP, CP, OC)}{P(BP, AP, CP, OC)} \quad (1)$$

Here, P(X) is the historical conditional probability of X, and BP, AP, CP, and OC represent the phase of the four NP4 modes as forecast by the dynamical model, which can be positive, negative, or unknown. The weather impact X is any weather outcome driven by atmospheric circulation in this region. To determine P(X), we identify days in the historical record when the same mode phase combination occurred, and then compile observed outcomes on those days to quantify the historical probability of different weather impacts for different locations. The statistical model will vary in complexity for each forecast and lead time. Some forecasts will have four modes available as predictors while other forecasts will only use one, two, or three modes due to uncertainty in the remaining modes. Sometimes there can be uncertainty about all modes; in those cases, we are not able to issue a forecast. We focus on predicting AR landfalls at different West Coast latitudes (32.5–55°N), SAWs in Southern California, and heat extremes over the Sierra Nevada, Coastal Southern California, and the San Francisco Bay area. A graphical summary of the hybrid methodology is provided in Figure S4 in Supporting Information S1.

4. Example Realtime Application From WY2022

Figure 1c shows an example of the filtering process for WY2022, highlighting correct, incorrect, and uncertain mode phase forecasts. The filtered forecasts are overwhelmingly correct (red/blue shading) with very little error (yellow shading). However, in weeks 3–4 there is an increasing number of forecasts classified as uncertain (gray shading). Despite information removal by filtering, useful information remains at the longer lead times. For

example, the dynamical model skillfully predicted the negative phase of the Alaskan-Pacific mode 3–4 weeks in advance during December-February, which in physical terms is characterized by a ridge over the Gulf of Alaska (c.f. Figure 1a). In December, the positive phase of the Canadian-Pacific mode was skillfully predicted at 3–4 weeks lead time, which is associated with a trough over British Columbia. There was also skill in forecasting the negative phase of the Offshore-California mode in week 3 during January and weeks 3–4 during February, which is associated with a ridge offshore from California. This persistent ridge during January-February was responsible for the extremely dry conditions that occurred in California and contributed to the continuation of the drought during WY2022 (Figure S5 in Supporting Information S1; Hatchett et al., 2023).

For comparison, Figure S6 in Supporting Information S1 shows the verified mode phase forecasts using the ensemble mean with no filtering. Most of the forecasts for weeks 1–2 are correct (red/blue shading) but at longer lead times, an increasing number of forecasts are incorrect (yellow). The filtering process (Figure 1c) was effective at removing many of these incorrect forecasts and re-classifying them as uncertain. The result is that most forecasts are correct after filtering (Figure 1c). To summarize, applying uncertainty filtering in this hybrid dynamical-statistical framework decreases the number of incorrect forecasts obtained by the raw dynamical model. Although the number of forecasts issued in this hybrid framework are lower, those issued can be interpreted with much higher confidence and reliability.

Figure 1d shows an example of real-time AR landfall probability forecasts from WY2022 for four coastal regions: British Columbia (50–55°N), Pacific Northwest (42.5–47.5°N), Northern California (37.5–40°N), and Southern California (32.5–35°N), where the input from Figure 1c was used to predict AR landfalls shown in Figure 1d. The top panels show the observed coastal AR IVT from 1 November to 28 February, and the bottom panels show the AR probability forecasts using the hybrid model. The forecasts represent the observed AR landfall behavior that occurred in WY2022 very well with 3 weeks lead, and with some skill at 4 weeks. In general, above (below) normal AR forecasts were issued for days when AR landfalls occurred (did not occur). For example, the mid-winter dry spell during December in British Columbia was correctly forecast, along with the wet periods that preceded and followed. Also notable are wet December conditions followed by a dry January-February along the US West coast, which was well represented in the forecasts for the Pacific Northwest and Northern/Southern California.

