Random Forest Approach for Improving Nonconvective High Wind Forecasting across Southeast Wyoming

Matthew D. Brothers[a] and Christopher L. Hammer[a]

[a] *NOAA/NWS Weather Forecast Office, Cheyenne, Wyoming*

ABSTRACT: High winds are one of the key forecast challenges across southeast Wyoming. The complex mountainous terrain across the region frequently results in strong gap winds in localized areas, as well as more widespread bora and chinook winds in the winter season (October–March). The predictors and general weather patterns that result in strong winds across the region are well understood by local forecasters. However, no single predictor provides notable skill by itself in separating warning-level events from others. Random forest (RF) classifier models were developed to improve upon high wind prediction using a training dataset constructed of archived observations and model parameters from the North American Regional Reanalysis (NARR). Three locations were selected for initial RF model development, including the city of Cheyenne, Wyoming, and two gap regions along Interstate 80 (Arlington) and Interstate 25 (Bordeaux). Verification scores over two winters suggested the RF models were beneficial relative to current operational tools when predicting warning-criteria high wind events. Three case studies of high wind events provide examples of the RF models' effectiveness to forecast operations over current forecast tools. The first case explores a classic, widespread high wind scenario, which was well anticipated by local forecasters. A more marginal scenario is explored in the second case, which presented greater forecast challenges relating to timing and intensity of the strongest winds. The final case study carefully uses Global Forecast System (GFS) data as input into the RF models, further supporting real-time implementation into forecast operations.

KEYWORDS: Complex terrain; Wind; Operational forecasting; Machine learning

1. Introduction

Locally enhanced winds due to complex topographic features have been observed and studied across mountainous regions around the world. Studies include foehn winds in the European Alps (Drechsel and Mayr 2008; Sprenger et al. 2017), bora winds near the Adriatic Sea (Smith 1987; Alpers et al. 2009), downslope and gap winds near the Strait of Juan de Fuca (Colle and Mass 2000), and the "Hirodo-kaze" downslope wind near the base of Mount Nagi, Japan (Fudeyasu et al. 2008). Most of these complex topographic features are on the subgrid scale of global numerical weather prediction (NWP); therefore, appropriate meteorological conditions are often not resolved accurately. This is still true for mesoscale NWP with fine-scale topographic features remaining unresolved, resulting in underforecast winds (Collins et al. 2020). Multiple NWP postprocessing techniques have been developed to address these issues with the prediction of locally enhanced winds (Lindsey et al. 2011; Sprenger et al. 2017).

Strong winds are one of the key forecast challenges across southeast Wyoming during the winter months. This region includes nearly 400 combined miles of Interstates 80 and 25, which are among the busiest stretches of interstate in the United States with regard to commercial truck traffic. The complex mountainous terrain across the region frequently results in strong gap winds in localized areas, as well as more widespread bora and chinook (foehn) winds in the winter

season (October–March). The local high wind criteria include sustained winds of 35 kt (18.01 m s^{-1} or 40 mph) for one hour and/or wind gusts in excess of 50 kt (25.72 m s^{-1} or 58 mph). These wind speeds are known to cause frequent blow overs of light, high-profile vehicles, including campers and commercial trailers. In some cases, property damage has occurred due to wind gusts in excess of 65 kt (33.43 m s^{-1} or 75 mph).

Figure 1 shows a topographical map of southern Wyoming, which is a low point in the Rocky Mountain region at elevations around 1981–2286 m (6500–7500 ft). A narrower channel exists between the Sierra Madre and Snowy Ranges, and the Shirley/Ferris/Seminoe Mountains that industry has targeted for wind turbine placement over the years (Martner and Marwitz 1982). This channel continues through the lower terrain of the Central Laramie Range before emptying out into the high plains of southeast Wyoming and the North Platte River Valley. On an even finer scale, localized narrow canyons and gaps, such as near Arlington and Bordeaux, can further enhance winds and pose a threat to motorists along adjacent roadways.

Gap winds form as statically stable air approaches a mountain barrier and can only pass the obstacle through low-lying areas in terrain. The strongest gap winds are typically observed in the gap exit region as flow spreads out. This creates an area of lower pressure in addition to background synoptic-scale pressure gradients, increasing the pressure gradient force through the gap (Overland and Walter 1981). Additionally, downslope winds into gap channels can further enhance gap wind flow in complex terrain. Mountain waves forming above and downwind of mountain barriers can lead to strong winds along the leeside slopes (Vosper 2004). These are frequent

