Extended-Range Probabilistic Fire-Weather Forecasting Based on Ensemble Model Output Statistics and Ensemble Copula Coupling

ROCHELLE P. WORSNOP AND MICHAEL SCHEUERER

Cooperative Institute for Research in the Environmental Sciences, University of Colorado Boulder, and NOAA/Earth System Research Laboratory, Physical Sciences Division, Boulder, Colorado

THOMAS M. HAMILL

NOAA/Earth System Research Laboratory, Physical Sciences Division, Boulder, Colorado

(Manuscript received 26 June 2019, in final form 6 September 2019)

ABSTRACT

Probabilistic fire-weather forecasts provide pertinent information to assess fire behavior and danger of current or potential fires. Operational fire-weather guidance is provided for lead times fewer than seven days, with most products only providing day 1–3 outlooks. Extended-range forecasts can aid in decisions regarding placement of in- and out-of-state resources, prescribed burns, and overall preparedness levels. We demonstrate how ensemble model output statistics and ensemble copula coupling (ECC) postprocessing methods can be used to provide locally calibrated and spatially coherent probabilistic forecasts of the hot–dry–windy index (and its components). The univariate postprocessing fits the truncated normal distribution to data transformed with a flexible selection of power exponents. Forecast scenarios are generated via the ECC-Q variation, which maintains their spatial and temporal coherence by reordering samples from the univariate distributions according to ranks of the raw ensemble. A total of 20 years of ECMWF reforecasts and ERA-Interim reanalysis data over the continental United States are used. Skill of the forecasts is quantified with the continuous ranked probability score using benchmarks of raw and climatological forecasts. Results show postprocessing is beneficial during all seasons over CONUS out to two weeks. Forecast skill relative to climatological forecasts depends on the atmospheric variable, season, location, and lead time, where winter (summer) generally provides the most (least) skill at the longest lead times. Additional improvements of forecast skill can be achieved by aggregating forecast days. Illustrations of these postprocessed forecasts are explored for a past fire event.

1. Introduction

Fire-weather forecasting for decision-making is a complex process that involves knowledge of atmospheric conditions, local topography, and state of vegetation/fuels. Wildfires are a global phenomena that occur often in Australia, Brazil, Canada, China, Greece, Portugal, Russia, South Africa, United States, among many other countries (Shvidenko and Schepaschenko 2013; McGee et al. 2015; Sharples et al. 2016; Palaiologou et al. 2018; Moreira and Pe'er 2018; Tedim et al. 2018; Eugenio et al. 2019). Although wildfires are an integral component to the natural ecosystem (Hutto 2008; Bowman et al. 2011), they lead to great loss of lives, property, and vulnerable habitats. Between 2008 and 2018, wildfires burned over 30.4 million ha of U.S. lands (NIFC 2019) and more than 29.1 million ha of land in Canada from 2007 to 2017 (NRCAN 2019). Wildfires also generate toxic emissions and smoke that can disperse hundreds or even thousands of kilometers from a fire affecting transportation visibility and human health (Reid et al. 2016; Black et al. 2017). In these ways, the effects of wildfires cross geographical and political boundaries making them a threat to much of the global population.

The effects of wildfires are expected to increase as changes in climate lead to longer fire seasons, more frequent wildfires and lightning discharges, and larger

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/MWR-D-19-0217.s1.

Corresponding author: Rochelle P. Worsnop, rochelle.worsnop@noaa.gov

DOI: 10.1175/MWR-D-19-0217.1

fire-burned areas (Flannigan et al. 2000, 2013; Westerling et al. 2006). Additionally, recent demographic shifts for new residential development at the wildland–urban interface further bolsters the threat that wildfires can have on the human population (Radeloff et al. 2018). Fire-weather forecasting may not prevent wildfires from starting, but a skillful forecast with ample lead time could help prevent loss of life and property by informing the decision-making process.

Information about the uncertainty of a forecast can provide decision-makers with a range of possible outcomes and the amount of confidence associated with a particular event (Krzysztofowicz 2001), which is valuable for deciding if, when, and how many precautionary measures should be taken. For fire- and land-management agencies, these decisions could include opting for or against performing a prescribed burn, issuing public alerts or restrictions, prepositioning nonlocal resources for reinforcement, and assessing overall preparedness levels.

The fire-forecasting community typically relies on deterministic forecasts (also called point forecasts), but a shift to probabilistic forecasts of fire-weather indices is beginning to appear in the literature (Di Giuseppe et al. 2016; Srock et al. 2018). Probabilistic forecasts aim to predict the uncertainty of a quantity or event of interest in the form of full predictive probability distributions (Gneiting and Katzfuss 2014) rather than single-valued or point forecasts. These aforementioned studies calculated fire indices with different forecast ensemble members out to a maximum of 6 or 10 days ahead. However, these forecasts are not postprocessed to account for systematic biases in the model and loss of skill with longer lead times. Therefore, we focus this paper on the generation and validation of postprocessed fire-weather forecasts in the extended range (defined here as the time range between 3 and 14 days).

Many meteorological centers around the world are now running numerical ensemble prediction systems to generate probabilistic forecasts spanning lead times of a few hours to several months. Typically, an NWP ensemble is generated by making slight modifications to the initial conditions, stochastically perturbing the model physics, and sometimes by running separate ensemble members with different physics schemes and/or dynamic cores to create a blend of models (Buizza et al. 1999; World Meteorological Organization (WMO) 2012). Although ensemble forecasts are available for lead times in the extended range and beyond, it can be difficult to get skillful predictions at subseasonal to monthly time scales (Hudson et al. 2011; White et al. 2017). At these scales, forecasts not only include influences from initial model conditions but also conditions that evolve on slower time scales such as soil moisture and sea surface temperatures (White et al. 2017).

Advances in data assimilation techniques, model initialization, physics parameterizations, and spatial and temporal resolution over the last decade now allow researchers to explore forecasts in the middle range between short-range weather forecasts and climate forecasts, known as subseasonal-to-seasonal forecasts (herein, we focus on predictions in the extended range, which include subseasonal forecasts out to two weeks). Even still, unmodified or raw ensemble forecasts are often underdispersive (Hamill and Colucci 1997; Raftery et al. 2005; Stauffer et al. 2017); they do not capture the full range of forecast scenarios and therefore yield insufficient estimates of the total uncertainty associated with a forecast. Raw ensemble forecasts are also often biased as a result of insufficient model resolution, especially over regions of complex terrain (Stauffer et al. 2017) and from deficiencies in physical model assumptions and data assimilation procedures (Buizza et al. 2005). These errors become even more problematic at longer lead times as the forecast moves farther away from initial constraints provided by the data assimilation system and as small errors in the initial conditions compound with each forecast integration (Hamill and Colucci 1997).

The presence of these errors necessitates statistical postprocessing of the raw ensemble to yield calibrated and sharp probabilistic forecasts, the overall goal of probabilistic forecasting (Gneiting et al. 2007). Calibrated refers to the statistical consistency between forecasts from the predictive distribution and the corresponding verifying observations whereas sharpness refers to the spread of the forecast ensemble (Gneiting et al. 2007). A well-calibrated forecast would suggest a probability of an event that is consistent with the average proportion of time that the event is observed, while sharpness implies that the forecast is as specific (e.g., short prediction intervals, event probabilities close to zero or one) as possible. Herein, we focus on postprocessing methods that use raw ensemble forecasts to generate full predictive cumulative distribution functions (CDFs) for any grid location on the model domain which are then turned into finite calibrated ensemble forecasts.

Some univariate statistical postprocessing methods that are commonly used on meteorological ensemble forecasts are Bayesian model averaging (Raftery et al. 2005), and nonhomogeneous Gaussian regression, also referred to as ensemble model output statistics (EMOS, Gneiting et al. 2005). These approaches use historical forecast–observation pairs as training data to fit regression models. With these data pairs, the goal is then to identify and correct biases in the raw ensemble

by generating predictive distributions that do not suffer from the same deficiencies. These methods define the parameters of the predictive distribution for each lead time and location. However, forecasters often want to know how an event will unfold over several lead times or locations, which requires multivariate statistical postprocessing. Multivariate postprocessing accounts for the spatial and temporal correlations between lead times and locations while preserving the initial skill gained by the univariate postprocessing steps.

Defining the relationship between the independent univariate predictive distributions to generate a multivariate ensemble can be done with a nonparametric sampling-reordering approach. This approach samples from a marginal predictive distribution and then reorders the samples at each lead time according to the rank structure of a specified dependence template. Some dependence templates rely on past observations (e.g., Schaake shuffle and adaptions thereof, Schefzik 2016; Scheuerer et al. 2017; Worsnop et al. 2018), while others rely on the raw forecast ensemble (ECC, Schefzik et al. 2013).

In this paper, we use 11-member ensemble reforecasts from an operational version of the ECMWF model and reanalysis data from ERA-Interim (data described in section 3) to generate and validate extended-range probabilistic forecasts of the hot–dry–windy index (HDWI) (described in section 2), which relies solely on values of temperature, moisture, and wind speed. We use the EMOS approach (described in section 4) to first generate calibrated and sharp univariate predictive distributions of meteorological variables that are used to calculate the HDWI. We then apply a variant of the ECC method (described in section 5) to produce coherent multivariate ensemble forecasts of the HDWI and its components out to two weeks. Performance of the postprocessing methods against raw and climatological forecasts is evaluated with estimates of the skill of the continuous rank probability scores (CRPSS) in section 6. Examples of how the postprocessed forecasts could be used for forecasting is discussed in section 7 and are followed with conclusions and a discussion of the utility of these methods for operational fire-weather forecasting in section 8.

2. Fire-weather metric: Hot–dry–windy index (HDWI)

HDWI and the corresponding HDWI climatology were developed by Srock et al. (2018) and McDonald et al. (2018), respectively based on the understanding of how the atmosphere physically affects wildfires. Although fuels, ignition agents, weather/climate, and humans all affect wildfire activity, weather/climate is the most important natural factor that influences daily wildfire danger (Flannigan et al. 2005). Wind, temperature, and moisture are the atmospheric components that most influence the amount of fire-burned area (Flannigan and Harrington 1988; McDonald et al. 2018). The HDWI is a purely meteorological index that includes these three components and was therefore selected as our predictand. Of course, even in the most fire-primed conditions, a fire will not start unless fuels are ready and available to burn and an ignition occurs. The intended purpose of the HDWI is to bring awareness of days/locations that may encounter a lesser or greater potential for fire based on the weather conditions. Complementing it with another index that encompasses fuel information would be helpful, and will be addressed in a separate study.

The HDWI (1) is the product of the maximum wind speed $U(\mathrm{m\,s^{-1}})$ in a layer between the surface and 500 m above ground level (AGL) and the maximum vapor pressure deficit VPD(Pa), in that same layer. Hereafter, we denote the maximum U and VPD found within this *vertical* layer with a tilde accent. The subscript "max" indicates that the HDWI is calculated with the daily maximum of $\tilde{U}$ and $\widetilde{\mathrm{VPD}}$ over a 24-h period (calculation of the daily maxima is discussed in section 5b). With these definitions we obtain the following:

$$\mathrm{HDWI} = \tilde{U}_{\max} \times \widetilde{\mathrm{VPD}}_{\max}(T, q). \tag{1}$$

The initial calculation of $\widetilde{\mathrm{VPD}}$ for the HDWI (before taking the daily maximum) is a function of temperature T brought down adiabatically to the surface and specific humidity q, which is conserved as the parcel is lowered to the surface. VPD is defined as the difference between the saturation vapor pressure e_s and the vapor pressure e. The VPD, in contrast to the ratio of e_s and e (i.e., relative humidity), gives a better depiction of the moisture evaporation rate that results from hot and dry conditions and therefore the evaporative potential of fuels (Leighly 1937; Thornthwaite 1940; Simard 1968). The daily maxima in (1) are only calculated in section 5b, until then we focus on quantities $\tilde{U}$ and $\widetilde{\mathrm{VPD}}$ that are specific to particular times of the day.

3. Data description

a. Ensemble reforecasts

Reforecasts are based on one model version that is generally rerun to generate past daily forecasts over a period of decades. For our analysis, we use medium-range reforecasts of an 11-member ensemble produced from the Cycle43r3 version of ECMWF's operational

model (ECMWF 2019). Initial conditions of the reforecasts are defined by ERA-Interim analyses, while the ensemble members are created by perturbing the initial states and model physics of the control run (Buizza et al. 1999). ECMWF ran the Cycle43r3 version so that it produced reforecasts for dates corresponding to every Monday and Thursday between 13 July 2017 and 4 June 2018 with the 0000 UTC initial conditions. For each of those dates, the model was rerun to generate daily ensemble reforecasts for the previous 20 years (e.g., because 4 June 2018 fell on a Thursday, daily reforecasts for 4 June over the previous 20 years (1998–2017) were generated). We downloaded 12-hourly reforecasts of weather variables over CONUS at a regular latitude–longitude grid of 0.75° (~80 km) from the ECMWF MARS archive system to match the resolution of the reanalysis data. Since we are postprocessing forecasts out to week two, we downloaded data for each date and 0000 UTC initialization out to lead times of 360 h. We use the forecasted variables in Table 1 to calculate the maximum VPD and the maximum U in a 500-m layer above the surface before interpolating those values at each location to match up with locations on the reanalysis grid (see section 3b) via conservative regridding (Jones 1999).

b. Reanalysis

We use the same meteorological variables in Table 1 from reanalysis data to train and verify the postprocessed HDWI forecasts and forecasts of its components. Reanalysis data are taken as the truth, because they are a combination of the NWP model and quality-controlled observations from weather stations, ships, buoys, satellites, etc. through data assimilation techniques (Dee et al. 2011). The major advantage of reanalysis data is that they are available at every grid point and integration of the NWP model, making them ideal for comparison with reforecasts.

For the analysis herein, we use regular latitude–longitude gridded data output from the global ERA-Interim (ERA-I) reanalysis system of the ECMWF (Dee et al. 2011). ERA-I data are available from 1979 to near–real time and are output every 3 h. The data have a spatial resolution of ~80 km (0.75°). We use data every 6 h starting at 0000 UTC from 13 July 1997 (earliest year associated with our reforecast dataset) to 4 June 2017 (latest year associated with our reforecast dataset).

4. Forecast calibration via univariate postprocessing

The gray-shaded boxes (steps 1–9) in Fig. 1 show a conceptual diagram of the univariate postprocessing methods outlined throughout the subsections below.