5. Hindcast Skill Assessment Methodology

5.1. Mode Phase Forecasts

As a first skill measure, we evaluate if the phases of the NP4 modes were accurately predicted by the ECMWF model over the 2001–2020 hindcast record, and if the consensus filtering methodology is effective at removing error. We use the forecast accuracy (A), defined as the fraction of forecasts in the correct category as $A = \frac{1}{N}\sum_{i=1}^{K} n(F_i O_i)$, where N is the number of forecasts, n is the number of forecasts in the ith category, and K is the number of categories (here K = 2 for the positive and negative phase). We first calculate the accuracy of the filtered forecasts (using the 70% consensus criteria). Next, for comparison, we calculate the accuracy of the mode phase forecasts using the ensemble mean without filtering. We also calculate the accuracy of the removed forecasts, which are the forecasts that were issued using the ensemble mean and were removed by our filtering process for not meeting 70% ensemble agreement criteria. Significant skill is determined by comparing the forecast accuracy with the accuracy obtained using random resampling (bootstrapping). Forecasts with accuracy exceeding the 95th percentile of the accuracy calculated by random resampling are considered significantly skillful. We also measured skill using the Heidke Skill Score and received a similar result (not shown).

5.2. Probabilistic Forecasts of Weather Impacts

To evaluate skill of the probabilistic forecasts for heat extremes, SAWs, and ARs, we compare the forecast conditional probability P(X) with the observed frequency for each type of weather outcome. We aggregate these probabilistic forecasts into three categories defined relative to the historical climatology as "low probability," "above normal probability," and "much above normal probability" as: <50% of climatology, 120%–160% of climatology, and >160% of climatology, respectively. We then measure the frequency of a weather event following each forecast category and use random resampling (bootstrapping) to compare the reliability of the forecasts

against the climatological reference. The climatological reference is represented by 1,000 random samples of n observations, where n is the number of forecasts in a category. A "low probability" forecast is considered skillful if the observed frequency of a weather event following these forecasts is low relative to climatology (falls below the 10th percentile of the resampled distribution). The "above normal" and "much above normal" forecasts are considered skillful if the observed frequency exceeds the 90th percentile.

We additionally provide the Brier Skill Score (BSS) as a secondary measure of skill, where the BSS quantifies the improvement of a forecast relative to a reference forecast by comparing the Brier scores of each (Weigel et al., 2007). The Brier score is defined as

$$\mathrm{BS} = \frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{K}} (p_i - o_i)^2,$$

where N is the sample size, K is the number of probability bins, p_i is the forecast probability for the ith bin, and o_i is the corresponding observed frequency. The BSS is defined as

$$\mathrm{BSS} = 1 - \frac{\mathrm{BS}}{\mathrm{BS}_{\mathrm{reference}}},$$

where we use the resampled climatology over the reforecast period as the reference. A result of BSS = 1 indicates a perfect forecast and BSS $\leq$ 0 indicates no skill relative to climatology.

6. Hindcast Skill Assessment Results

6.1. NP4 Modes

The hindcast skill assessment for the NP4 modes is shown in Figure 2a for the filtered forecasts (orange) and the ensemble mean (purple). Additionally, we show the accuracy of the forecasts removed by the filtering process (green). The forecasts using the ensemble mean or the consensus filtering approach are significantly skillful out to 30 days relative to climatology (95% level using random resampling). However, the filtered forecasts improve both accuracy (difference along the y-axis) and lead time (difference along the x-axis). For example, in week 3 for the Alaskan-Pacific mode, accuracy improves by nearly 10% at 16-days lead (see vertical arrow). Also in week 3, the filtered model achieves the same accuracy at a 21-day lead that the ensemble mean achieves at a 16-days lead (see horizontal arrow). The information removed by filtering (green) is skillful overall (usually above or near the 95% significance line) but contains a large amount of error (40%–60% accuracy). This highlights the effectiveness of the filtering method. At shorter lead times the percentage of forecasts removed by filtering is small (e.g., 2%–3% at day 2, values shown as text in plot), while at longer leads the amount of uncertainty and therefore the percentage of forecasts removed by filtering grows substantially (e.g., 54%–67% at day 21). At shorter leads (~week 1) uncertainty often results from a mode phase transition, where ensemble members fall on different sides of the zero line as the mode shifts from one phase to the other (e.g., see Figure 1b, Baja-Pacific mode, day 6), whereas at longer leads (~weeks 3–4) there is more general dispersion that is difficult to diagnose. Overall, the filtering methodology is effective at removing error, and the remaining information is more accurate across lead times after filtering.