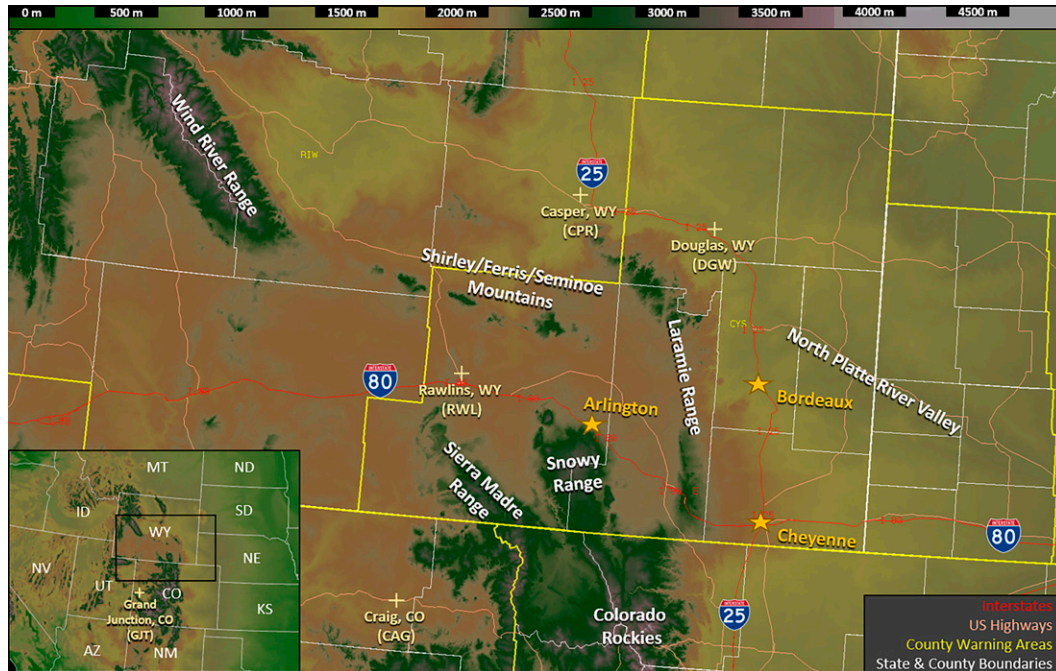*Corresponding author*: Matthew Brothers, matthew.brothers@noaa.gov

Fig. 1. Topographical map of southern Wyoming with select geographical features labeled in addition to the Arlington and Bordeaux gap locations along Interstates 80 and 25, respectively, and the city of Cheyenne.

along the Rocky Mountain Front Range as high pressure builds west of the Continental Divide, creating a strong west–east-oriented mean sea level pressure (MSLP) gradient as noted in previous studies such as Cotton et al. (1995). These strong MSLP gradients can create favorable conditions for bora winds on the leeside slopes of the Rocky Mountains. Bora winds are a type of downslope wind that occur due to strong cold air advection over a mountain barrier, and are enhanced by the higher density of the colder upstream air mass (Smith 1987; Alpers et al. 2009). This is similar to a foehn (chinook) wind, but the resulting adiabatic warming is offset by the presence of cold air advection aloft (Defant 1951; Brinkmann 1974; Reed 1981).

a. Summary of current operational tools

The predictors and general weather patterns that result in strong winds are well understood. Local forecasters use low-level height gradients, winds aloft, omega (vertical motion) fields, and low-level lapse rates to identify such threats. However, no single predictor provides notable skill by itself in separating warning-level events from others. Local studies at the Cheyenne National Weather Service (NWS) office have developed tools for high wind prediction dating back to the 1970s. An early 2000s local study analyzed the correlation between height gradients and wind gusts at Arlington (KARL). More specifically, a height gradient from Craig, Colorado (CAG), to Casper, Wyoming (CPR), was determined to be effective for high wind prediction at Arlington, using a threshold of 50–60 m at both the 700- and 850-mb levels (1 mb = 1 hPa). Arlington averages 38.6 days each winter with wind gusts exceeding high wind criteria and 2.2 days with wind gusts exceeding 75 mph.

A 2015 local study developed logistic regression models similar to Lindsey et al. (2011) to assign probabilities of exceeding 58-mph wind gusts at Cheyenne (KCYS) and Bordeaux (KBRX) (Hammer 2015; Finch and Lindsey 2016). Bordeaux averages 30.5 days each winter with wind gusts exceeding high wind criteria and 2.9 days with wind gusts exceeding 75 mph. High winds are less frequent at Cheyenne, which averages 7.6 days each winter with wind gusts exceeding high wind criteria. The Bordeaux logistic regression model was trained using the 850-mb CAG–CPR height gradient (m), ARL–BRX MSLP gradient (mb), and 800-mb wind speed (kt; 1 kt $\approx$ 0.51 m s^{-1}). The Cheyenne logistic regression model uses the MSLP gradient (mb) from Grand Junction, Colorado (GJT), to CYS, 725-mb omega ($-\mu$b s^{-1}), surface (SFC)–650-mb lapse rate (°C km^{-1}), and 650-mb wind speed (kt) and direction (°). The models helped to improve forecast skill, but still have shortcomings with differentiating wind events that approach local high wind criteria and those actually exceeding the 58-mph wind gust threshold. An optimal probability of 26% was originally obtained for the Bordeaux logistic regression model, meaning this threshold had the best skill when the model probability was used as the sole trigger point for high wind prediction for cases in the validation dataset. For Cheyenne, an optimal probability of 12% was obtained. These logistic regression models were trained using North American Regional Reanalysis (NARR) data and then implemented into real-time operations using Global Forecast System (GFS) data.