TABLE 1. Model variables output as ECMWF reforecasts and used for the calculation of and postprocessing of the hot–dry–windy index forecasts.

Model level	Variables
Surface	2-m dewpoint temperature, 2-m temperature, 10-m u-wind speed component, 10-m v-wind speed component, geopotential, mean sea level pressure
Pressure	Geopotential, specific humidity, temperature, u-wind speed component, v-wind speed component

a. Ensemble model output statistics (EMOS)

We postprocess forecasts of the HDWI components (i.e., $\tilde{U}$ and $\widetilde{\text{VPD}}$) at each forecast lead time and location on the reanalysis grid using the EMOS approach (Gneiting et al. 2005). The EMOS approach fits a probability distribution model to the raw ensemble model's output statistics. The result is a calibrated and sharp predictive distribution function of a continuous weather variable that is corrected for forecast biases and ensemble dispersion errors. Linear regression equations are used to define the first and second moments of the distributions. EMOS, originally implemented by Gneiting et al. (2005) to produce probabilistic forecasts from a Gaussian predictive PDF, has since been extended for the truncated normal, gamma, and truncated logistic distributions (Thorarinsdottir and Gneiting 2010; Scheuerer and Möller 2015). The truncated distributions have a lower bound set to zero, which makes them, along with the gamma distribution, suitable for use with nonnegative quantities like $\tilde{U}$ and $\widetilde{\text{VPD}}$.

Since the ensemble members of the ECMWF reforecasts are created from random perturbations to the initial conditions, they are considered exchangeable (i.e., they produce equally likely scenarios of a future state and are not systematically distinguishable from one another; Bröcker and Kantz 2011). Because the ensemble members are exchangeable, we use a modified version of the original EMOS multiple linear regression equation. This modification is based solely on the forecasted ensemble mean and variance instead of weighted contributions from the individual members (Gneiting et al. 2005). Through exploratory analysis, we found that for most of CONUS and all seasons and lead times, data transformation in combination with the truncated normal distribution results in a satisfactory representation of the forecast uncertainty about $\tilde{U}$ and $\widetilde{\text{VPD}}$. For this reason, we focus our analysis in the paper on a modified version of the truncated normal distribution (modification is discussed in section 4d) and test

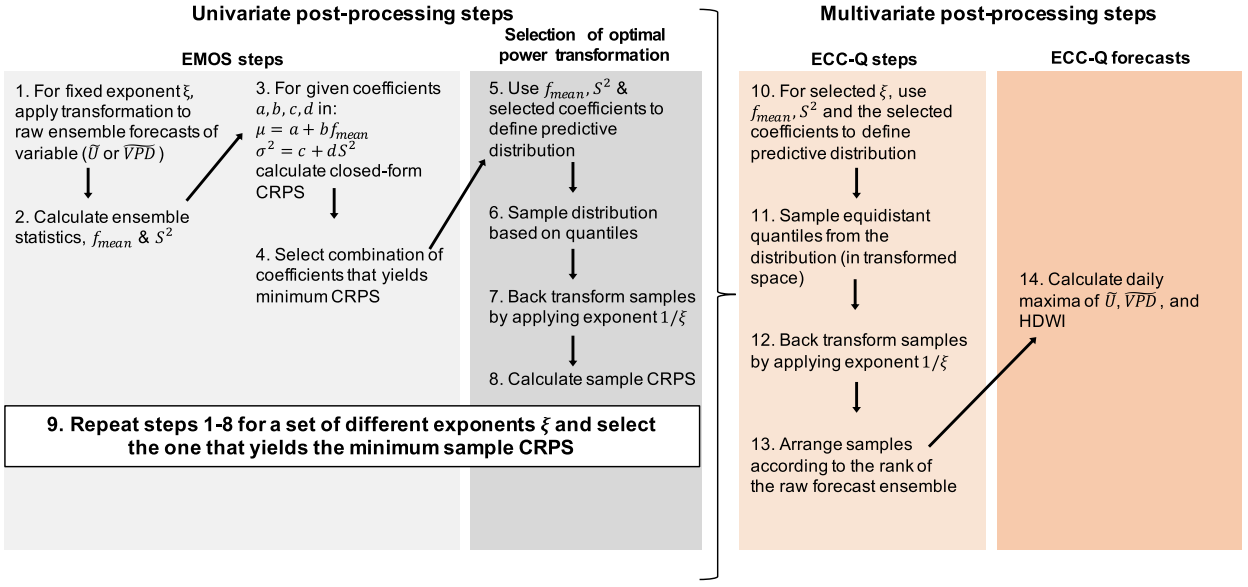

FIG. 1. Schematic of the general concept of the univariate and multivariate postprocessing methods. The steps are numbered in the order in which they are performed. Technical details for each step are discussed in sections 4 and 5. The whole process outlined in this schematic refers to one particular month, year, lead time, and location, except for Step 14 which requires several lead times in a day to calculate the daily maxima.

different power transformations of the data fitted to this distribution (discussed in section 4d).

b. *Training period*

Training data are comprised of past forecast–observation pairs; in our case, we use the 11-member ECMWF reforecasts for past forecasts and ERA-I reanalysis data as a proxy for observations. For each month/year combination in this 20-yr forecast–observation record, the rolling training period consists of all forecasts from that same month and from all years excluding the current year. The training period for each month/year combination has ~152–171 forecast–observation pairs given that there are two ECMWF reforecasts per week and 19 years included in the training period. For each forecast lead time and location, raw forecasts of $\tilde{U}$ and $\widetilde{\text{VPD}}$ from the training data are used to calculate the ensemble statistics used in the EMOS regression equation described next.

c. *EMOS model fitting*

For a given power transformation (if applicable) and an m-member ensemble of exchangeable forecasts $f_1, \ldots, f_m$, estimates of the EMOS regression coefficients a, b, c, and d define the distribution mean μ and distribution variance σ^2 of the truncated normal predictive distribution $\mathcal{N}_0(\mu, \sigma^2)$. The moments of the distribution are defined as

$$\mu = a + bf_{\text{mean}} \quad \text{and} \quad \sigma^2 = c + dS^2, \tag{2}$$

where $f_{\text{mean}} = 1/m\sum_{k=1}^{m} f_k$ denotes the raw ensemble mean and $S^2 = 1/m\sum_{k=1}^{m}(f_k - f_{\text{mean}})^2$ denotes the raw ensemble variance of the power-transformed forecasts. These processes correspond to step 1 and step 2 in Fig. 1.

The EMOS coefficients for the distribution in (2) are fitted by selecting the values that minimize the mean continuous ranked probability score (CRPS; Hersbach 2000)—averaged over all forecast dates in the training period—when evaluated with the observations in that same training period. The CRPS is a proper scoring rule that summarizes the sharpness and calibration of a predictive distribution (Gneiting et al. 2005; Gneiting and Raftery 2007). CRPS is negatively oriented and is defined for a given univariate predictive cumulative distribution function (CDF) F and a verifying observation y as

$$\text{CRPS}(F, y) = \int_{-\infty}^{\infty} [F(t) - H(t - y)]^2 \, dt, \tag{3}$$

where H is the Heaviside step function that equals 1 when its argument ≥0 and equals 0 otherwise.

To fit EMOS coefficients for data transformed with a given power transformation, we employ a closed-form expression of the CRPS (detailed in section 4d). This step corresponds to step 3 in Fig. 1. Power transformations of the data (typically with an exponent ξ between 0 and 1) are often used to reduce skewness and achieve spread that is independent of the forecast magnitude (i.e., homoscedastic). However, strong transformations

(i.e., small ξ) have the negative side effect of emphasizing small forecast values at the expense of larger ones (which are often more important in applications) when the mean CRPS of the power-transformed forecasts and observations is calculated. We try to minimize this effect by using a closed-form expression that evaluates the CRPS on a scale that is as similar as possible to the original (untransformed) scale. The combination of EMOS coefficients that minimize the mean CRPS for a given power transformation define the μ and σ^2 of the predictive distribution in (2). This step corresponds to step 4 in Fig. 1. Special care is needed to choose a closed-form expression of the CRPS that will work for a range of candidate power transformations; we detail this process next.

d. *Selection of closed-form CRPS expression and data transformation*

We are not aware of any closed-form expressions available for an arbitrary exponent ξ in combination with the truncated normal distribution, however, Taillardat et al. (2016) provided a closed-form expression of the CRPS for the *square root–transformed* truncated normal distribution. This expression represents the CRPS of a CDF and the verifying observation on the original scale when a predictive truncated normal distribution is fitted to square root-transformed data ($\xi = 0.5$). However, exploratory analysis of our data showed that the optimal power ξ can be either below or above 0.5. It is possible to test a range of other ξ exponents using the Taillardat et al. (2016) formulation by first applying a pretransformation to the data. If a power $\xi < 0.5$ or $\xi > 0.5$ is required, a pretransformation with an exponent $\eta = \xi/0.5$ is applied to the data and then the Taillardat et al. (2016) expression of the CRPS for the square root-transformed truncated normal distribution is used with the pretransformed forecasts and observations.

We test different power transformations $\xi \in [0.2, 0.3, 0.4, 0.5, 1.0]$, which correspond to pretransformations $\eta \in [0.4, 0.6, 0.8, 1.0, 2.0]$. Using the training forecast and observation pairs for each year, month, lead time, and location, the EMOS coefficients a, b, c, and d are selected separately for each candidate power transformation by CRPS minimization. The result is an optimal univariate calibration for each candidate power transformation. Next, we determine which of the candidate power transformations ξ yields the overall best calibration for each year, month, lead time, and location.

e. *Selection of optimal power transformation*

The optimal exponent ξ is also chosen by CRPS minimization, but is performed after the best EMOS coefficients for each candidate ξ are determined. Since mean CRPS values obtained with different pretransformations are not directly comparable, the CRPSs have to be compared on the original scale. We are not aware of any closed-form expression for an arbitrary exponent ξ, and only 5 possible choices have to be evaluated, so we use the sample CRPS, which is calculated in the following steps:

1) Using the power-transformed ensemble forecasts (for each $\xi \in [0.2, 0.3, 0.4, 0.5, 1.0]$) from the training period and the EMOS coefficients estimated as described above, calculate the calibrated predictive distributions according to Eq. (2) for each candidate ξ. This step corresponds to step 5 in Fig. 1.
2) Calculate a CRPS-optimal sample (e.g., Bröcker 2012) from the calibrated predictive distribution for each ξ by choosing $K = 20$ equidistant quantile levels $\tau_k = (k - 0.5)/K, k = 1, \ldots, K$. This step corresponds to step 6 in Fig. 1.
3) Convert this sample of transformed forecasts back to the original scale for each candidate ξ using a power transformation exponent ξ^{-1}. This step corresponds to step 7 in Fig. 1.
4) Calculate the sample CRPS (Grimit et al. 2006) from these (inverse) power-transformed quantile forecasts and the (untransformed) verifying observation. This step corresponds to step 8 in Fig. 1.

After we calculate the sample CRPS using the steps above for each month, we define a monthly mean sample CRPS by averaging the sample CRPS over all training dates in a rolling 3-month period centered on a given month. Monthly mean sample CRPS values are calculated with the training data for each month, year, lead time, and location, and for each candidate power transformation. The data transformation ξ that produces the minimum sample CRPS is selected for that month, year, lead time, and location. The selection of ξ based on the minimum sample CRPS corresponds to step 9 in Fig. 1.

Note that we select the power transformation based on the minimum mean sample CRPS for each location, year, month, lead time, and atmospheric variable $\tilde{U}$ and $\widetilde{\text{VPD}}$ separately. The necessity of this flexible model approach is illustrated in Figs. 2 and 3, which show that the optimal power transformation is dependent on all of those factors. We provide a discussion on how this flexible model approach could be modified to run in an operational setting in the conclusion section. In either case, a flexible distribution model may not be needed for smaller regions or countries, but to produce the most skillful forecasts

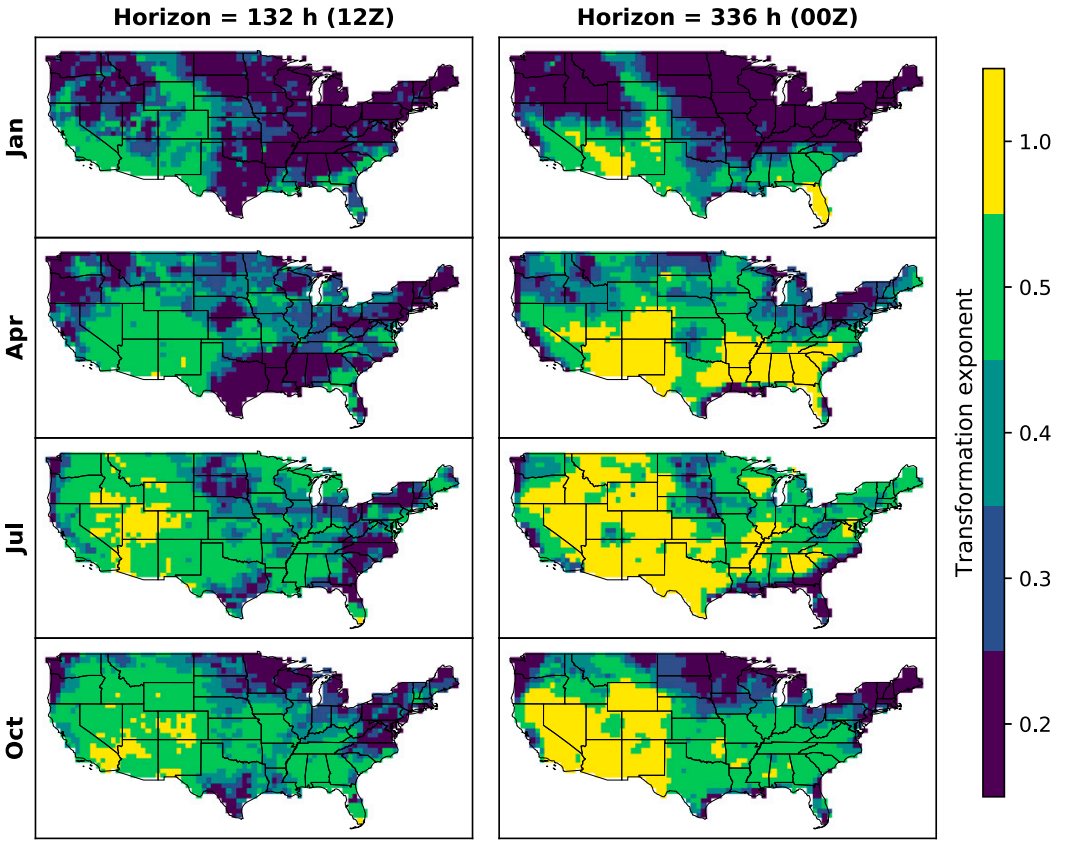

FIG. 2. Power transformation exponents selected by the flexible distribution model for $\widetilde{\text{VPD}}$. The middle month of each season and two forecast horizons starting from a 0000 UTC initialization are shown. The time of day associated with each lead time is written in parentheses.

over all areas of CONUS and for all seasons, this approach is critical.

f. *Verification of calibrated univariate forecasts*

To verify that the univariate forecasts obtained with the selected power transformation and fitted EMOS coefficients yield calibrated probabilistic forecasts during the verification period, we use probability integral transform (PIT) (Gneiting et al. 2007; Dawid 1984). PITs are calculated for the predictive CDFs F_i of the square root-transformed truncated normal distribution defined by the corresponding flexible selection of power transformations. The F_i and its verifying observation y_i define the flexible PIT values as $F_i(y_i)$ for each month, year, lead time, and location. A histogram of the resulting PIT values for all years, all locations, and for all verification days within a given month will be uniformly distributed if the forecasts from the predictive distribution are perfectly calibrated. Verification PIT histograms in Fig. 4 show that the flexible predictive distribution model produces calibrated forecasts for $\tilde{U}$ and $\widetilde{\text{VPD}}$; therefore, we proceed with this flexible model framework.