6.2. Extreme Heat

Figure 2b shows hindcast skill for heat extremes. In general, "low probability" forecasts are followed by a low frequency of extreme heat occurrence, whereas "above normal" and "much above normal" forecasts are followed by a much higher frequency of occurrence. In weeks 1–3, the forecasts are significantly skillful relative to climatology using the resampling method (i.e., blue markers fall below the 10th percentile of the resampled distribution and orange/red markers fall above the 90th percentile of the resampled distribution). The sign of the BSS (shown along the bottom of Figure 1b as ±) remains positive into or throughout week 3. There is evidence of skill in predicting extreme out-of-season heat into week 4 for the Central Sierra Nevada where the forecasts are generally on the right side of the climatology, but the skill is not reliable (i.e., more data points fall within the 10th–90th percentiles of the resampled distribution and the BSS is mostly negative). The magnitude of the BSS for heat extremes is provided in Figure S7 in Supporting Information S1.

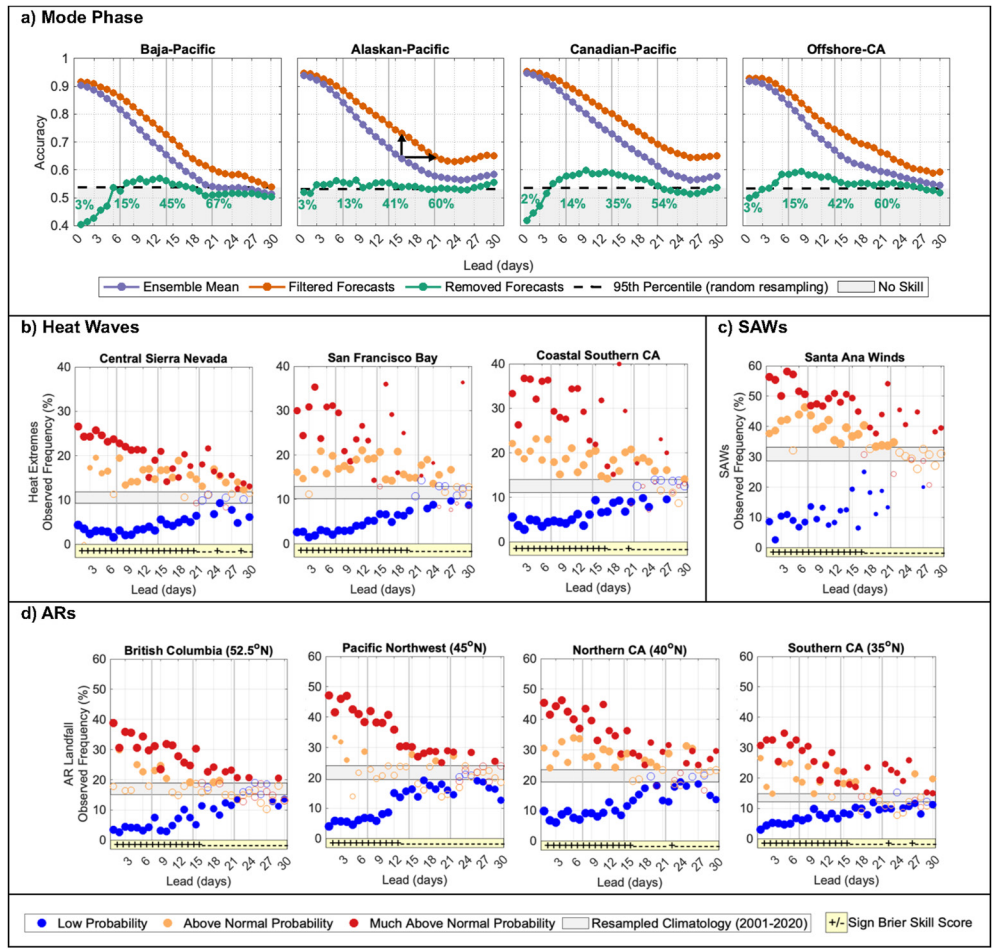

Figure 2. (a) Accuracy of the mode phase forecasts out to 30 days, where purple shows accuracy using the ensemble mean, orange shows accuracy of the filtered forecasts, green shows accuracy of forecasts removed by filtering, the black dashed line gives the 95th percentile of the climatological distribution using random resampling, and gray shading shows the area below the 95th percentile indicating no significant skill. Lines have been smoothed using a 3-day running mean. Black arrows are discussed in the text. (b–d) Hindcast skill assessment of heat waves in three regions, SAWs in Southern California, and ARs at four coastal latitudes, respectively. The *y*-axis shows observed event frequency following three forecast categories: low (blue), above normal (orange), and much above normal (red) probability. The sign of the BSS (±) is shown along the bottom, highlighted in yellow. The gray shaded area in (b–d) gives the 10th–90th percentiles of the resampled distribution over the hindcast period (2001–2020). Red, orange, and blue markers are weighted by the log of the sample size (n = 10 to 451), and filled markers indicate statistically significant skill (90% level using resampling).

6.3. Santa Ana Winds