b. Problem statement

Machine learning techniques for weather and hazard forecasting have recently been a growing topic of interest (McGovern

et al. 2017). Random forest (RF) models have been developed to create severe weather outlooks and excessive rainfall outlooks similar to those products issued by the Storm Prediction Center and Weather Prediction Center (Herman and Schumacher 2018a,b; Hill et al. 2020; Schumacher et al. 2021). Researchers at the National Severe Storms Laboratory (NSSL) used machine learning techniques to develop the New Tornado Detection Algorithm and New Mesocyclone Detection Algorithm to assist with warning operations and tested these tools as part of the Hazardous Weather Testbed (Sandmael et al. 2020). Several NWS forecast offices have been developing machine learning tools for fog detection on webcams (Wilson and Cote 2020), predicting the development of radiation fog at Terminal Aerodrome Forecast (TAF) sites (Carr et al. 2020), and predicting mountain pass closures and potential car crashes (Dang 2020). Winds have also become a recent candidate for machine learning applications because of NWP limitations on the finer scale. This includes machine learning methods for postprocessing of ensemble wind gusts (Schulz and Lerch 2022), and identification of foehn winds in the Reuss Valley in Switzerland using an AdaBoost algorithm (Sprenger et al. 2017).

This study explores the use of machine learning techniques, specifically RFs, to improve upon existing local tools for predicting nonconvective high winds and increase forecaster confidence when issuing high wind warnings. The goal is to not only improve forecast accuracy and lead time for high wind events, but also to develop a tool that can enhance decision support messaging for high wind events to core partners and the public. Additionally, this study can serve as a proof of concept to be applied at other wind-prone areas where high-resolution forecast models still severely underforecast strong winds.

2. Data and methods

a. Initial site selection

The initial exploration into RF modeling focuses on three primary locations in southeast Wyoming: Arlington, Bordeaux, and Cheyenne. Each site is a unique wind microclimate, influenced by different physical and dynamic processes. Gap winds are common occurrences at both Arlington and Bordeaux, while the city of Cheyenne primarily experiences downslope windstorms, including both chinook and bora-type events. Bordeaux is in a unique position where it experiences both gap and downslope winds. All three locations are prone to severe mountain wave activity. These key differences will be important to lay the foundation for predicting high winds in multiple climate and terrain regimes. These three sites were also the focus of numerous past local studies (Hammer 2015; Finch and Lindsey 2016), which led to a variety of forecasting tools still used extensively in operations today. This existing guidance will be paramount in providing a critical baseline comparison to determine the validity of the proposed techniques, and may help to raise forecaster confidence in future tools.

b. Period of study

The selected period of study covers ten winters from 1 October 2010 to 31 March 2020. Figure 2 shows the frequency of high wind occurrences for all three sites based on hour of the day and month of the year, using the observational data described in section 2c. Nonconvective high wind occurrences generally peak between December and February for the three sites of interest, which was also seen across southeast Wyoming in Klink (1999). The warm-season months of April–September were not considered in order to limit potential contamination from convectively induced high wind events, which are extremely difficult to differentiate without an extensive analysis of radar and satellite data.

c. Observational data and classification of winds

Wind observations for the Cheyenne Airport Automated Surface Observing System (ASOS) were obtained using archived Meteorological Aerodrome Reports (METAR) from the Iowa Environmental Mesonet (IEM). Peak wind gusts were often found to be observed in between routine observations taken at 5-min or 1-h intervals. This was addressed by parsing the maximum peak wind (PK WND) from all METARs reported in a given time range. If no individual METARs reported a PK WND, the peak wind gust was set to the highest observed sustained wind speed for that period. The Wyoming Department of Transportation owns and operates the observation sites at Arlington and Bordeaux. Wind data for non-ASOS sites such as these were collected from Utah MesoWest (Horel et al. 2002). These sites typically report every 2–5 min, an acceptable time scale to consider individual observations representative of local peak winds. Time matching was necessary to accurately compare wind observations to the predictor data from the NARR. The observational data were split into 3-hourly time blocks by analyzing data ±90 min from the official NARR times. For instance, NARR data for 1800 UTC would be matched to observations occurring between 1630 and 1930 UTC. Calculations of maximum wind gust and maximum 1-h sustained winds were made for each time block. Based on these calculations, each time block was then assigned a wind classification defined in Table 1. The high wind classification represents local high wind criteria. While not a warning-level event, elevated winds are still significant to forecast operations and provide a necessary step between the high wind and none classification. The wind speeds and gusts defining the elevated classification more similarly represent the NWS wind advisory criteria. Additionally, including an elevated classification helps to alleviate the class imbalance observed in the dataset.