5. Multivariate postprocessing

The orange-shaded boxes (steps 10–14) in Fig. 1 show a conceptual diagram of the multivariate postprocessing methods outlined throughout the subsections below.

a. *Ensemble copula coupling (ECC)*

The calculation of the HDWI involves different weather variables ($\tilde{U}$ and $\widetilde{\text{VPD}}$) and different lead times (daily maxima of these quantities are required). The forecasts of $\tilde{U}$, $\widetilde{\text{VPD}}$, and subsequently HDWI should be spatially and temporally coherent. Sampling-reordering techniques, such as ECC restore dependencies between variables, lead times, and spatial locations that were lost in the univariate postprocessing steps. Each member of the raw ensemble is based on a physical model and can thus be expected to have a realistic spatial, temporal, and intervariable structure. ECC seeks to transfer this structure to the calibrated forecasts by imposing the

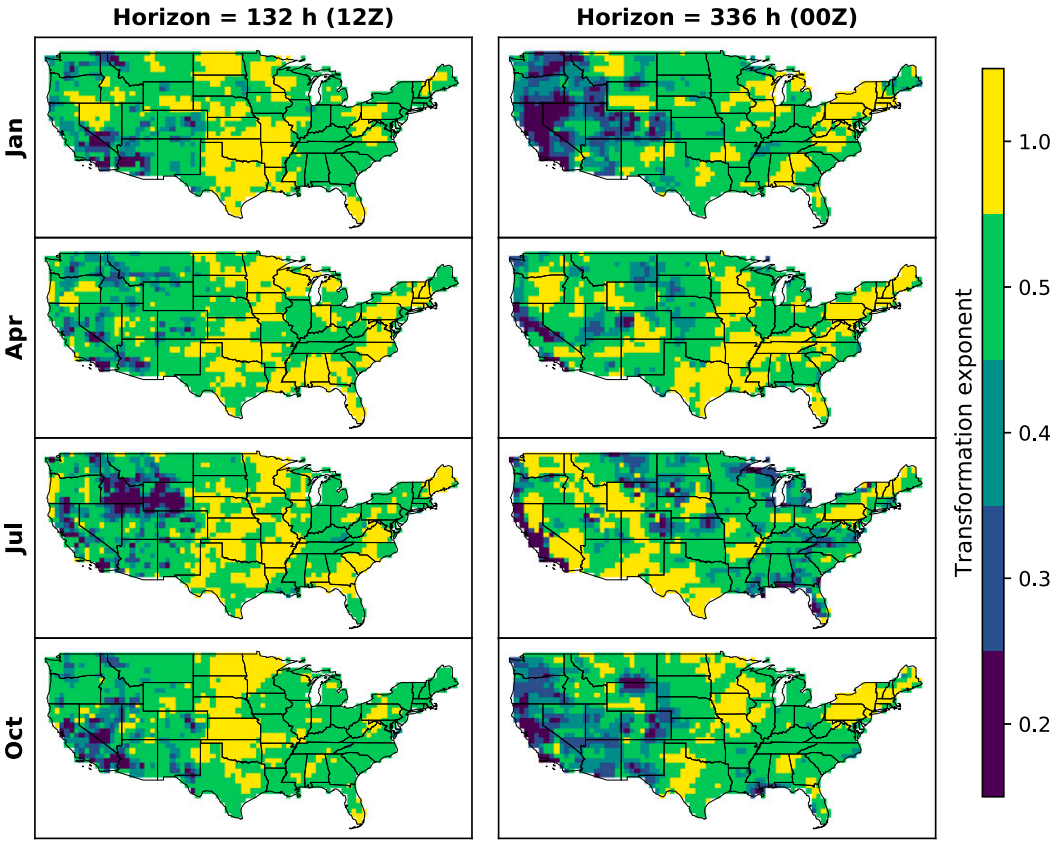

FIG. 3. As in Fig. 2, but for $\tilde{U}$.

rank of an exchangeable raw ensemble onto samples drawn from a postprocessed marginal predictive distribution (like the flexibly transformed truncated Gaussian distribution).

The ECC approach generates a postprocessed ensemble of the same size, m, as the raw ensemble. Various methods of sampling m members from the postprocessed marginal predictive distributions include the random draw (ECC-R), transformation (ECC-T), and the equidistant quantile (ECC-Q) approaches (Schefzik et al. 2013). The quantization used in the ECC-Q approach is the standard and recommended method to generate a representative sample of the postprocessed marginal predictive distribution (Schefzik et al. 2013), so we proceed with this method.

We construct the postprocessed marginal ensembles $x^{j,k} = (x_1^{j,k}, \ldots, x_m^{j,k})$ in the transformed space for each lead time j and location k by sampling m quantiles from the marginal distribution $\mathcal{N}_{0j,k}$ that resulted from the EMOS univariate postprocessing steps for each variable, month, and year. The quantile levels are defined in the same way as τ_k in section 4e, now with $K = m = 11$. These processes correspond to step 10 and step 11 in Fig. 1. After back-transformation to the original data space (step 12 in Fig. 1), $x^{j,k}$ is rearranged and matched according to the rank of the raw forecast ensemble (step 13 in Fig. 1).

b. Calculating daily maxima

Recall that the reforecast data are available at 12-h increments starting from 0000 UTC initialization out to 2 weeks. However, the HDWI calculation calls for the maximum value over 1200, 1800, and 0000 UTC (i.e., daylight hours in CONUS). Therefore, we do the same analysis for an estimate of the 1800 UTC forecast, which we calculated as the average $\tilde{U}$ and $\widetilde{\text{VPD}}$ at 1200 UTC "today" and 0000 UTC the "next day". These average forecasts likely suffer from systematic biases, but with ERA-I reanalysis data being available at 6-h intervals, we were able to correct any systematic biases in our estimate of the 1800 UTC reforecast through the postprocessing steps in sections 4 and 5. From the 1200, 1800, and 0000 UTC data, we calculate the maximum daily $\tilde{U}$ and $\widetilde{\text{VPD}}$ for each postprocessed ensemble member, which we denote as $\tilde{U}_{\max}$ and $\widetilde{\text{VPD}}_{\max}$, respectively. We then calculate the resulting product of those values to yield the daily maximum HDWI.

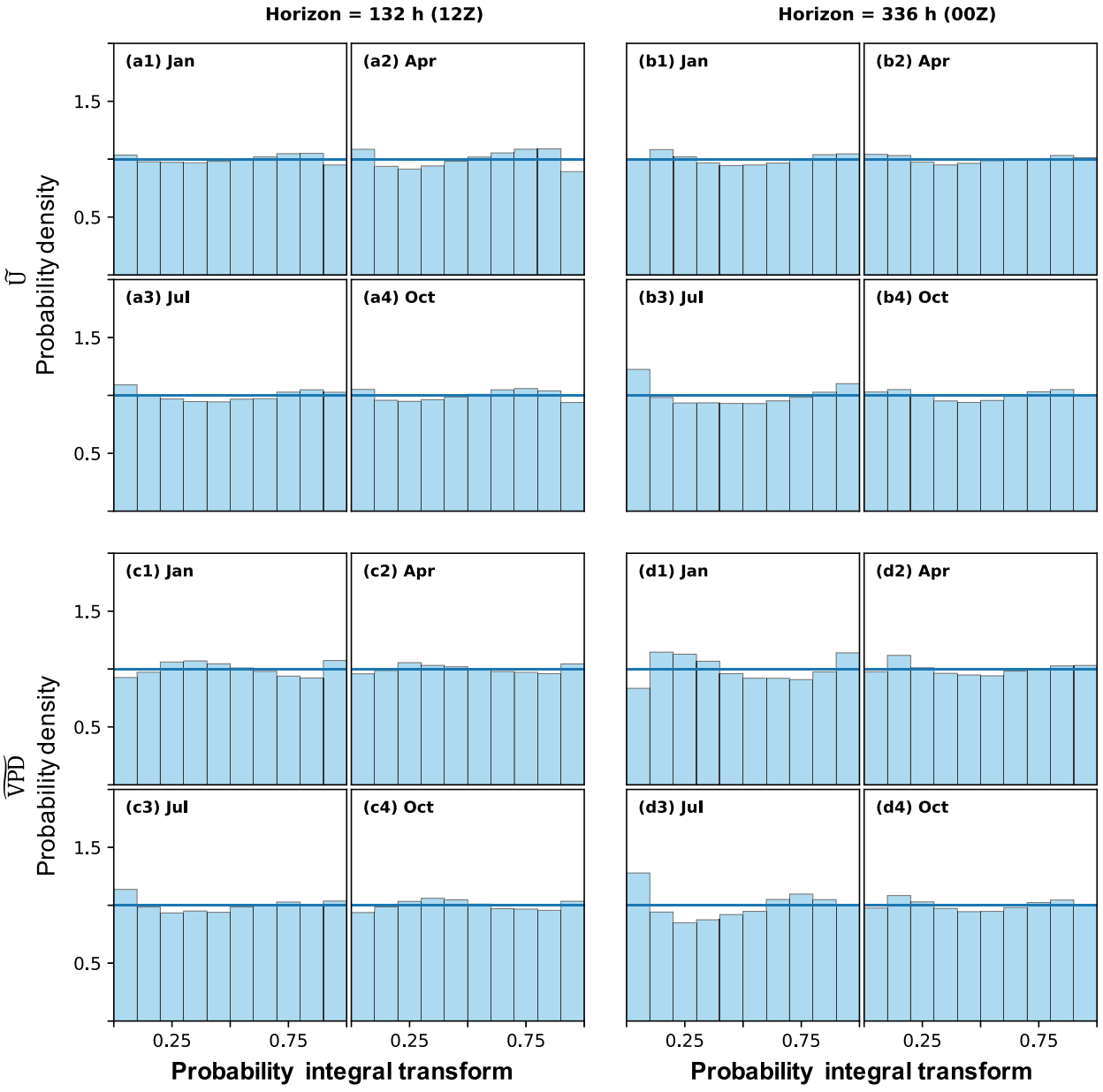

FIG. 4. Verification PIT histograms of (a),(b) $\tilde{U}$ and (c),(d) $\widetilde{\mathrm{VPD}}$ for the middle month of each season and for two forecast horizons. The horizontal blue line indicates uniformity.

Calculation of these daily maxima corresponds to step 14 in Fig. 1.

c. *Verification of calibrated multivariate ensemble forecasts*

Even though the HDWI indicates a hazard for significant fire-weather conditions, a fire will not start if there is not an ignition. For this reason, verifying HDWI forecasts with observed fires may result in many false alarms. Instead, the 11 members of the ECC-Q forecasts of $\tilde{U}_{\max}$, $\widetilde{\mathrm{VPD}}_{\max}$, and HDWI were evaluated against reanalysis data of the same variable for every date in each year, month, and day ahead. The skill score, namely the skill of the CRPS (CRPSS) is used to evaluate the overall performance of the postprocessed forecasts to a defined benchmark. While the score is negatively oriented (smaller positive and negative values are better than larger values), the skill score is positively oriented so that larger positive values are better than smaller positive or any negative values. The CRPSS is calculated as $\mathrm{CRPSS} = -[(\mathrm{CRPS}_{\mathrm{fcst}} - \mathrm{CRPS}_{\mathrm{ref}})/\mathrm{CRPS}_{\mathrm{ref}}]$, where $\mathrm{CRPS}_{\mathrm{fcst}}$ is the CRPS of the ECC-Q ensemble forecast

TABLE 2. Sharpness and calibration results for the raw ensemble and postprocessed ensemble (ECC) for day 6, 8, 10, and 12 lead times during the winter.

83.33% prediction interval: Median width relative to climatological width					83.33% prediction interval: Coverage (%)				
Method	D6	D8	D10	D12	Method	D6	D8	D10	D12
Raw	0.44	0.54	0.63	0.68	Raw	69.42	72.87	75.0	76.73
ECC	0.61	0.71	0.78	0.82	ECC	84.04	83.96	84.04	84.13
66.66% prediction interval: Median width relative to climatological width					66.66% prediction interval: Coverage (%)				
Method	D6	D8	D10	D12	Method	D6	D8	D10	D12
Raw	0.45	0.57	0.65	0.69	Raw	52.12	55.15	57.5	59.23
ECC	0.64	0.75	0.82	0.85	ECC	67.69	67.33	67.31	67.69

and $CRPS_{ref}$ is the CRPS of the reference ensemble forecast. Herein, we mostly consider an ensemble of climatological forecasts as references for comparison, but in some cases we use the raw ensemble forecast as a reference in order to quantify the improvement due to statistical postprocessing. A CRPSS value between 0 and 1 translates to improved forecasting skill over the reference forecast.