Figure 2c shows hindcast skill for SAWs over Southern California. The results are similar as for heat extremes, where the outcomes for the three forecast categories are separated and significantly skillful relative to climatology at lead times of 1–3 weeks (using random resampling). The forecast is unreliable at 4 weeks lead, and the BSS becomes negative around day 17.

6.4. Atmospheric Rivers

Figure 2d shows hindcast skill for AR landfalls at four different coastal latitudes, where each location represents the probability that an AR impacted the 2.5° × 2.5° R1 grid cell area centered over the specified location. The hybrid forecasts are skillful at 1–2 weeks lead for all locations. Resolution is poor for British Columbia and the Pacific Northwest as seen by the many orange data points representing the "above normal" forecast category falling within the range of climatology, suggesting that AR landfalls are over-estimated. Interestingly, the more confident "much above normal" forecast category shows higher skill. The resolution is better for Northern and

Southern California where there is evidence of skill relative to climatology into week 3 for all categories (forecasts perform better than climatological resampling, significant at the 90% level). The BSS becomes negative at day 13–17, depending on region.

7. Discussion

We have described and evaluated a new statistical-dynamical hybrid model that produces skillful probabilistic forecasts of heat extremes, Santa Ana winds, and landfalling ARs over California at subseasonal timescales. The methodology is skillful into week 3 for most impacts and regions studied. Week 4 shows evidence of skill relative to climatology for some impacts and regions, though the forecasts lack reliability (BSS < 0) at these longer timescales.

Because we use a hybrid approach, the potential skill is not strictly limited by the dynamical model skill. Near-future improvements are possible through continued development of the statistical model. For example, signals from lower frequency climate teleconnections could be incorporated at longer lead times when the dynamical model is uncertain. Impacts of climate-scale teleconnections on offshore atmospheric ridges (Gibson et al., 2020a) as well as the NP4 modes (GGR'20) have been identified, and these relationships could be incorporated to extend skill and lead time using this hybrid approach. The skill could be sensitive to thresholds or definitions used for SAWs or ARs. Possible improvements could be made by focusing on the more extreme or persistent events. For heat waves, non-stationarity due to climate change will become an important consideration since we rely on historical probabilities in our methodology.