d. North American Regional Reanalysis data

Meteorological parameters at multiple pressure levels were obtained from the NARR (Mesinger et al. 2006) for 3-hourly intervals throughout the period of study to develop a large initial dataset. The data were extracted from the closest grid point on the 32-km NARR grid to each location, including sea level pressures, winds, heights, temperatures, and omega at all available pressure levels. This information was then used to derive new parameters such as height and sea level pressure gradients between points, temperature lapse rates, and vertical potential temperature gradients to expand upon
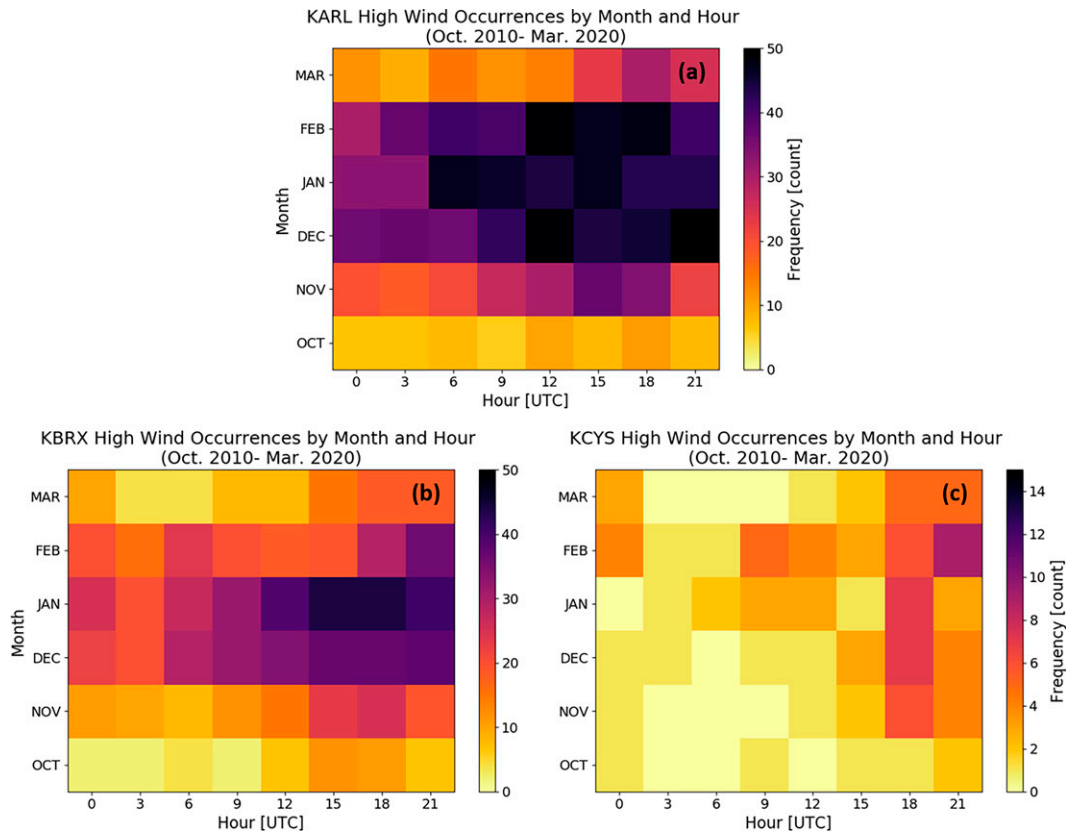
FIG. 2. Heat map of frequency of high wind occurrences (wind gusts ≥ 58 mph and/or 1-h sustained winds ≥ 40 mph) at (a) Arlington, (b) Bordeaux, and (c) Cheyenne based on hour (UTC) of the day (*x* axis) and month of the year (*y* axis). Note the scale change in (c) with the maximum value at 15 compared to 50 for (a) and (b).

the dataset of predictors to select from for each RF model. Section 2e details methods used for specific predictor selection. The height and pressure gradients are of particular importance, especially for gap flows as it implies the character of the general weather pattern, including the position and strength of areas of high and low pressure. Various gradients were selected for investigation based on gap orientations, terrain, and synoptic climatology. The 700-mb height gradient between CAG and CPR is an example of a derived predictor between points. Predictors derived from vertical profiles, such as the surface–700-mb temperature lapse rate (°C km^{-1}), were also important to imply the thermodynamic profile at a given point and time in the absence of a sounding dataset.

Since the NARR is a reanalysis dataset, a separate system was needed to deploy these models into real-time operations

for forecasters. Past studies for current operational tools used GFS data for real-time operations with success (Hammer 2015; Finch and Lindsey 2016). Additionally, training on a re-analysis dataset avoids the need to rebuild the dataset and retrain the RF models each time a model system, such as the GFS, completes an update. In real-time operations, GFS data for selected predictors (section 2e) were extracted from the closest grid point to each location on the 13-km grid.

e. Predictor background analysis and selection

The extracted and derived NARR predictors were compared individually against the observed winds to determine the predictability of warning-criteria wind gusts. Predictors for each site's RF model were selected based on a thorough background analysis of correlations and critical success index (CSI) scores of individual predictors at various thresholds. Many meteorological parameters, such as CAG–CPR height gradients, were already well known to be strong predictors of significant winds from previous local studies, while others were entirely experimental based on forecaster experience and general understanding of high winds. CSI scores were calculated [see Eq. (3) in section 2g] and analyzed for all available predictors in the NARR dataset for each site. This included various gradient orientations as well as data for

TABLE 1. Classifications assigned to observational data at 3-hourly intervals based on maximum wind gusts and maximum 1-h sustained winds.