6. Verification results and discussion

a. *Calibration and sharpness of the raw and ECC-Q forecasts*

We first quantify the performance of the postprocessed ensemble forecasts using the methods described in the previous sections by calculating the widths and empirical coverages of two predictions intervals for each season. The 83.33% and 66.66% prediction intervals shown in Table 2 (winter) and Table 3 (summer) are bounded by the values of the 1st and 11th and the 2nd and 10th ordered members of the forecast ensemble, respectively. The widths of the prediction intervals allow one to quantify the sharpness of the forecasts and the coverages represent the relative frequencies of the observations lying within those intervals (in the case of optimal calibration, these frequencies would match the nominal levels of the prediction intervals). To normalize the widths so that we can aggregate across all CONUS grid points, we divide the ensemble forecast widths at each location for a given season by the corresponding climatological widths.

Results in Tables 2 and 3 for different lead times show that the raw ensemble is underdispersive, as indicated by the narrow widths and poor coverage of the prediction intervals. The ECC-Q forecast ensemble has wider prediction intervals which capture more possible observed outcomes in the future. Note that the ECC-Q widths are still more sharp than the climatological widths as indicated by width values less than one. For these reasons, the coverage and therefore forecast calibration of the postprocessed forecasts increases. We found similar results during spring and fall (tables are available in the online supplemental material A).

b. *Skill scores of HDWI at select locations*

Skill of the ECC-Q ensemble forecasts compared to the raw ensemble forecasts is evaluated with the CRPSS verification tool. To quantify the uncertainty of the

TABLE 3. As in Table 2, but for the summer.

83.33% prediction interval: Median width relative to climatological width					83.33% prediction interval: Coverage (%)				
Method	D6	D8	D10	D12	Method	D6	D8	D10	D12
Raw	0.59	0.68	0.74	0.78	Raw	73.13	75.31	76.41	77.19
ECC	0.76	0.86	0.90	0.93	ECC	85.31	85.31	85.31	85.63
66.66% prediction interval: Median width relative to climatological width					66.66% prediction interval: Coverage (%)				
Method	D6	D8	D10	D12	Method	D6	D8	D10	D12
Raw	0.61	0.71	0.77	0.80	Raw	55.31	57.81	59.06	59.69
ECC	0.80	0.89	0.93	0.96	ECC	68.75	68.75	69.06	68.75

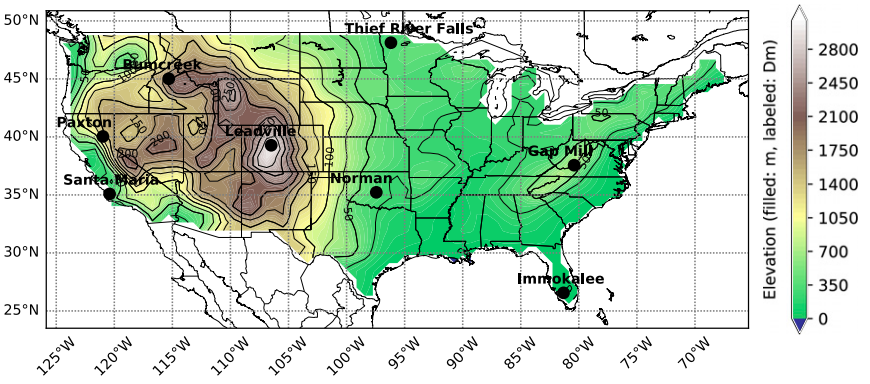

FIG. 5. ERA-I reanalysis elevation (shaded, lines contoured in 50-m intervals) and the eight locations discussed throughout the paper (black dots). Elevations above 500 m are labeled in decameters.

CRPSS, we calculate 100 uniform random bootstrap samples with replacement of the $\text{CRPS}_{\text{fcst}}$ and CRPS_{ref} values. The bootstrap dataset included all CRPS values for a given month, day horizon (there is now one forecast per day–the daily maximum), and all years. For each bootstrap set, the same indices of the $\text{CRPS}_{\text{fcst}}$ and CRPS_{ref} arrays were used to provide a fair comparison across all locations.

Select locations to evaluate the behavior of the CRPSS are plotted with the ERA-I reanalysis elevation in Fig. 5. Wildfires often occur in regions surrounding these locations, which include the Rocky Mountains, Sierra Nevada, Great Plains, coastal areas affected by downslope wind-driven fires, and forested lands across the southeastern United States, upper Midwest, and Appalachian Mountains. Peak fire season varies at these locations. Based on a 24-yr climatology of wildfire occurrence (Short 2017; SPC 2019), fires are more likely to occur during winter in central and southern Florida, late winter and early spring in Oklahoma, spring and fall in the Appalachians, mid-to-late spring near the upper Midwest, and during the summer in the western United States. Wildfires can and do occur in many other regions across CONUS. However, these locations provide a representative subset to explore variations in forecast skill with elevation, location, and season.

Climatological forecasts can often provide the best estimate of the state of the atmosphere when forecasts from numerical weather prediction are unavailable, especially at longer lead times. We therefore compare the CRPS values of the raw and ECC-Q forecasts to the CRPS values of a climatological reference forecast. Climatologies for $\tilde{U}_{\max}$, $\widetilde{\text{VPD}}_{\max}$, and HDWI were calculated from the ERA-I reanalysis data using a ±7-day sliding window around each day of year for all available years from 1997 to 2017. Data from these dates were then divided into equidistant quantiles of the same size as the ECC-Q ensemble (i.e., $K = m = 11$ for τ_k in section 4e).

The median of the CRPSS bootstrap samples of the raw and ECC-Q HDWI forecasts using the HDWI climatology as a benchmark is shown in Fig. 6 (full uncertainty information of the ECC-Q forecasts is shown in the online supplemental material B). At the locations in Fig. 5, the raw forecasts (dashed lines) typically perform worse than climatology during the summer. At some locations during the winter (e.g., Santa Maria, California; Norman, Oklahoma; Immokalee, Florida; Thief River Falls, Minnesota; Gap Mill, West Virginia; Bumcreek, Idaho), the raw forecasts can perform better than or at least as good as climatology. However, at other locations during the winter (e.g., Paxton, California, and Leadville, Colorado), the raw forecasts perform worse than climatology. These results reveal that some locations across CONUS need postprocessing to improve skill over the local climatology, especially during their various peak fire seasons.

The postprocessed forecasts (solid lines in Fig. 6) overall perform better than the raw forecasts at these locations, especially during the summer. Postprocessing at these locations during the spring and fall (plots available in the online supplemental material B) also yields improvements over the raw forecasts. These results show that statistical postprocessing is necessary to provide the most skillful forecasts of HDWI across the CONUS in the extended range. The skill of the ECC-Q forecasts (relative to the climatological forecasts) decreases with lead time for all seasons and locations. For winter, spring, and fall, improved skill over the climatological forecasts of HDWI generally persists up to Day 10. This increased skill for later lead times is particularly valuable for locations in the Great Plains, the upper Midwest, the Appalachian Mountains, and the Southeast where the corresponding peak fire seasons

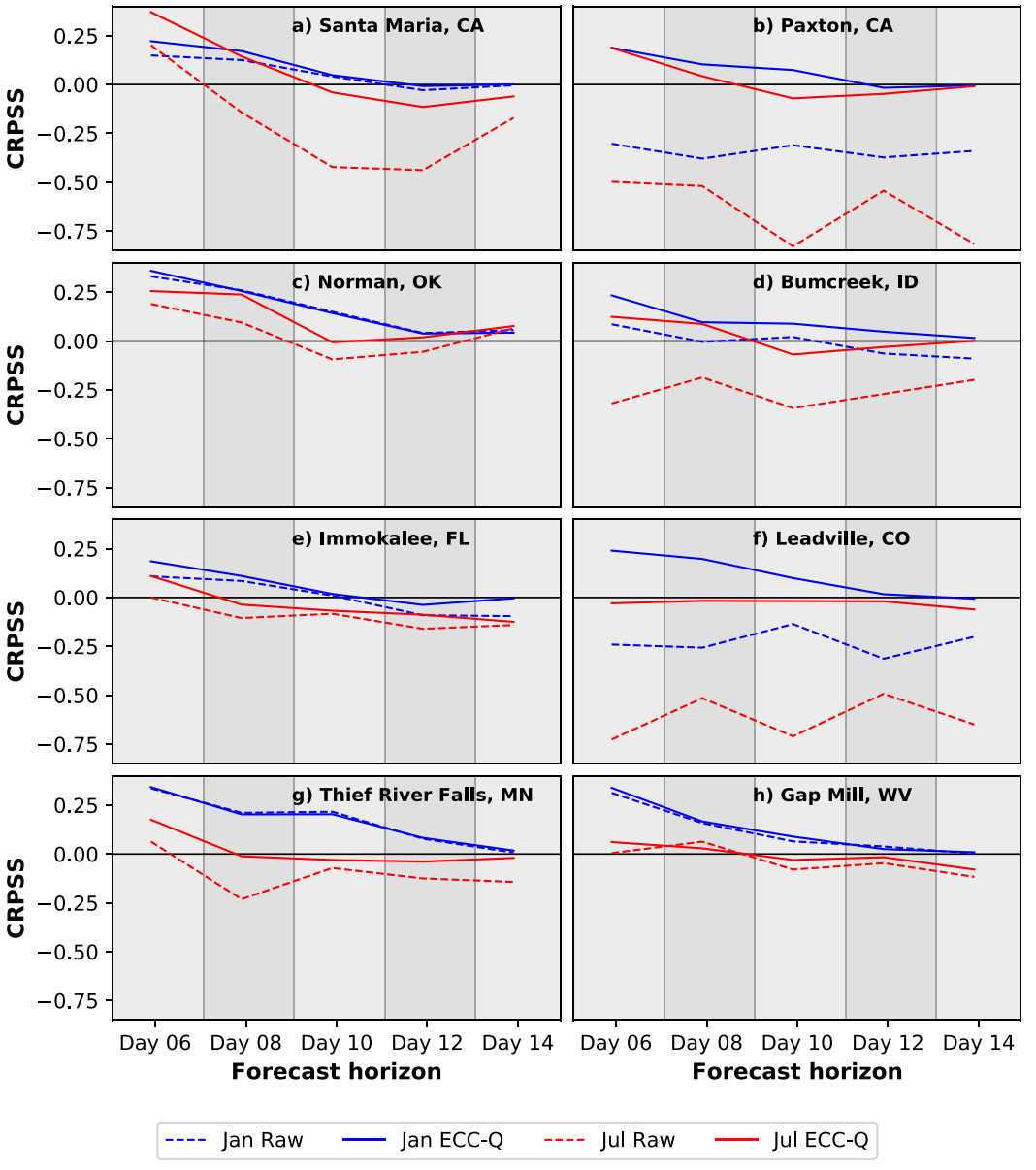

FIG. 6. Median CRPSS for the raw (dashed lines) and ECC-Q forecasts (solid lines) of HDWI at the locations shown in Fig. 5 calculated with a climatological reference. CRPSS is shown for the middle month of winter (blue lines) and summer (red lines) and for forecast horizons out to day 14 in 2-day increments (gray rectangles). Values above the solid black line show greater skill than the climatological forecast.

occur outside of the summer. In summer, the improvement degrades below climatology beyond day 8.

c. *Skill scores of* $\tilde{U}_{\max}$, $\widetilde{\mathrm{VPD}}_{\max}$, *and HDWI for all CONUS*

CRPSS values show improvements of the ECC-Q forecasts over raw reference forecasts of HDWI over most of CONUS (Fig. 7). Median CRPSS values of HDWI (not shown) calculated from using CRPSS values from all CONUS grid points reveal that the most improvement over the raw forecasts happens during the spring and winter at Day 1 (CRPSS is ~0.3). By Day 8, the improvement (i.e., the skill) converges to ~0.1 for all seasons and then remains steady out to Day 14. Across all of CONUS, the maximum improvement of ECC-Q forecasts over the raw forecasts at Day 6 is ~0.5 in spring and winter.

From Fig. 7a, the areas with the most skill over the raw reference forecasts occur in regions of complex terrain across the western United States. In these areas,

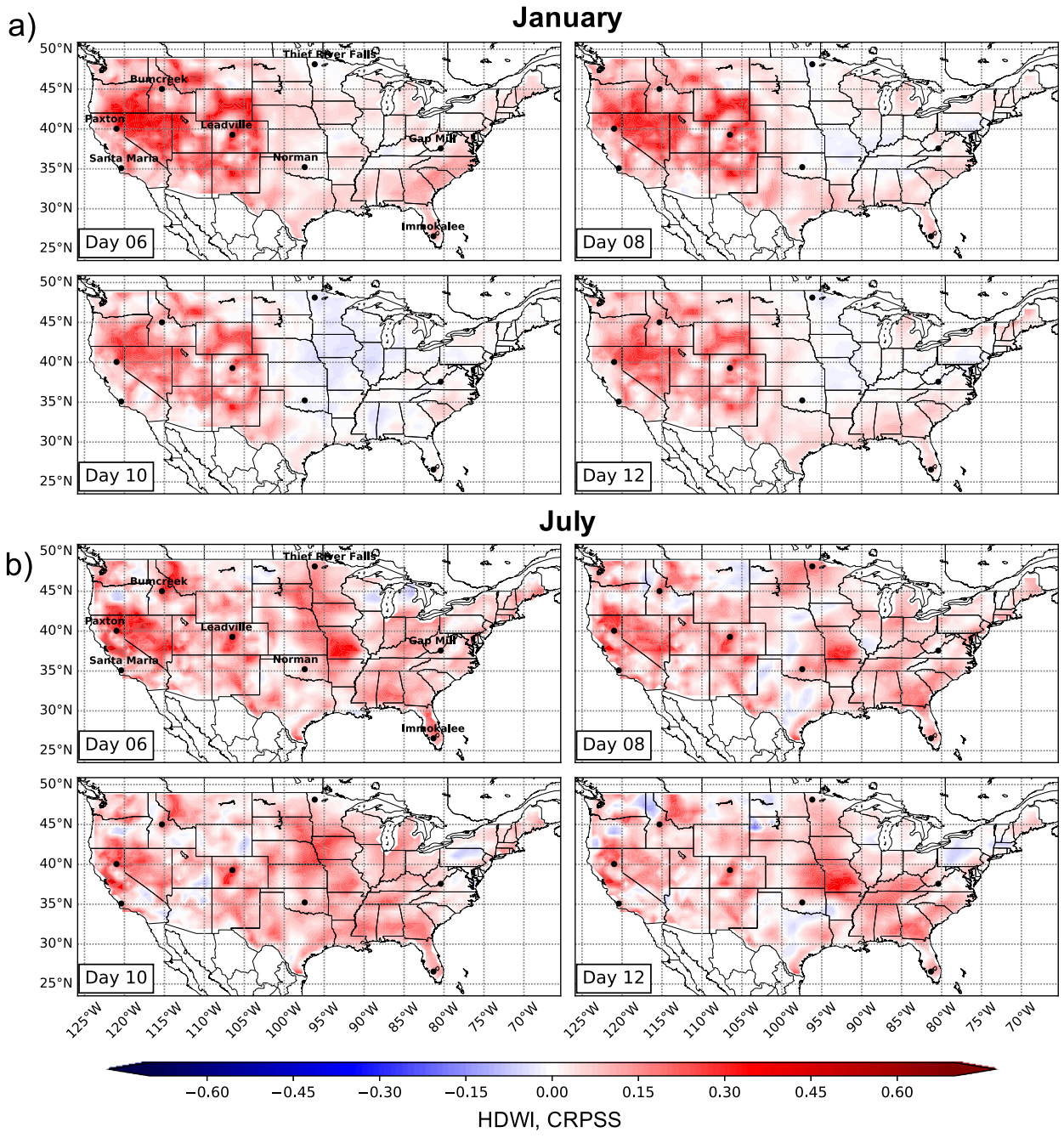

FIG. 7. CRPSS values of HDWI calculated with the ECC-Q ensemble relative to a raw ensemble reference forecast for the middle month of (a) winter and (b) summer for 6, 8, 10, and 12 days ahead. Red shaded regions indicate where the skill is improved over the raw forecast. Black dots correspond to the locations in Fig. 5.