We focused on three impact types (heat extremes, SAWs, and ARs). However, other applications are possible including cold extremes and hot/cold temperature anomalies more generally, as well as conditions driving mid-winter snowmelt (Hatchett et al., 2023), high snow level storms (Shulgina et al., 2023) with rain-on-snow (Heggli et al., 2022), and other decision-specific variables not represented in dynamical forecasts. The offshore wind forecasts—essential ingredients for Southern California's wildfires—focused on Santa Ana winds, which were amenable to this analysis given the SAW record of Guzman-Morales and Gershunov (2019). Related applications could include Diablo and Sundowner winds of Northern and Central California; however, the seasonality of these winds requires an extension of the methodology into earlier fall and later spring (Abatzoglou et al., 2021). Identifying relationships between the GGR'22 weather regimes and observed live/dead fuel moisture could inform long-lead significant fire potential models and improving situational awareness using this hybrid approach.

Predictability of extreme precipitation in a probabilistic sense could also be explored. This is especially important as both extreme winter precipitation and early winter wildfires may become more common (Cayan et al., 2022; Gershunov et al., 2019), raising the possibility of compound extreme events such as short-duration high-intensity rainfall, which can cause devastating post-fire debris flows (Oakley et al., 2017, 2018a) and landslides (Oakley et al., 2018b; Rengers et al., 2020), rain-on-snow flooding (Haleakala et al., 2023), as well as other precipitation patterns driving mass movements such as avalanches (Hatchett et al., 2017). Improving lead time to prepare for these types of events and likelihood of occurrence is crucial to prevent loss of life and mitigate damage to property (Oakley et al., 2023). This approach could be applied in climate change studies, where changes to atmospheric circulation identified in global climate models along with thermodynamic responses could be linked to future impacts (Michaelis et al., 2022; Rhoades et al., 2023). We envision these results and future updates will form the basis for real-time forecast tools with possible applications for early warning systems and decision support across many sectors including water resources, public health, land, and fire management in a varying and changing climate.

Data Availability Statement

The NP4 data set is from Guirguis et al. (2020b). The Santa Ana winds regional index is from Guzman-Morales and Gershunov (2019). The SIO-R1 AR catalog is from Gershunov et al. (2017). The gridMET temperature data set is from Abatzoglou (2013). The NCEP/NCAR reanalysis is from Kalnay et al. (1996). The ECMWF hindcast data set is from Vitart et al. (2017).

Acknowledgments
This research was funded by the U.S. Department of the Interior via the Bureau of Reclamation (USBR-R15AC00003) and the California Department of Water Resources (4600010378 UCOP2-11). This study also contributes to the Regional Integrated Sciences and Assessments (RISA) California–Nevada Climate Applications Program, the International Research Applications Program of the National Oceanic and Atmospheric Administration (NA17OAR4310284), the Southwest Climate Adaptation Science Center (G18AC00320), and the National Science Foundation (NSF CoPe award no. 2209058). We thank two anonymous reviewers for their helpful feedback during the review process.

References

Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling [Dataset]. *International Journal of Climatology*, *33*(1), 121–131. https://doi.org/10.1002/joc.3413

Abatzoglou, J. T., Hatchett, B. J., Fox-Hughes, P., Gershunov, A., & Nauslar, N. J. (2021). Global climatology of synoptically-forced downslope winds. *International Journal of Climatology*, *41*(1), 31–50. https://doi.org/10.1002/joc.6607

Aguilera, R., Corringham, T., Gershunov, A., Leibel, S., & Benmarhnia, T. (2021). Fine particles in wildfire smoke and pediatric respiratory health in California. *Pediatrics*, *147*(4), e2020027128. https://doi.org/10.1542/peds.2020-027128

Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, *525*(7567), 47–55. https://doi.org/10.1038/nature14956

Cayan, D. R., DeHaan, L., Gershunov, A., Guzman-Morales, J., Keeley, J. E., Mumford, J., & Syphard, A. D. (2022). Autumn precipitation – The competition with Santa Ana winds in determining fire outcomes in Southern California. *International Journal of Wildland Fire*, *31*(11), 1056–1067. https://doi.org/10.1071/WF22065