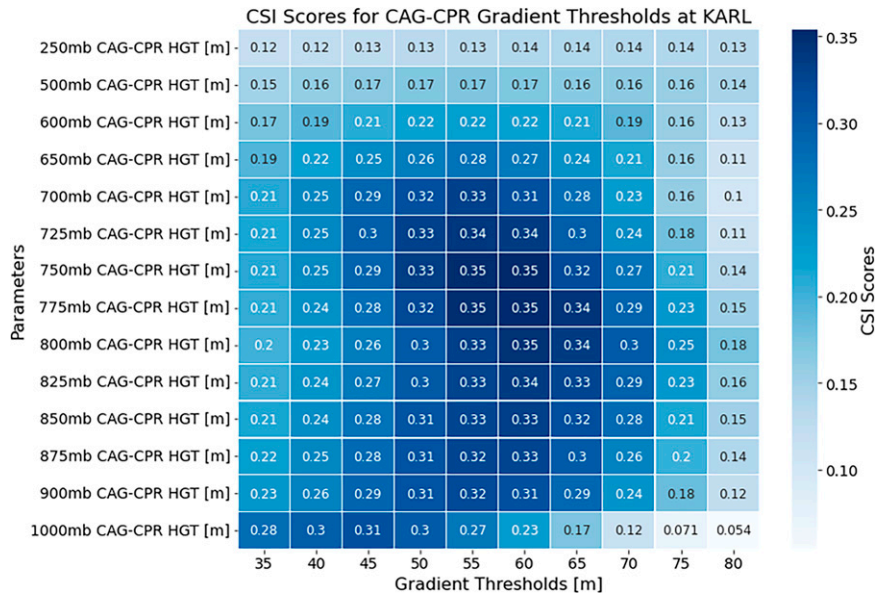| Classification	Wind gust		1-h sustained winds
High wind	≥58 mph	and/or	≥40 mph
Elevated	45–57 mph	and/or	35–39 mph
None	<45 mph	and	<35 mph

FIG. 3. Heat map showing CSI scores for predicting high winds at Arlington (KARL) by only considering CAG–CPR height gradients (m) at various thresholds (x axis) and pressure levels (y axis). Darker blue colors indicate higher CSI scores.

multiple pressure levels and layers. The CSI scores were determined under the assumption that an individual predictor threshold is solely used for a "yes" or "no" decision on high wind prediction. Figure 3 shows an example of this single predictor analysis for CAG–CPR height gradient thresholds (m) for multiple pressure levels at Arlington. In this example, we confirm that a 750-mb CAG–CPR gradient exceeding 55–60 m is a strong predictor of high winds at Arlington, and therefore was selected as a predictor for the RF model.

The best predictor candidates for the CSI score analysis were correlated reasonably well (r^2 greater than 0.5) to the observed wind gust without accounting for any other predictor variables. In some cases, CSI scores were not necessarily the best method of determining the value of a predictor. Those predictors with r^2 values of 0.25 or less were generally insufficient to independently predict maximum wind gusts, but could still have value when considering the overall environment. The temperature lapse rate is known to be an important predictor for high winds; as it indicates the degree of low-level mixing as well as the presence of stable layers near mountain tops (Vosper 2004). Temperature lapse rates are not particularly useful as an independent predictor, however, this does not mean they are not valuable or should not be included in a model. Figure 4c for Cheyenne shows that high winds rarely occur when a surface-based temperature inversion is present (i.e., negative lapse rate), while cases increase as lapse rates steepen toward dry adiabatic. Figure 4a for Arlington shows a greater number of cases for lower lapse rates between 0° and 6°C km^{-1}, thus implying the importance of a statically stable layer aloft. These tendencies strongly supported the inclusion of lapse rates in the RF models. Omega also showed very little skill as an independent predictor, but

could still be a valuable consideration in an otherwise favorable high wind environment. The vast majority of high wind cases in southeast Wyoming have been associated with sinking motion, which promotes downward momentum transfer of stronger winds from aloft. This suggests omega is an extremely important predictor for the RF models. Additional standard meteorological parameters, such as 700-mb temperatures and MSLP, were included despite their overall weak correlation (r^2 of 0.05 or less). These predictors generally did not show any distinct tendencies when high winds occurred, but were still considered important as they provide a complete assessment of the atmospheric profile. These analyses and forecaster input provided a diverse and physically sound list of predictors (see Table A1 in the appendix) for the initial development of the RF models discussed herein.