systematic biases of wind and temperature with height are common in raw forecasts by NWP models. Improvements through postprocessing are more evenly spread across CONUS in the summer (Fig. 7b). This result is because the raw model has a harder time forecasting conditions in the summer months, even away from complex terrain (not shown). While initial good skill of raw forecasts will translate to enhanced skill via postprocessing, poor initial skill from systematic biases leaves more room for improvement by the postprocessing methods.

When comparing the ECC-Q forecasts of HDWI to climatology, the skill gained depends on lead time and season (Fig. 8). The median (maximum) CRPSS

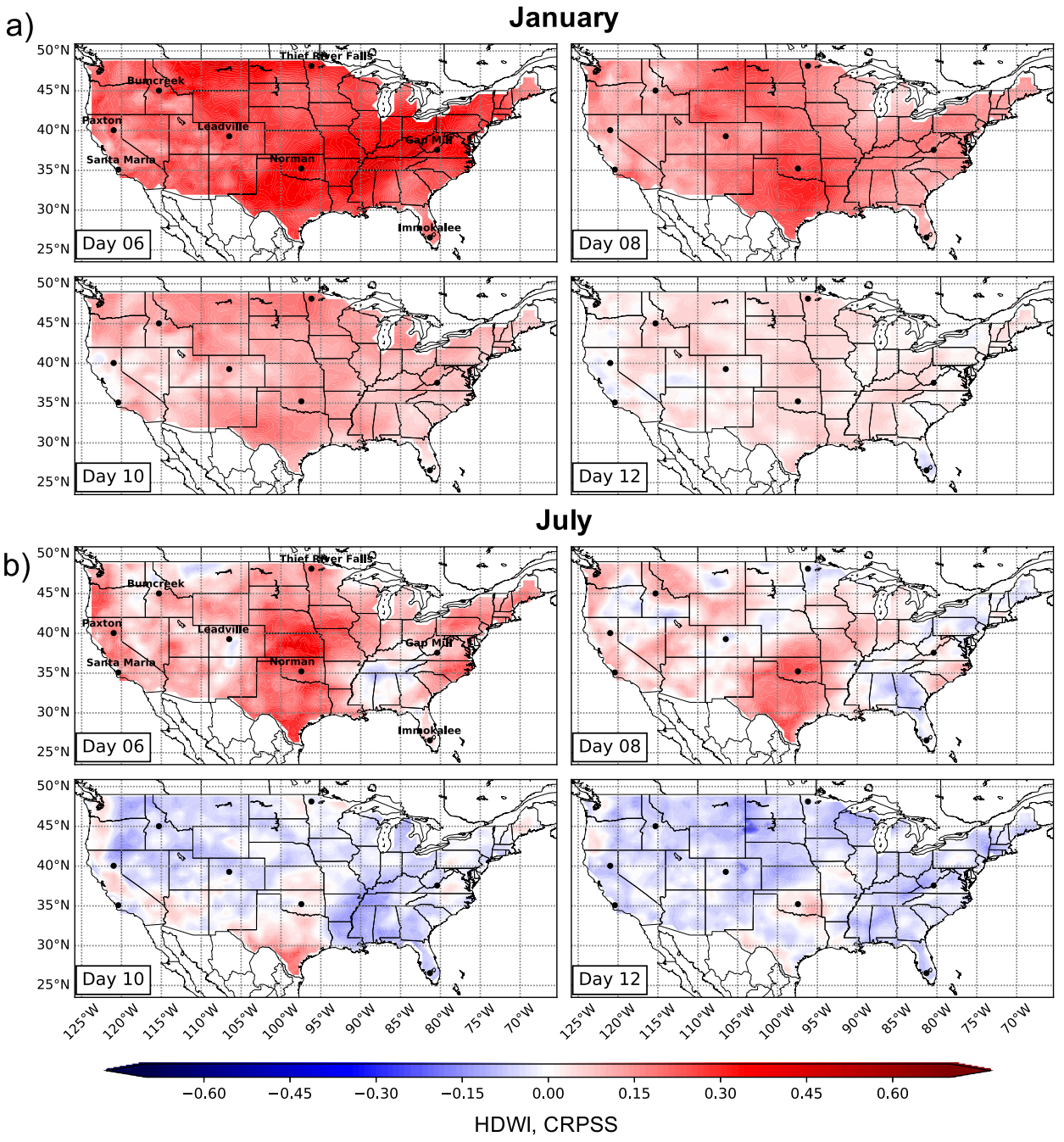

FIG. 8. CRPSS values of HDWI calculated with the ECC-Q ensemble and the climatological reference forecast for the middle month of (a) winter and (b) summer for 6, 8, 10, and 12 days ahead. Red shaded regions indicate where the skill is improved over a climatological forecast. Black dots correspond to the locations in Fig. 5.

values of all CONUS grid points in winter range from >0.3 (0.45) at Day 6 to 0.05 (0.1) by Day 12 across the majority of CONUS. Conversely, for summer, the median (maximum) values at Day 6 are > 0.15 (0.35) and fall to negligible improvements by Day 9, except for Texas, Oklahoma, and Arizona. When considering improvement over the majority of CONUS, one could expect skill of the HDWI out to Day 7 in the summer, which is particularly valuable for the western U.S. fire seasons and up to Day 11 in the winter, which is valuable for fire predictions during the peak season in Florida. Spring and fall show improved skill over climatology for most of CONUS up to Day 9 (plots available in the online supplemental material B), again adding enhanced

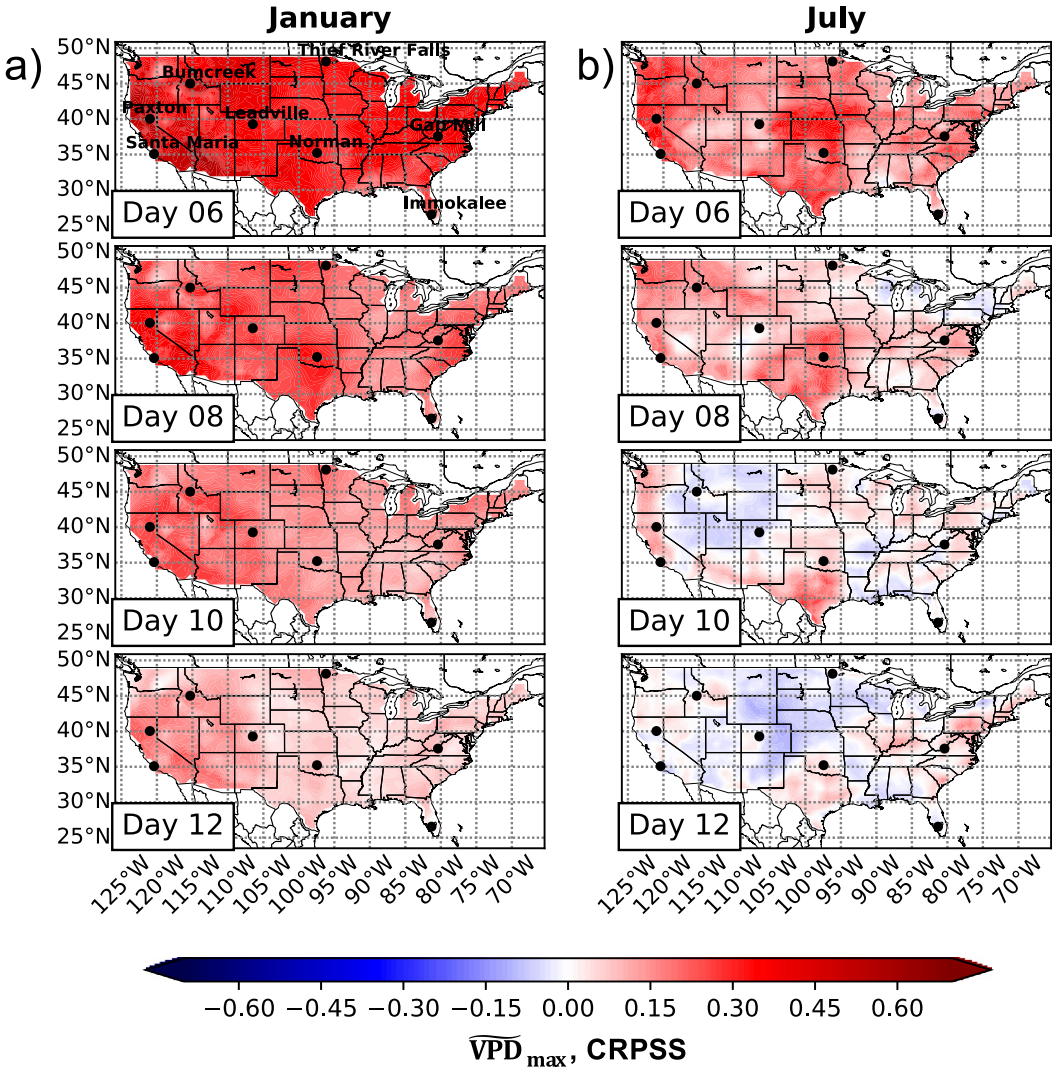

FIG. 9. CRPSS values of daily maximum $\widetilde{\text{VPD}}$ of the ECC-Q ensemble forecasts relative to a climatological reference forecast for the middle month of (a) winter and (b) summer for 6, 8, 10, and 12 days ahead. Red shaded regions indicate where the skill is improved over climatology. Black dots correspond to the locations in Fig. 5.

skill during peak fire seasons in the Great Plains, upper Midwest, and along the Appalachian Mountains.

The daily maximum HDWI is evaluated as a whole, but looking at the skill of the individual components can reveal limits of predictability. The $\widetilde{\text{VPD}}_{\text{max}}$ (Fig. 9) in the winter yields the most skill relative to climatology when compared to $\tilde{U}_{\text{max}}$ (Fig. 10) or summer months across the majority of the CONUS domain. The median CRPSS values from all grid points of $\widetilde{\text{VPD}}_{\text{max}}$ is ~0.4 in the winter, spring, and fall at Day 6 and stays above 0 out to ~ Day 14 (purple, green, and gray solid lines in Fig. 11a). For the summer (orange solid line in Fig. 11a), the median value for $\widetilde{\text{VPD}}_{\text{max}}$ by Day 6 is half that found in other seasons. Summer median skill also drops to zero three days earlier than the other seasons. Although not as high as $\widetilde{\text{VPD}}_{\text{max}}$, the median skill of $\tilde{U}_{\text{max}}$ (dashed lines in Fig. 11a) follows a similar seasonal and lead-time pattern as $\widetilde{\text{VPD}}_{\text{max}}$. Median skill for $\tilde{U}_{\text{max}}$ drops to 0 around Day 11 for winter, spring, and fall, and by Day 9 for summer. Overall, when comparing the median skill over climatology for all of CONUS, the $\widetilde{\text{VPD}}_{\text{max}}$ yields approximately 2–3 more days of positive skill than $\tilde{U}_{\text{max}}$.