Corringham, T. W., McCarthy, J., Shulgina, T., Gershunov, A., Cayan, D. R., & Ralph, F. M. (2022). Climate change contributions to future atmospheric river flood damages in the western United States. *Scientific Reports*, *12*, 13747. https://doi.org/10.1038/s41598-022-15474-2

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

DeFlorio, M. J., Ralph, F. M., Waliser, D. E., Jones, J., & Anderson, M. L. (2021). Better subseasonal-to-seasonal forecasts for water management. *EOS*, *102*. https://doi.org/10.1029/2021EO159749

DeFlorio, M. J., Waliser, D. E., Ralph, F. M., Guan, B., Goodman, A., Gibson, P. B., et al. (2019). Experimental subseasonal-to-seasonal (S2S) forecasting of atmospheric rivers over the western United States. *Journal of Geophysical Research: Atmosphere*, *124*(21), 11242–11265. https://doi.org/10.1029/2019jd031200

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

Dettinger, M. D., Ralph, F. M., Das, T., Neiman, P. J., & Cayan, D. R. (2011). Atmospheric rivers, floods and the water resources of California. *Water*, *3*(2), 445–478. https://doi.org/10.3390/w3020445

Ferranti, L., Corti, S., & Janousek, M. (2015). Flow-dependent verification of the ECMWF ensemble over the Euro-Atlantic sector. *Quarterly Journal of the Royal Meteorological Society*, *141*(688), 916–924. https://doi.org/10.1002/qj.2411

Gershunov, A., Guzman-Morales, J., Hatchett, B., Aguilera, R., Shulgina, T., Guirguis, K., et al. (2021). Hot and cold flavors of southern California's Santa Ana winds: Their causes, trends, and links with wildfire. *Climate Dynamics*, *57*(7–8), 2233–2248. https://doi.org/10.1007/s00382-021-05802-z

Gershunov, A., Shulgina, T., Ralph, F. M., Lavers, D., & Rutz, J. J. (2017). Assessing the climate-scale variability of atmospheric rivers affecting the west coast of North America [Dataset]. Geophysical Research Letters, *44*(15), 7900–7908. https://doi.org/10.1002/2017gl074175

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

Gibson, P. B., Waliser, D. E., Goodman, A., DeFlorio, M. J., Delle Monache, L., & Molod, A. (2020b). Subseasonal-to-seasonal hindcast skill assessment of ridging events related to drought over the Western United States. *Journal of Geophysical Research: Atmosphere*, *125*(22), e2020JD033655. https://doi.org/10.1029/2020jd033655

Gibson, P. B., Waliser, D. E., Guan, B., DeFlorio, M. J., Ralph, F. M., & Swain, D. L. (2020a). Ridging associated with drought across the western and Southwestern United States: Characteristics, trends and predictability sources. *Journal of Climate*, *33*(7), 2485–2508. https://doi.org/10.1175/jcli-d-19-0439.1

Guirguis, K., Gershunov, A., Clemesha, R. E. S., Shulgina, T., Subramanian, A. C., & Ralph, F. M. (2018). Circulation drivers of atmospheric rivers at the North American West Coast. *Geophysical Research Letters*, *45*(22), 12–576. https://doi.org/10.1029/2018gl079249

Guirguis, K., Gershunov, A., DeFlorio, M. J., Shulgina, T., Delle Monache, L., Subramanian, A. C., et al. (2020a). Four atmospheric circulation regimes over the North Pacific and their relationship to California precipitation on daily to seasonal timescales. *Geophysical Research Letters*, *47*(16), e2020GL087609. https://doi.org/10.1029/2020gl087609

Guirguis, K., Gershunov, A., DeFlorio, M. J., Shulgina, T., Delle Monache, L., Subramanian, A. C., et al. (2020b). Data from: Four atmospheric circulation regimes over the North Pacific and their relationship to California precipitation on daily to seasonal timescales [Dataset]. UC San Diego Library Digital Collections. https://doi.org/10.6075/J0154FJJ