f. Machine learning methods

The Scikit-Learn module in Python (Pedregosa et al. 2011) was used to train a RF classifier model (Breiman 2001) on the dataset to predict the occurrences of "high," "elevated," and "none" wind intensity labels as defined in Table 1. Separate three-class RF classification models were trained using a unique set of predictors for each site: Arlington, Bordeaux, and Cheyenne. RF models are an ensemble of decision trees. Each unique decision tree is trained to identify the best thresholds at which to split predictor data in order to isolate labeled classes. For example, the initial split node of a decision tree will select a predictor, such as the 750-mb CAG–CPR height gradient, to split the data into two separate branches. A threshold for this predictor (such as a value greater than 60 m) is determined, which best isolates the labeled classes into separate subsets of the training data. The subset of the data that
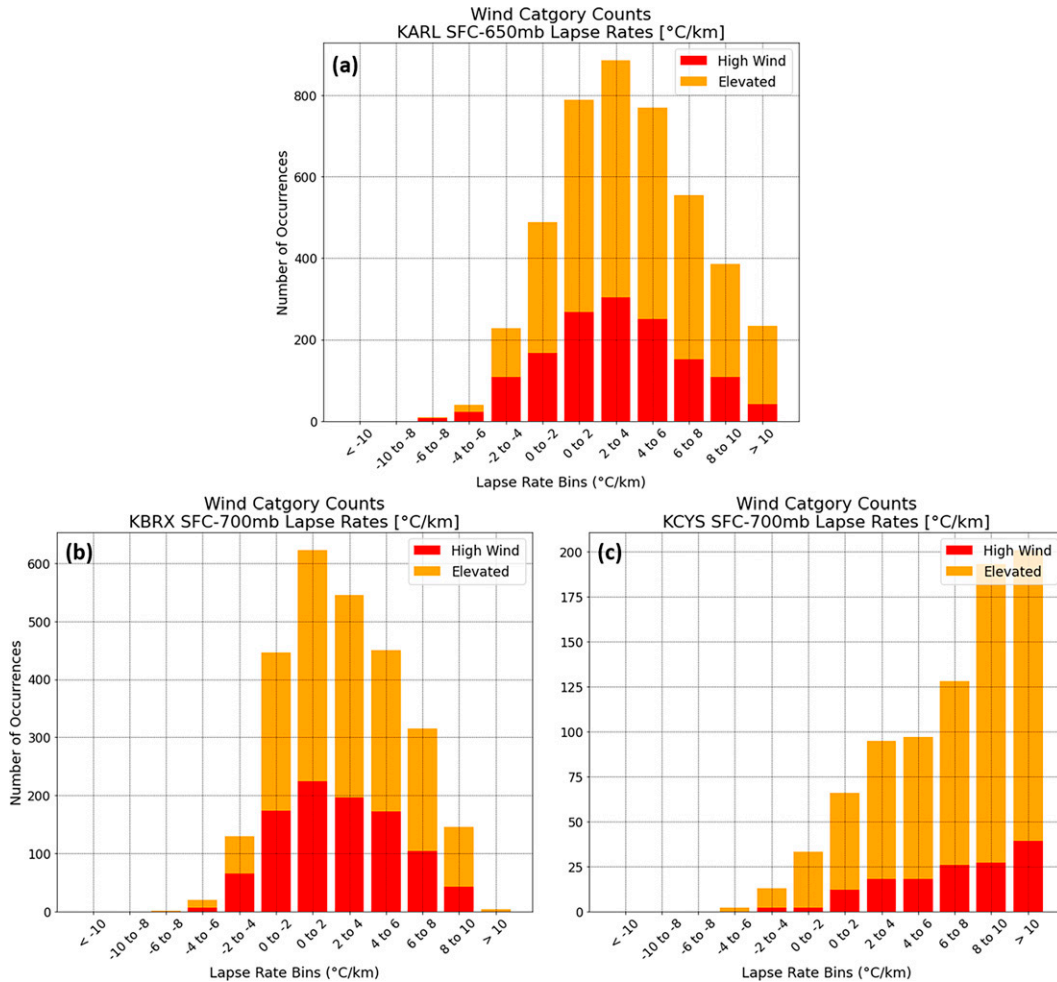
Fig. 4. Stacked histograms of SFC–650-mb lapse rate (°C km^{-1}) at (a) Arlington and SFC–700-mb lapse rate (°C km^{-1}) at (b) Bordeaux and (c) Cheyenne from the NARR for high wind (red) and elevated (orange) observations.

meets the defined threshold follows one branch to the next split node, and the other subset follows the opposite decision tree branch. This process repeats at each split node with new predictors and thresholds until a leaf node is reached. A leaf node is the final branch of the decision tree; and is reached when either a class has been completely isolated, the maximum depth of the tree has been reached, or there are not enough samples in the training data to make another split. After training, new input data are run through each decision tree within the RF classifier model to determine probabilities for each classification. These probabilities are based on the percentage of training samples assigned to each label at the leaf node reached by the new input data. The probabilities are averaged over all the trees in the forest. Deterministic predictions in this study were based on the classification with the highest probability.