Even though the median skill can begin to drop below 0 by Day 9 for summer and later for the other seasons, we can expect positive skill out to Day 14 at some locations (Fig. 11b). The longest-lived high skill scores for both variables tend to be across the western United States in winter and across Texas and Oklahoma in the summer. At these locations, a forecaster could utilize

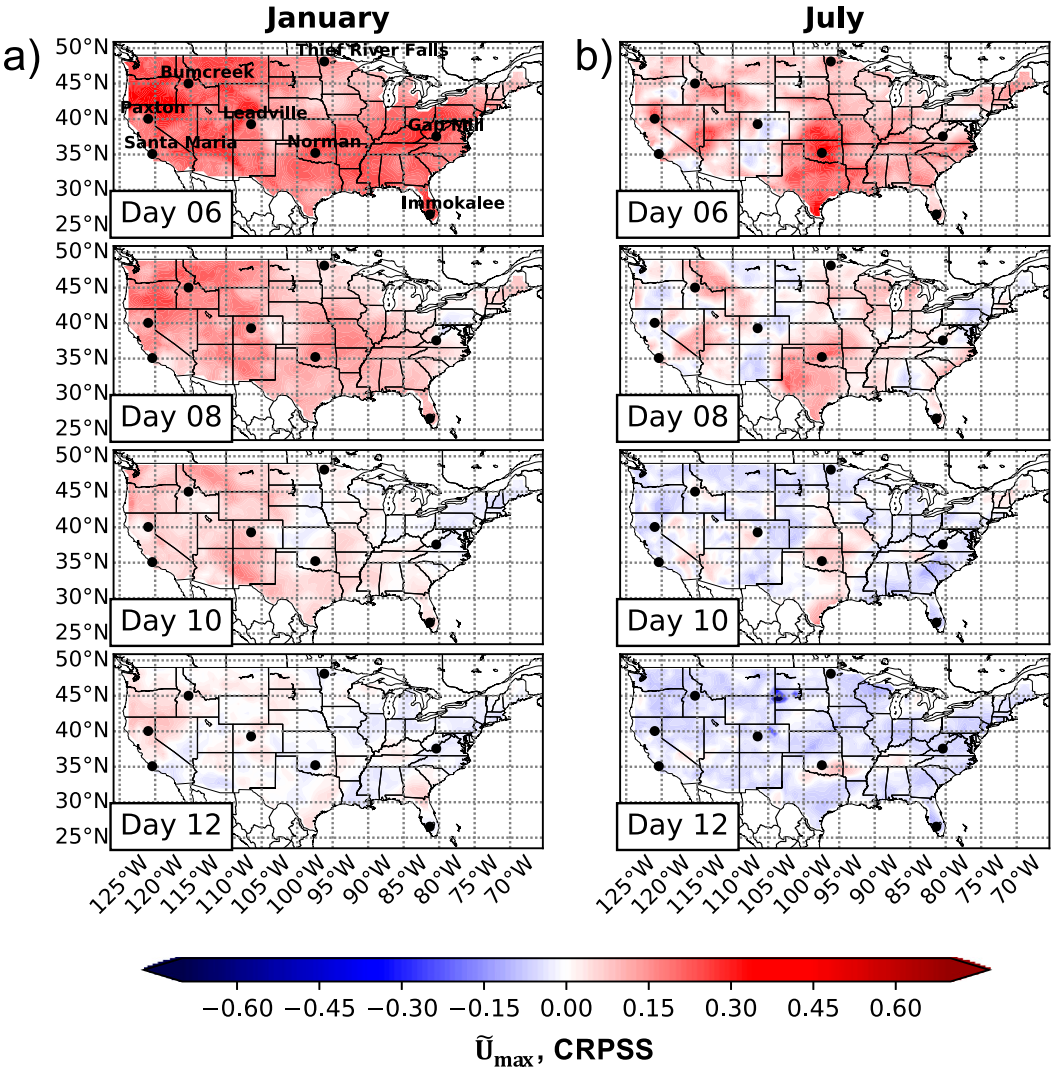

FIG. 10. CRPSS values of the daily maximum $\tilde{U}$ of the ECC-Q ensemble forecasts relative to a climatological reference forecast for the middle month of (a) winter and (b) summer for 6, 8, 10, and 12 days ahead. Red shaded regions indicate where the skill is improved over climatology. Black dots correspond to the locations in Fig. 5.

the longevity of improved skill to make 2-week-ahead forecasts that can be more skillful than climatology.

d. Effect of aggregating forecast days on skill scores

Not only do decision-makers need skillful forecasts out to two weeks at select locations, they also need them for all of CONUS, especially when making regional or national decisions. To enhance the predictability discussed in the previous section, we aggregate forecast days. At longer lead times, forecasts inherently struggle with accurately predicting instantaneous states of the atmosphere (Epstein 1988) and therefore must rely on collective statistics (e.g., averages) to generate the most skill. Even though the HDWI and its components are not strictly instantaneous, but rather daily maxima quantities calculated over the course of 24 h, we can still gain more skill through further aggregation. By averaging $\tilde{U}_{max}$, $\widetilde{VPD}_{max}$, or HDWI over a calendar period, we can relieve the burden of the forecast to get the timing of meteorological conditions just right. National prediction centers such as the Climate Prediction Center (CPC) and the Storm Prediction Center (SPC) already utilize aggregated forecast days to create ''outlooks'' of anomalous weather for a given calendar period. Notable examples include the CPC's 6–10 and 8–14 day temperature and precipitation outlooks. We explore the advantages and limitations of averaging $\tilde{U}_{max}$, $\widetilde{VPD}_{max}$, and HDWI over these outlook periods.

To create a 6–10 day outlook, for each reforecast date, we average Day 6 up to and including Day 10 ECC-Q

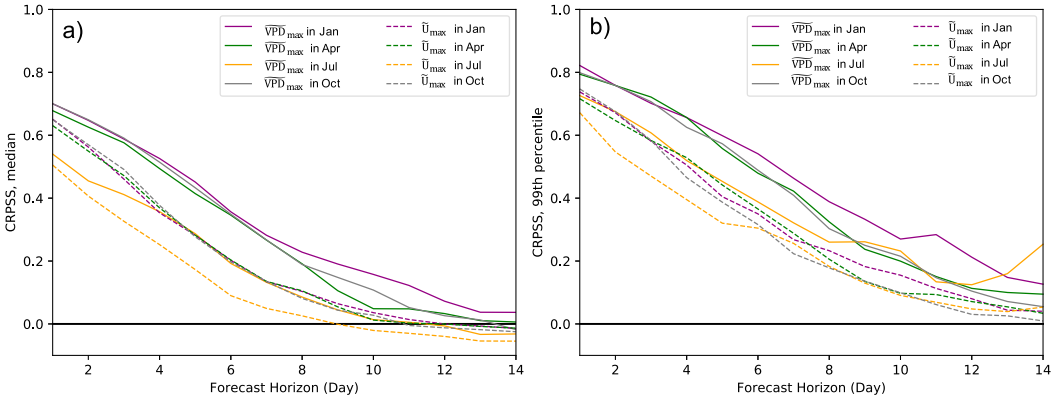

FIG. 11. (a) The median and (b) the 99th percentile of all CRPSS values in CONUS for day 1–14. $\widetilde{\mathrm{VPD}}_{\mathrm{max}}(\tilde{U}_{\mathrm{max}})$ values are shown for the middle months of each season in solid (dashed) lines. The reference forecast is climatology. Values above the black horizontal line show greater skill than climatological forecasts.

forecasts for each ensemble member. The same is done for Day 8–14 outlook, but using the corresponding lead times. Climatological reference forecasts are created in the same way for each date of the ERA-I reanalysis that falls within each reforecast date-outlook period combination (e.g., if a reforecast date is Monday 9 January 2017 and the outlook period of interest is Day 6–10, then a aggregated climatology is created using all ERA-I reanalysis years during 15–19 January). This process results in 20 years of average values for each outlook period starting from each day of the year associated with the reforecast dataset. Equidistant quantiles (i.e., $K = 11$ for τ_k in section 4e) from these climatological averages make up the reference ensemble forecast when calculating CRPSS.

When comparing the aggregated 8–14 day outlooks of HDWI (Fig. 12) to the single-day forecasts in Fig. 8, we see that the outlooks are comparable in skill to ~ Day 8 (Day 9) forecasts in winter (summer). Considering a longer time window enhances the skill of the ECC-Q forecasts, which is particularly valuable during summer when the single-day forecasts of the HDWI components, $\widetilde{\mathrm{VPD}}_{\mathrm{max}}$ (Fig. 9) and $\tilde{U}_{\mathrm{max}}$ (Fig. 10), begin diminishing beyond Day 8 and Day 6, respectively. Overall, we can expect strong skill over climatology across most of CONUS for aggregated Day 8–14 (Day 6–10) HDWI forecasts during the winter (summer). Similar to the single-day forecasts, some locations have positive skill for even longer aggregated forecast periods (e.g., Texas and Oklahoma show skill at Day 8–14 for HDWI in the summer). Aggregated forecasts during the spring and fall show improved skill over climatology as well (plots are available in online supplemental material C), with the Day 8–14 HDWI forecasts not having as much improvement as in the winter, but overall more improvement than in the summer.

7. Fire event case study

a. *Illustration of single-day forecasts*

The results in the previous sections show that statistical postprocessing is advantageous to obtain skillful extended-range forecasts of HDWI, $\tilde{U}_{\mathrm{max}}$, and $\widetilde{\mathrm{VPD}}_{\mathrm{max}}$. We now explore how these postprocessed forecasts along with climatology can be used to generate probabilistic fire-weather and fire-indicator forecasts for a real fire event. We created the images in Fig. 13 based on the climatology figures in McDonald et al. (2018) and the type of figure used in the experimental web version of the HDWI (HDWI 2019). However, there are some key differences. We calculate the climatology with ERA-I reanalysis data while their climatology was produced with the Climate Forecast System Reanalysis (CFSR, Saha et al. 2010) dataset. Because these datasets differ in resolution (0.75° grid spacing compared to their 0.5°), model physics, and data assimilation schemes, the weather produced in each of these datasets should not be compared. Therefore, this case study illustrates a new climatological dataset for HDWI, $\tilde{U}_{\mathrm{max}}$, and $\widetilde{\mathrm{VPD}}_{\mathrm{max}}$. These climatologies can be used to infer the severity of HDWI, $\tilde{U}_{\mathrm{max}}$, and $\widetilde{\mathrm{VPD}}_{\mathrm{max}}$ quantities forecasted by the ECMWF model with the same grid spacing. Another difference between their analysis and ours is that we not only look at the HDWI, but also the individual components $\tilde{U}_{\mathrm{max}}$ and $\widetilde{\mathrm{VPD}}_{\mathrm{max}}$. Last, we extend our forecasts out to two weeks.

The climatology for HDWI, $\tilde{U}_{\mathrm{max}}$, and $\widetilde{\mathrm{VPD}}_{\mathrm{max}}$ in Fig. 13a is calculated the same way as in section 6a using a ±7-day sliding window around each day of the year and all available years between 1997 and 2017. Instead of separating the daily climatological distributions into equidistant quantiles, we evaluate the 0.25, 0.50, 0.75, 0.90, and 0.95 quantiles levels. The daily

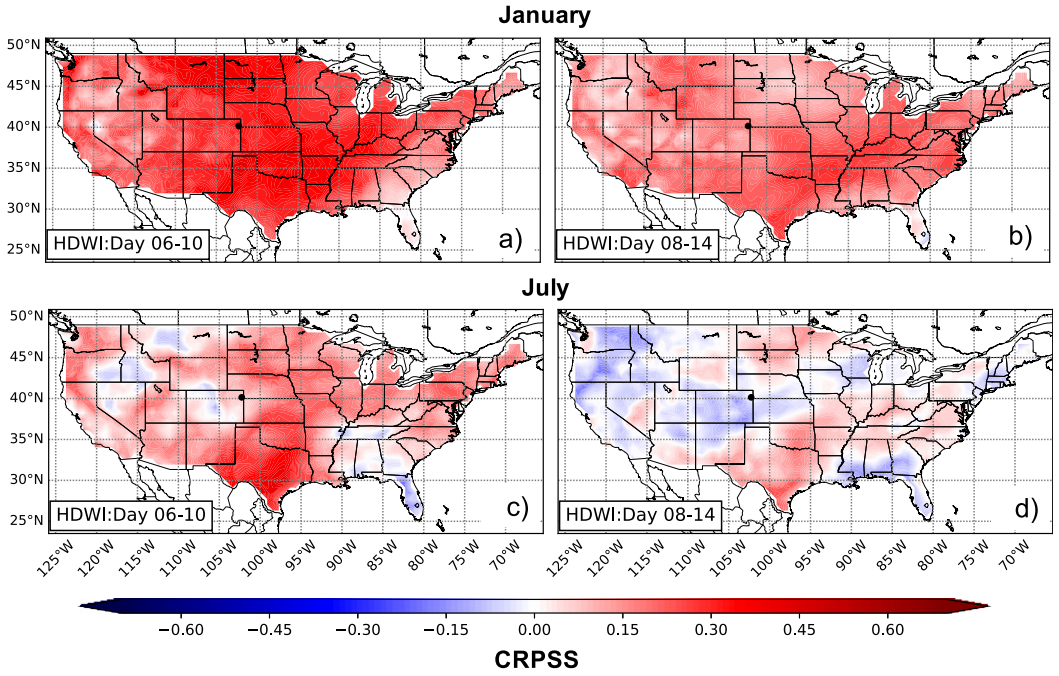

FIG. 12. CRPSS values of HDWI for the middle month of (a),(b) winter and (c),(d) summer. The CRPSS was calculated with averaged ECC-Q ensemble forecasts and averaged climatological reference forecasts for two calendar periods: day 6–10 and day 8–14. Red shaded regions indicate where the skill is improved over climatology. The black dot in northeast Colorado marks the location of the Heartstrong fire discussed in section 7.

climatology of the HDWI, $\tilde{U}_{\max}$, and $\widetilde{\mathrm{VPD}}_{\max}$ are shown (Fig. 13a) at the location of the Heartstrong fire, which ignited on 18 March 2012 in Yuma County, Colorado (the location is marked as the black dot in Fig. 12). The climatologies show a strong seasonal cycle. At this location, the peak HDWI and $\widetilde{\mathrm{VPD}}_{\max}$ happen during June and July while $\tilde{U}_{\max}$ is at its minimum. Since these climatologies depend on location and time of year, they are helpful to assess if a particular HDWI, $\tilde{U}_{\max}$, and $\widetilde{\mathrm{VPD}}_{\max}$ forecast is high for that day and location. On the day of ignition, the fire burned approximately 10 000 ha leading to three firefighter injuries, two destroyed homes, and loss of livestock (Gabbert 2012). The fire was most threatening on 18 March, as firefighters achieved 100% containment by the next day.

The developers of the HDWI intended for the index to alert a forecaster to days that need to be looked at further to understand the meteorological conditions that may make a fire difficult to manage. Figure 13b illustrates why this more in-depth analysis is necessary. The Heartstrong fire occurred over cured western perennial grasslands during a surge in wind speed (> 95th percentile) that followed several days of anomalously high (> 95th percentile) $\widetilde{\mathrm{VPD}}_{\max}$. Although most of the HDWI ECC-Q ensemble members showed spikes during this day, inclusion of the individual components helps explain why there is a peak. Knowing which factors may dominate the fire conditions could help fire managers decide on the best mitigation and suppression tactics.