Guirguis, K., Gershunov, A., Hatchett, B., Shulgina, T., DeFlorio, M. J., Subramanian, A. C., et al. (2022). Winter wet–dry weather patterns driving atmospheric rivers and Santa Ana winds provide evidence for increasing wildfire hazard in California. *Climate Dynamics*, *60*(5–6), 1729–1749. https://doi.org/10.1007/s00382-022-06361-7

Guzman-Morales, J., & Gershunov, A. (2019). Climate change suppresses Santa Ana Winds of Southern California and sharpens their seasonality. [Dataset]. Geophysical Research Letters, *46*(5), 2772–2780. https://doi.org/10.1029/2018gl080261

Guzman-Morales, J., Gershunov, A., Theiss, J., Li, H., & Cayan, D. R. (2016). Santa Ana Winds of southern California: Their climatology, extremes, and behavior spanning six and a half decades. *Geophysical Research Letters*, *43*(6), 2827–2834. https://doi.org/10.1002/2016gl067887

Haleakala, K., Brandt, W. T., Hatchett, B. J., Li, D., Lettenmaier, D. P., & Gebremichael, M. (2023). Watershed memory amplified the Oroville rain-on-snow flood of February 2017. *PNAS Nexus*, *2*, 1–15. https://doi.org/10.1093/pnasnexus/pgac295

Hatchett, B. J., Burak, S., Rutz, J. J., Oakley, N. S., Bair, E. H., & Kaplan, M. L. (2017). Avalanche fatalities during atmospheric river events in the western United States. *Journal of Hydrometeorology*, *18*(5), 1359–1374. https://doi.org/10.1175/jhm-d-16-0219.1

Hatchett, B. J., Koshkin, A. L., Guirguis, K., Rittger, K., Nolin, A. W., Heggli, A., et al. (2023). Midwinter dry spells amplify post-fire snowpack decline. *Geophysical Research Letters*, *50*(3), e2022GL101235. https://doi.org/10.1029/2022gl101235

Heggli, A., Hatchett, B., Schwartz, A., Bardsley, T., & Hand, E. (2022). Toward snowpack runoff decision support. *iScience*, *25*(5), 104240. https://doi.org/10.1016/j.isci.2022.104240

Huang, X., & Swain, D. (2022). Climate change is increasing the risk of a California megafloood. *Science Advances*, *8*(32), eabq0995. https://doi.org/10.1126/sciadv.abq0995

Hughes, M., & Hall, A. (2010). Local and synoptic mechanisms causing Southern California's Santa Ana winds. *Climate Dynamics*, *34*(6), 847–857. https://doi.org/10.1007/s00382-009-0650-4

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., et al. (1996). The NCEP/NCAR 40-year reanalysis project [Dataset]. *Bulletin of the American Meteorological Society*, *77*(3), 437–471. https://doi.org/10.1175/1520-0477(1996)077<0437:tnyrp>2.0.co;2

Kanamaru, H., & Kanamitsu, M. (2007). Fifty-seven-year California reanalysis downscaling at 10 km (CaRD10). Part II: Comparison with North American regional reanalysis. *Journal of Climate*, *20*(22), 5572–5592. https://doi.org/10.1175/2007jcli1522.1

Michaelis, A. C., Gershunov, A., Weyant, A., Fish, M. A., Shulgina, T., & Ralph, F. M. (2022). Atmospheric river precipitation enhanced by climate change: A case study of the storm that contributed to California's Oroville Dam crisis. *Earth's Future*, *10*(3), e2021EF002537. https://doi.org/10.1029/2021ef002537

Oakley, N., Cannon, F., Munroe, R., Lancaster, J., Gomberg, D., & Ralph, F. M. (2018a). Brief communication: Meteorological and climatological conditions associated with the 9 January 2018 post-fire debris flows in Montecito and Carpentaria, California, USA. *Natural Hazards and Earth System Sciences*, *18*(11), 3037–3043. https://doi.org/10.5194/nhess-18-3037-2018