The dataset was split into two groups: 1) training dataset covering eight winters of data from 1 October 2010 to 31 March 2018, and 2) testing dataset covering two winters of data from 1 October 2018 to 31 March 2020. The training

dataset was used for an eightfold cross-validation process for hyperparameter tuning, where seven winters of data were used for training and one winter was used for validation eight different times. Hyperparameters are configuration settings for the machine learning algorithm that can be set up to adjust the model's learning process, such as the number of trees or maximum depth of the trees within a random forest. Multiple settings were tested for each individual hyperparameter, using default values for all other hyperparameters to evaluate their influence on the forecast skill using CSI scores [see Eq. (3) in section 2g]. From these results, a confined list of hyperparameter values for each site was compiled for further tuning using a grid search method (Pedregosa et al. 2011). These results can be seen in Table 2, with all tested values listed and the selected hyperparameters for each RF model shown in bold. Overall, the majority of hyperparameters tested from the confined list showed little variation in forecast skill (not shown). Increasing the class weights for high wind showed the most impact by strongly penalizing the RF models for mistakes related to the high wind classification. These hyperparameters

TABLE 2. Number of predictors and list of hyperparameters tested using grid search method for each RF model. Hyperparameters selected for the RF models trained at each site are set bold. Specific details about each hyperparameter can be found in online documentation provided by Scikit-Learn.

Site	No. of predictors	No. of trees	Max tree depth	Min samples at leaf node	Max predictors per split node	High wind class weight
KARL	28	200	6 8 **10** 12	**8** 10 15 30	**5** ("**sqrt**") 8 12	**2** 3
KBRX	27	200	6 8 **10** 12	**6** 8 10 15 30	**5** ("**sqrt**") 8 12	**2** 3
KCYS	23	150	4 6 8 **10**	**8** 10 15 30	4 8 **14**	**5** 7

were used to train the RF model on the entire eight years of data in the training dataset. The model was applied to the testing dataset, containing only unseen data. All verification statistics presented in section 3 are based on model performance using the testing dataset.

g. RF model verification methods

Since the desired predictions from the RF models were classification labels for wind intensity defined in Table 1, the verification statistics used to evaluate the RF models are those used for dichotomous forecasts: probability of detection (POD), false alarm ratio (FAR), critical success index (CSI), and frequency bias. The POD, FAR, CSI, and bias are defined as follows:

$$POD = \frac{hits}{hits + misses}, \qquad (1)$$

$$FAR = \frac{false\ alarms}{hits + false\ alarms}, \qquad (2)$$

$$CSI = \frac{hits}{hits + misses + false\ alarms}, \qquad (3)$$

$$bias = \frac{hits + false\ alarms}{hits + misses}. \qquad (4)$$

These verification scores can be acquired from the contingency table shown in Table 3. Elevated classifications were combined with the "none" classifications in the contingency table. These verification statistics were calculated using two methods: 1) single-observation verification and 2) event-based verification. For single-observation verification, the predicted and observed classifications were compared at only single time steps. For event-based verification, multiple times were grouped together based on occurrence and timing of high winds. High wind observations were identified within the

TABLE 3. Contingency table used for calculating POD, FAR, CSI, and frequency bias verification statistics defined above. "Elevated" and "none" classifications were grouped together since the focus of this study is on prediction of high winds.

	Observed high wind	Observed elevated or none
Forecast high wind	Hit	False alarm
Forecast elevated or none	Miss	Correct negative

dataset, then a search was conducted forward and backward in time until more than six consecutive hours of weak winds with gusts below 45 mph and sustained winds below 35 mph were identified. These times would then be grouped together and considered part of an event, and were scored based on the highest predicted classification over that time range.

Table 4 provides an example of how an event may be identified and evaluated using both methods. A high wind classification was observed at 1500 UTC 16 November 2018, but the RF model predicted an elevated classification. This would count as a miss using single-observation verification. At the next time step, 3 hours later, the RF model predicted a high wind classification, but the observed classification was elevated. This would be a false alarm using single-observation verification. However, if these times were grouped together, then this would be considered a hit as the RF model predicted high wind classification within a short window of a similar observation. This method closely resembles how operational forecasters would issue high wind warnings, and represents the true operational utility.

3. Results and discussion

a. RF model bulk performance statistics

Table 5 shows 3 × 3 contingency tables using single-observation verification for the RF models' predictions over the testing dataset for all three wind classifications. For each site, over 70% of the high wind false alarms issued by the RF models occurred when elevated winds were observed, and therefore would have been considered "close." The majority of missed high wind observations at Arlington were also considered to be close, as over 70% occurred when the RF model prediction was elevated. The opposite was true for the Bordeaux and Cheyenne RF models, as the majority of missed high wind observations occurred with a RF model prediction of none. For the remainder of this section, elevated and none classifications were grouped together, similar to what was shown in Table 3. This was done to evaluate the RF models' performance at predicting whether or not high winds would occur.