For this case, the postprocessed ensemble members also performed better than the raw forecasts (Fig. 13b). While the ECC-Q and raw members are similar in shape, the magnitude of the postprocessed forecasts are closer to the reality of the event, especially for $\widetilde{\mathrm{VPD}}_{\max}$.

b. *Illustration of aggregated forecast outlooks*

Postprocessed probabilistic forecasts of HDWI are also shown for aggregated forecast outlooks (Fig. 14) to illustrate how these outlooks could be used by a fire-weather forecaster. Similar figures for $\tilde{U}_{\max}$, and $\widetilde{\mathrm{VPD}}_{\max}$ are available in the online supplemental material C. For each forecast day (e.g., 11 March 2012 in Fig. 14a), lead times from a defined calendar period/time scale (e.g., Day 4–8 ahead in Fig. 14a) are averaged for each variable and ensemble member separately. ERA-I climatologies of these variables and calendar periods are calculated by taking the average during the same calendar period as the forecast (e.g., 15–19 March in Fig. 14a) for every year in the 20-yr record. The resulting climatological distribution is separated into terciles representing the lower, middle, and upper thirds of the climatology for that start date and calendar period. The members of the aggregated mean forecast are then

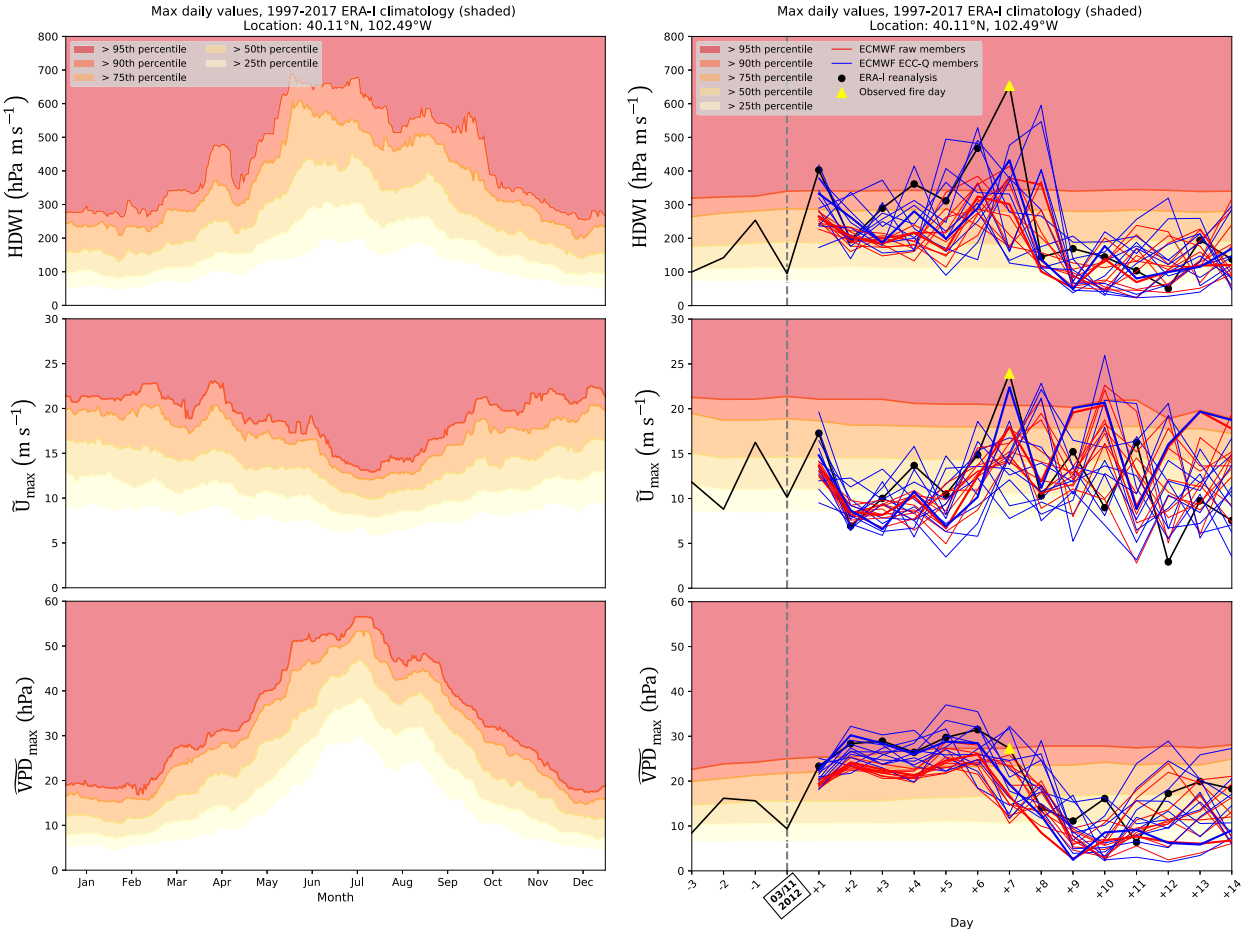

FIG. 13. (a) Percentiles of the daily ERA-I climatology of (top) HDWI, (middle) $\tilde{U}_{max}$, and (bottom) $\widetilde{VPD}_{max}$ at the location of the Heartstrong fire. (b) ECC-Q ensemble (blue lines) and raw ensemble (red lines) forecasts of the (top) HDWI, (middle) $\tilde{U}_{max}$, and (bottom) $\widetilde{VPD}_{max}$ compared to percentiles of the ERA-I climatology (shaded colors). The ERA-I reanalysis is also displayed by the black dotted line. The start day of the forecast is indicated by the gray dashed line, and the start day of the fire is shown as the yellow triangle. Bold red and blue lines indicate the control ensemble member.

marked according to where they fall in the categories: lower (≤33.3%), middle, or upper (≥66.6%) tercile of the aggregated climatology. Based on the percentage of aggregated forecast members that fall into those categories (see Table 4) determines whether the aggregated forecast is labeled as ''above normal'', ''below normal'', or ''normal/equal chance.'' Equal chance is assigned when the probability of the forecast being above or below normal is indistinguishable. Both normal and equal chance categories are masked out as white in the forecast probability plots in Fig. 14. Comparisons with observed anomalies (i.e., did the average observation during the same calendar period fall in the upper, middle, or lower tercile of climatology) detail how well the forecasted anomalies did (Fig. 14).

For the 11 March 2012 case in Figs. 14a and 14c, the forecasted probabilities of anomalous HDWI values performed well. The forecasted pattern of above normal and below normal values matches the pattern of observed anomalies for HDWI, owing to the good performance of its components, especially that of $\widetilde{VPD}_{max}$. The forecast also has high confidence (represented by dark red and blue colors in the left panel in Fig. 14a) in the outcome. As the calendar period includes later lead times, Day 8–14, the confidence in the upcoming anomalies decreases, but the overall forecasted pattern matches the observed pattern, except for $\tilde{U}_{max}$. For the 12 July 2012 case in Figs. 14b and 14d, the overall pattern of forecasted anomalies matches the observed anomalies for Day 4–8, but as expected, is less confident than the winter case (mostly because of the performance $\tilde{U}_{max}$). The Day 8–14 forecasted anomalies give an idea of below normal HDWI in the western United States, but for this case, a decision-maker should not put too much weight on the forecast to make decisions until the forecast becomes more confident. In both

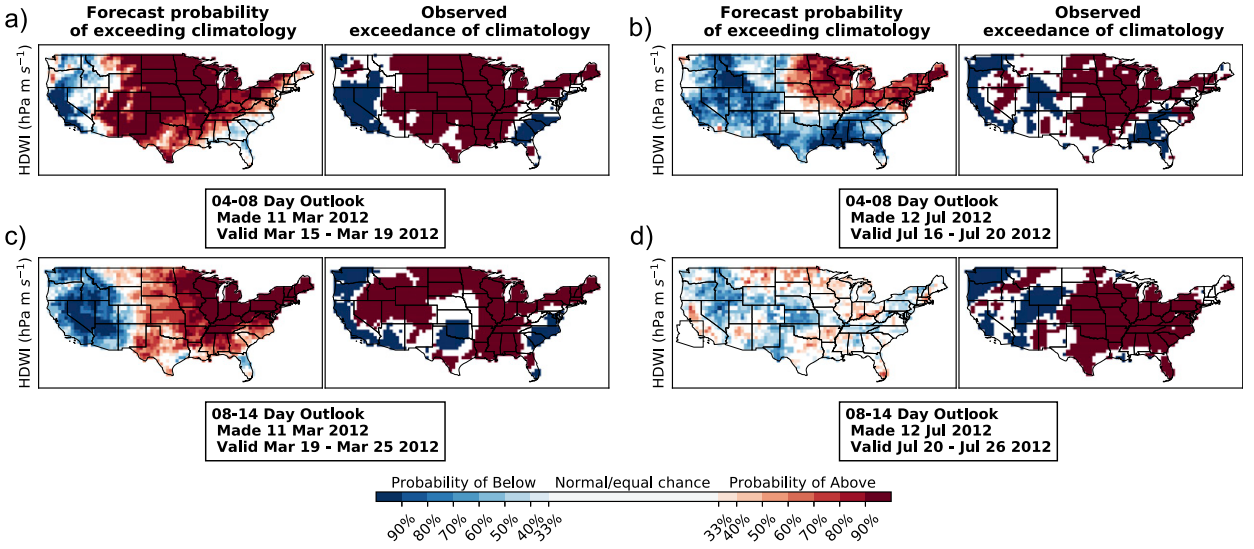

FIG. 14. Probability of HDWI being above (red colors), below (blue colors), or normal/equal chance (white) compared to climatology [left panels in (a)–(d)] for two aggregated calendar periods. Also shown are observed exceedances of climatology (red above, blue below, white normal) of HDWI [right panels in (a)–(d)] for the same aggregated calendar periods. The two calendar periods are (a),(b) day 4–8 and (c),(d) day 8–14 for (a),(c) a winter case and (b),(d) a summer case.

cases, although forecast confidence decreases with later lead times, aggregating the forecast days mostly yields the overall pattern of above and below normal fire-weather conditions for the coming weeks. This information is valuable to fire- and land-management agencies to make decisions regarding where to focus their attention on across CONUS and how many resources may be needed.

8. Conclusions

We employed univariate and multivariate statistical postprocessing methods to generate skillful forecasts of the HDWI, $\tilde{U}_{\max}$, and $\widetilde{\mathrm{VPD}}_{\max}$ out to two weeks ahead. We first corrected biases in the 20 years of reforecasts for each lead time, month, year, and location separately using the EMOS method. The EMOS regression coefficients were selected based on the minimization of the CRPS of the truncated normal distribution (with different power transformations applied to the data). Because the climatologies of $\tilde{U}$ and $\widetilde{\mathrm{VPD}}$ vary throughout CONUS, no single power transformation was optimal. For this reason, we allowed for a flexible distribution model that is based on the optimal power transformation for each particular lead time, month, year, location, and atmospheric variable. Minimum CRPS values and uniform PIT histograms verified that the flexible model is an appropriate distribution to represent $\tilde{U}$ and $\widetilde{\mathrm{VPD}}$. We then applied the multivariate ECC-Q approach to generate spatially and temporally coherent forecast scenarios across lead times. This approach used the ranks of the raw ensemble members to define the dependence template that links quantiles of the univariate predictive distributions across lead times.

Skill of the CRPS showed that the above postprocessing techniques yield improvements over the raw forecasts of HDWI, $\tilde{U}_{\max}$, and $\widetilde{\mathrm{VPD}}_{\max}$ and also over the corresponding climatological forecasts. Generally, there was more skill of the postprocessed forecasts relative to climatology in the winter, spring, and fall than in the summer. Similarly, there was more skill for $\widetilde{\mathrm{VPD}}_{\max}$ than for $\tilde{U}_{\max}$. However, even during the summer, the median forecast skill for a single day forecast was positive out to Day 9 (Day 8) for $\widetilde{\mathrm{VPD}}_{\max}$ ($\tilde{U}_{\max}$). Some locations maintained positive skill up to Day 14. We found that even more forecast skill is possible by aggregating lead times together for different lengths of calendar periods. This approach minimized timing errors and would allow a decision-maker to evaluate the general meteorological conditions up to two weeks ahead.

TABLE 4. Climatological categories for the aggregated ensemble forecasts. Categories are assigned based on the percentage of ensemble members that fall into the upper and lower terciles of climatology.

Category	Percentage of members in the upper tercile of climatology	Percentage of members in the lower tercile of climatology
Above normal	≥33.3%	≤33.3%
Normal	≤33.3%	≤33.3%
Equal chance	≥33.3%	≥33.3%
Below normal	≤33.3%	≥33.3%

Aggregating the forecasts for a calendar period of 8–14 days ahead often was equivalent to a Day 9 single-day forecast.

We then illustrated how these single-day forecasts and aggregated forecast outlooks of HDWI, $\tilde{U}_{\max}$, and $\widetilde{\mathrm{VPD}}_{\max}$ could be visualized and used by a forecaster. The single-day forecast figure showed the ECC-Q forecasts and raw forecasts for a past fire event compared to percentiles of climatology. Visualizing not only the HDWI, but also its components $\tilde{U}_{\max}$ and $\widetilde{\mathrm{VPD}}_{\max}$ (and likely other weather variables too), details the meteorological conditions that would drive a potential fire. This information is critical for wildfire mitigation and suppression strategies. We also showed how the enhanced skill from aggregating forecast lead times could be used to predict probabilistic HDWI, U, and VPD anomalies. The inclusion of later lead times in the aggregating calendar period made the forecast confidence decrease. However, we found that the overall pattern of climatological anomalies was still apparent and could be useful for decision-makers that want to see not only what the conditions two weeks ahead look like, but also how confident the ECC-Q forecasts are in that result.

The goal of this paper was to use statistical postprocessing methods to generate skillful extended-range forecasts of a fire-weather indicator (and its components) used by the fire community. We also wanted to show the limits of predictability of each of these variables. Based on these results, it is likely that forecasts from other similar fire-weather indicators would benefit from postprocessing as well.

If these data, variables, and methods were used in an operational setting, one could ensure that the forecast skill never drops below climatology by making a couple of changes. For the lead times that have skill below climatology, instead of using the EMOS method and truncated square root-transformed normal distribution, one could simply use quantiles of climatology for these select lead times. The multivariate method to generate forecast scenarios would be the same. A forecaster may also wish to know what the fire-weather conditions will be for a large region rather than a local forecast. One could spatially aggregate the postprocessed ensemble members in the end, which gives the option of local or regional forecasts.