Oakley, N. S., Lancaster, J. T., Hatchett, B. J., Stock, J., Ralph, F. M., Roj, S., & Lukashov, S. (2018b). A 22-year climatology of cool season hourly precipitation thresholds conducive to shallow landslides in California. *Earth Interactions*, *22*(14), 1–35. https://doi.org/10.1175/ei-d-17-0029.1

Oakley, N. S., Lancaster, J. T., Kaplan, M. L., & Ralph, F. M. (2017). Synoptic conditions associated with cool season post-fire debris flows in the Transverse Ranges of southern California. *Natural Hazards*, *88*(1), 327–354. https://doi.org/10.1007/s11069-017-2867-6

Oakley, N. S., Liu, T., McGuire, L. A., Simpson, M., Hatchett, B. J., Tardy, A., et al. (2023). Toward probabilistic post-fire debris-flow hazard decision support. *Bulletin America Meteorology Social*, *104*(9), E1587–E1605. https://doi.org/10.1175/BAMS-D-22-0188.1

Palmer, T. (2017). The primacy of doubt: Evolution of numerical weather prediction from determinism to probability. *Journal of Advances in Modeling Earth Systems*, *9*(2), 730–734. https://doi.org/10.1002/2017ms000999

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

Ralph, F. M., Neiman, P. J., Wick, G. A., Gutman, S. I., Dettinger, M. D., Cayan, D. R., & White, A. B. (2006). Flooding on California's Russian River: Role of atmospheric rivers. *Geophysical Research Letters*, *33*(13), L13801. https://doi.org/10.1029/2006GL026689

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

Rengers, F. K., McGuire, L. A., Oakley, N. S., Kean, J. W., Staley, D. M., & Tang, H. (2020). Landslides after wildfire: Initiation, magnitude, and mobility. *Landslides*, *17*(11), 2631–2641. https://doi.org/10.1007/s10346-020-01506-3

Rhoades, A. M., Zarzycki, C. M., Inda Díaz, H. A., Ombadi, M., Pasquier, U., Srivastava, A., et al. (2023). *Recreating the California New Year's flood event of 1997 in a regionally refined Earth system model*. ESS Open Archive.

Robertson, A. W., Vigaud, N., Yuan, J., & Tippett, M. K. (2020). Toward identifying subseasonal forecasts of opportunity using North American weather regimes. *Monthly Weather Review*, *148*(5), 1861–1875. https://doi.org/10.1175/mwr-d-19-0285.1

Rutz, J. J., Shields, C. A., Lora, J. M., Payne, A. E., Guan, B., Ullrich, P., et al (2019). The atmospheric river tracking method intercomparison project (ARTMIP): Quantifying uncertainties in atmospheric river climatology. *Journal of Geophysical Research: Atmospheres*, *124*(24), 13777–13802. https://doi.org/10.1029/2019JD030936

Schwarz, L., Malig, B. J., Guzman-Morales, J., Guirguis, K., Gershunov, A., Basu, R., & Benmarhnia, T. (2020). The health burden of fall, winter and spring heat waves in Southern California and contribution of Santa Ana Winds. *Environmental Research Letters*, *15*(5), 054017. https://doi.org/10.1088/1748-9326/ab7f0e

Shulgina, T., Gershunov, A., Hatchett, B. J., Guirguis, K., Subramanian, A. C., Margulis, S. A., et al. (2023). Observed and projected changes in snow accumulation and snowline in California's snowy mountains. *Climate Dynamics*, *61*(9–10), 4809–4824. https://doi.org/10.1007/s00382-023-06776-w

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., et al. (2017). The Subseasonal to seasonal (S2S) prediction project database. [Dataset]. Bulletin America Meteorology Social, *98*(1), 163–173. https://doi.org/10.1175/bams-d-16-0017.1

Weigel, A. P., Liniger, M. A., & Appenzeller, C. (2007). The discrete brier and ranked probability skill scores. *Monthly Weather Review*, *135*(1), 118–124. https://doi.org/10.1175/MWR3280.1