Additionally, recall that we fit the flexible distribution model during the EMOS steps for each variable, location, year, month, and lead time separately. In an operational setting, instead of using all dates from the current month and year (which in real time would not be available except on the last day of a given month), one could use the hindcasts associated with the last 8–10 reforecast dates (i.e., if using ECMWF reforecasts which are available twice per week) to fit the model. This ensures that there are enough training data to fit the model that are similar to the current date. The computational expense would also be substantially smaller than what was used in this research study, because in real-time, cross validation would not be necessary.

While this paper focused on the HDWI and its components, other meteorological variables as well as vegetation data could play a role in even more skillful extended-range fire forecasts. We intend to analyze this point in a follow-up project.

Acknowledgments. We are grateful for the many discussions with and suggestions from the fire community. These included several people from U.S. national prediction centers, the USDA Forest Service, and weather forecast offices. We are thankful for access to ECMWF's operational medium-range forecasts provided by Xavier Abellan and for Dave Allured of NOAA/PSD for updating and sharing the ERA-I reanalysis data. We greatly appreciate the two anonymous reviewers that provided suggestions that improved this paper. This research was supported by a grant from a Service Level Agreement between NOAA/ESRL/PSD and NOAA/NWS/CPC on the development of fire-weather related extended range forecast products for the CPC, Project OAR-CPO-NWS-CPC SLA 18-1-007.

REFERENCES

Black, C., Y. Tesfaigzi, J. A. Bassein, and L. A. Miller, 2017: Wildfire smoke exposure and human health: Significant gaps in research for a growing public health issue. *Environ. Toxicol. Pharmacol.*, **55**, 186–195, https://doi.org/10.1016/j.etap.2017.08.022.

Bowman, D. M. J. S., and Coauthors, 2011: The human dimension of fire regimes on Earth. *J. Biogeogr.*, **38**, 2223–2236, https://doi.org/10.1111/j.1365-2699.2011.02595.x.

Bröcker, J., 2012: Evaluating raw ensembles with the continuous ranked probability score. *Quart. J. Roy. Meteor. Soc.*, **138**, 1611–1617, https://doi.org/10.1002/qj.1891.

——, and H. Kantz, 2011: The concept of exchangeability in ensemble forecasting. *Nonlinear Processes Geophys.*, **18**, 1–5, https://doi.org/10.5194/npg-18-1-2011.

Buizza, R., M. Milleer, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908, https://doi.org/10.1002/qj.49712556006.

——, P. L. Houtekamer, G. Pellerin, Z. Toth, Y. Zhu, and M. Wei, 2005: A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems. *Mon. Wea. Rev.*, **133**, 1076–1097, https://doi.org/10.1175/MWR2905.1.

Dawid, A. P., 1984: Present position and potential developments: Some personal views: Statistical theory: The prequential approach. *J. Roy. Stat. Soc. Ser. Gen.*, **147**, 278–292, https://doi.org/10.2307/2981683.

Dee, D. P., and Coauthors, 2011: 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.

Di Giuseppe, F., F. Pappenberger, F. Wetterhall, B. Krzeminski, A. Camia, G. Libertá, and J. San Miguel, 2016: The potential predictability of fire danger provided by numerical weather prediction. *J. Appl. Meteor. Climatol.*, **55**, 2469–2491, https://doi.org/10.1175/JAMC-D-15-0297.1.

ECMWF, 2019: Cycle 43r3. ECWMF, accessed 24 May 2019, https://www.ecmwf.int/en/forecasts/documentation-and-support/evolution-ifs/cycles/cycle-43r3.

Epstein, E. S., 1988: Long-range weather prediction: Limits of predictability and beyond. *Wea. Forecasting*, **3**, 69–75, https://doi.org/10.1175/1520-0434(1988)003<0069:LRWPLO>2.0.CO;2.

Eugenio, F. C., A. R. dos Santos, B. D. Pedra, J. E. Macedo Pezzopane, R. G. Mafia, E. B. Loureiro, L. D. Martins, and N. S. Saito, 2019: Causal, temporal and spatial statistics of wildfires in areas of planted forests in Brazil. *Agric. For. Meteor.*, **266–267**, 157–172, https://doi.org/10.1016/j.agrformet.2018.12.014.

Flannigan, M. D., and J. B. Harrington, 1988: A study of the relation of meteorological variables to monthly provincial area burned by wildfire in Canada (1953–80). *J. Appl. Meteor.*, **27**, 441–452, https://doi.org/10.1175/1520-0450(1988)027<0441:ASOTRO>2.0.CO;2.

——, B. J. Stocks, and B. M. Wotton, 2000: Climate change and forest fires. *Sci. Total Environ.*, **262**, 221–229, https://doi.org/10.1016/S0048-9697(00)00524-6.

——, K. A. Logan, B. D. Amiro, W. R. Skinner, and B. J. Stocks, 2005: Future area burned in Canada. *Climatic Change*, **72**, 1–16, https://doi.org/10.1007/s10584-005-5935-y.

——, A. S. Cantin, W. J. de Groot, M. Wotton, A. Newbery, and L. M. Gowman, 2013: Global wildland fire season severity in the 21st century. *For. Ecol. Manage.*, **294**, 54–61, https://doi.org/10.1016/j.foreco.2012.10.022.

Gabbert, B., 2012: Update on the Heartstrong fire in Colorado. *Wildfire Today*, accessed 4 September 2019, https://wildfiretoday.com/2012/03/19/update-on-heartstrong-fire-in-colorado/.

Gneiting, T., and A. E. Raftery, 2007: Strictly proper scoring rules, prediction, and estimation. *J. Amer. Stat. Assoc.*, **102**, 359–378, https://doi.org/10.1198/016214506000001437.

——, and M. Katzfuss, 2014: Probabilistic forecasting. *Annu. Rev. Stat. Appl.*, **1**, 125–151, https://doi.org/10.1146/annurev-statistics-062713-085831.

——, A. E. Raftery, A. H. Westveld, and T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Wea. Rev.*, **133**, 1098–1118, https://doi.org/10.1175/MWR2904.1.

——, F. Balabdaoui, and A. E. Raftery, 2007: Probabilistic forecasts, calibration and sharpness. *J. Roy. Stat. Soc. Series B Stat. Methodol.*, **69**, 243–268, https://doi.org/10.1111/j.1467-9868.2007.00587.x.

Grimit, E. P., T. Gneiting, V. J. Berrocal, and N. A. Johnson, 2006: The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification. *Quart. J. Roy. Meteor. Soc.*, **132**, 2925–2942, https://doi.org/10.1256/qj.05.235.

Hamill, T. M., and S. J. Colucci, 1997: Verification of Eta–RSM short-range ensemble forecasts. *Mon. Wea. Rev.*, **125**, 1312–1327, https://doi.org/10.1175/1520-0493(1997)125<1312:VOERSR>2.0.CO;2.

HDWI, 2019: The Hot-Dry-Windy Index (HDW). Accessed 23 May 2019, https://www.hdwindex.org/.

Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, **15**, 559–570, https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

Hudson, D., O. Alves, H. H. Hendon, and A. G. Marshall, 2011: Bridging the gap between weather and seasonal forecasting: intraseasonal forecasting for Australia. *Quart. J. Roy. Meteor. Soc.*, **137**, 673–689, https://doi.org/10.1002/qj.769.

Hutto, R. L., 2008: The ecological importance of severe wildfires: Some like it hot. *Ecol. Appl.*, **18**, 1827–1834, https://doi.org/10.1890/08-0895.1.

Jones, P. W., 1999: First- and second-order conservative remapping schemes for grids in spherical coordinates. *Mon. Wea. Rev.*, **127**, 2204–2210, https://doi.org/10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2.

Krzysztofowicz, R., 2001: The case for probabilistic forecasting in hydrology. *J. Hydrol.*, **249**, 2–9, https://doi.org/10.1016/S0022-1694(01)00420-6.

Leighly, J., 1937: A note on evaporation. *Ecology*, **18**, 180–198.

McDonald, J. M., A. F. Srock, and J. J. Charney, 2018: Development and application of a Hot-Dry-Windy Index (HDW) climatology. *Atmosphere*, **9**, 285, https://doi.org/10.3390/atmos9070285.

McGee, T., B. McFarlane, and C. Tymstra, 2015: Wildfire: A Canadian perspective. *Wildfire Hazards, Risks and Disasters*, J. F. Shroder and D. Paton, Eds., Elsevier, 35–58 https://doi.org/10.1016/B978-0-12-410434-1.00003-8.

Moreira, F., and G. Pe'er, 2018: Agricultural policy can reduce wildfires. *Science*, **359**, 1001–1001, https://doi.org/10.1126/science.aat1359.

NIFC, 2019: Total wildland fires and acres (1926-2017). Accessed 23 May 2019, https://www.nifc.gov/fireInfo/fireInfo_stats_totalFires.html.

NRCAN, 2019: Forest area burned and number of forest fires in Canada, 2007-2017. Accessed 23 May 2019, https://www.nrcan.gc.ca/forests/report/disturbance/16392.

Palaiologou, P., A. A. Ager, M. Nielsen-Pincus, C. R. Evers, and K. Kalabokidis, 2018: Using transboundary wildfire exposure assessments to improve fire management programs: A case study in Greece. *Int. J. Wildland Fire*, **27**, 501–513, https://doi.org/10.1071/WF17119.

Radeloff, V. C., and Coauthors, 2018: Rapid growth of the US wildland-urban interface raises wildfire risk. *Proc. Natl. Acad. Sci. USA*, **115**, 3314–3319, https://doi.org/10.1073/pnas.1718850115.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, **133**, 1155–1174, https://doi.org/10.1175/MWR2906.1.

Reid, C. E., B. Michael, F. H. Johnston, J. Michael, J. R. Balmes, and C. T. Elliott, 2016: Critical review of health impacts of wildfire smoke exposure. *Environ. Health Perspect.*, **124**, 1334–1343, https://doi.org/10.1289/ehp.1409277.

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1058, https://doi.org/10.1175/2010BAMS3001.1.

Schefzik, R., 2016: A similarity-based implementation of the Schaake Shuffle. *Mon. Wea. Rev.*, **144**, 1909–1921, https://doi.org/10.1175/MWR-D-15-0227.1.

——, T. L. Thorarinsdottir, and T. Gneiting, 2013: Uncertainty quantification in complex simulation models using ensemble copula coupling. *Stat. Sci.*, **28**, 616–640, https://doi.org/10.1214/13-STS443.

Scheuerer, M., and D. Möller, 2015: Probabilistic wind speed forecasting on a grid based on ensemble model output

statistics. *Ann. Appl. Stat.*, **9**, 1328–1349, https://doi.org/10.1214/15-AOAS843.

——, T. M. Hamill, B. Whitin, M. He, and A. Henkel, 2017: A method for preferential selection of dates in the Schaake shuffle approach to constructing spatiotemporal forecast fields of temperature and precipitation. *Water Resour. Res.*, **53**, 3029–3046, https://doi.org/10.1002/2016WR020133.

Sharples, J. J., and Coauthors, 2016: Natural hazards in Australia: Extreme bushfire. *Climatic Change*, **139**, 85–99, https://doi.org/10.1007/s10584-016-1811-1.

Short, K. C., 2017: Spatial wildfire occurrence data for the United States, 1992-2015. [FPA_FOD_20170508] (4th ed.), Forest Service Research Data Archive, https://doi.org/10.2737/RDS-2013-0009.4.

Shvidenko, A. Z., and D. G. Schepaschenko, 2013: Climate change and wildfires in Russia. *Contemp. Probl. Ecol.*, **6**, 683–692, https://doi.org/10.1134/S1995425513070010X.

Simard, A. J., 1968: The moisture content of forest fuels—I: A review of the basic concepts. Department of Forestry and Rural Development, 50 pp., http://www.cfs.nrcan.gc.ca/bookstore_pdfs/24782.pdf.

SPC, 2019: Probability of wildfire ≥ 100 acres within 25 miles (1992-2015). Accessed 3 September 2019, https://www.spc.noaa.gov/new/FWclimo/climo.php?parm=100ac.

Srock, A. F., J. J. Charney, B. E. Potter, and S. L. Goodrick, 2018: The Hot-Dry-Windy Index: A new fire weather index. *Atmosphere*, **9**, 279, https://doi.org/10.3390/atmos9070279.

Stauffer, R., N. Umlauf, J. W. Messner, G. J. Mayr, and A. Zeileis, 2017: Ensemble postprocessing of daily precipitation sums over complex terrain using censored high-resolution standardized anomalies. *Mon. Wea. Rev.*, **145**, 955–969, https://doi.org/10.1175/MWR-D-16-0260.1.

Taillardat, M., O. Mestre, M. Zamo, and P. Naveau, 2016: Calibrated ensemble forecasts using quantile regression forests and ensemble model output statistics. *Mon. Wea. Rev.*, **144**, 2375–2393, https://doi.org/10.1175/MWR-D-15-0260.1.

Tedim, F., and Coauthors, 2018: Defining extreme wildfire events: Difficulties, challenges, and impacts. *Fire*, **1**, 9, https://doi.org/10.3390/fire1010009.

Thorarinsdottir, T. L., and T. Gneiting, 2010: Probabilistic forecasts of wind speed: Ensemble model output statistics by using heteroscedastic censored regression. *J. Roy. Stat. Soc. Ser. A Stat. Soc.*, **173**, 371–388, https://doi.org/10.1111/j.1467-985X.2009.00616.x.

Thornthwaite, C. W., 1940: Atmospheric moisture in relation to ecological problems. *Ecology*, **21**, 17–28, https://doi.org/10.2307/1930614.

Westerling, A. L., H. G. Hidalgo, D. R. Cayan, and T. W. Swetnam, 2006: Warming and earlier spring increase western U.S. forest wildfire activity. *Science*, **313**, 940–943, https://doi.org/10.1126/science.1128834.

White, C. J., and Coauthors, 2017: Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteor. Appl.*, **24**, 315–325, https://doi.org/10.1002/met.1654.

World Meteorological Organization (WMO), 2012: Guidelines on ensemble prediction systems and forecasting. WMO, 32 pp., https://library.wmo.int/pmb_ged/wmo_1091_en.pdf.

Worsnop, R. P., M. Scheuerer, T. M. Hamill, and J. K. Lundquist, 2018: Generating wind power scenarios for probabilistic ramp event prediction using multivariate statistical post-processing. *Wind Energy Sci.*, **3**, 371–393, https://doi.org/10.5194/wes-3-371-2018.