Performance diagrams (Roebber 2009) were used for model evaluation, as they display four measures of forecast performance (POD, FAR, CSI, and frequency bias) on a single diagram. Figures 5a and 5b show performance diagrams with verification scores for the RF models (circles) compared to the current operational tools (crosses) for single-observation verification and event-based verification, respectively. Verification scores for current operational tools (700-mb CAG–CPR

TABLE 4. Example high wind event from dataset showing verification for single-observation and event-based methods.

Event ID	Date and time	Observed class	Predicted class	Single-observation verification	Event-based verification
0	0600 UTC 16 Nov 2018	None	None	Correct negative	Correct negative
1	0900 UTC 16 Nov 2018	None	None	Correct negative	Hit
1	1200 UTC 16 Nov 2018	Elevated	None	Correct negative	Hit
1	1500 UTC 16 Nov 2018	**High wind**	Elevated	Miss	Hit
1	1800 UTC 16 Nov 2018	Elevated	**High wind**	False alarm	Hit
1	2100 UTC 16 Nov 2018	None	None	Correct negative	Hit
0	0000 UTC 17 Nov 2018	None	None	Correct negative	Correct negative

height gradients for KARL and logistic regression models for KBRX and KCYS) were based on optimal thresholds defined earlier. The RF trained for Arlington showed the best performance for both verification methods, with CSI scores of 0.666 and 0.384 for the event-based and single-observation methods, respectively. The model for Bordeaux only trailed behind slightly, with an event-based CSI of 0.609 and a single-observation CSI of 0.369. At Cheyenne, the RF model performed much more poorly, with 0.313 event-based and 0.195 single-observation CSI scores, respectively.

Compared to current operational forecast tools, these RF models at Arlington and Bordeaux were more skillful, with higher CSI scores for both single-observation and event-based verification. These results for single-observation verification were statistically significant with 95% confidence, using 1000 random bootstrapped samples with replacement over the testing dataset, while differences for the Cheyenne models were not statistically significant. Additionally, the frequency bias was minimized with the RF models at each site, compared to the overforecasting (KARL) and underforecasting (KBRX and KCYS) current operational tools.

The weaker performance of the RF model at Cheyenne could be attributed to multiple factors. The frequency of high wind occurrences in the training and testing datasets is much

TABLE 5. Contingency table showing predictions and observations of all three wind classifications over the testing dataset (two winters) for the RF models at Arlington, Bordeaux, and Cheyenne.

		Observation		
		High wind	Elevated	None
	KARL			
	High wind	145	84	22
	Elevated	93	257	99
	None	36	243	1902
	KBRX			
	High wind	111	75	29
Forecast	Elevated	31	65	26
	None	55	220	2277
	KCYS			
	High wind	7	12	3
	Elevated	4	10	8
	None	9	101	2757

greater at Arlington and Bordeaux (9.97% and 6.86%, respectively) compared to Cheyenne (1.01%). This results in fewer high wind occurrences in the training dataset for the Cheyenne RF model to learn from, and thus yields a lower chance that a similar event exists within the testing dataset. Additionally, the predictor variables selected for the Cheyenne RF model do not display individual CSI scores as high in the background analysis, compared to those selected for Arlington and Bordeaux. Cheyenne is also primarily impacted by downslope windstorms, which can be more difficult to predict due to a greater number of possible influencing parameters. The gap areas of Arlington and Bordeaux are generally more predictable, as these sites are primarily influenced by pressure gradients and flow aloft.

Reliability diagrams (Bröcker and Smith 2007) were used to assess the uncertainty calibration of the RF models by looking at how often a forecast high wind probability actually occurred. Figures 6a–c show reliability diagrams for high wind probability of the RF models trained at Arlington, Bordeaux, and Cheyenne, respectively. The most frequent forecast for the Arlington model was between 0% and 10%, consisting of 1990 out of 2881 forecast times in the testing dataset, as seen in Fig. 6a. However, the Arlington RF model still produced a considerable number of high wind forecast probabilities in the higher probability bin ranges. The Arlington RF model's high wind probabilities have a slight overforecasting bias, but overall performed fairly well through all bin ranges.

Similar to the Arlington model, the Bordeaux RF model in Fig. 6b shows a slight overforecasting bias of high wind probabilities, but still follows the 1:1 line fairly well throughout the high wind forecast probability bin ranges. The reliability diagram for the Cheyenne RF model suggests the presence of a severe overforecasting issue, as high winds were only observed 40% of the time when RF high wind probabilities were between 60% and 100%. This could also be an artifact of the lack of high wind probability forecasts above 30%. A high wind forecast probability between 0% and 10% was produced 2802 out of 2911 forecast times in the testing dataset, while only 27 forecast times yielded a probability greater than 30%.

Overall, these verification scores provide encouraging insight into the RF model's performance at Arlington and Bordeaux. Despite these encouraging results from a bulk statistics perspective, case studies are needed to truly evaluate the usefulness of each RF model to operational forecasters in select situations. Section 3c will investigate three case studies of events with varying intensities and durations to evaluate the RF model
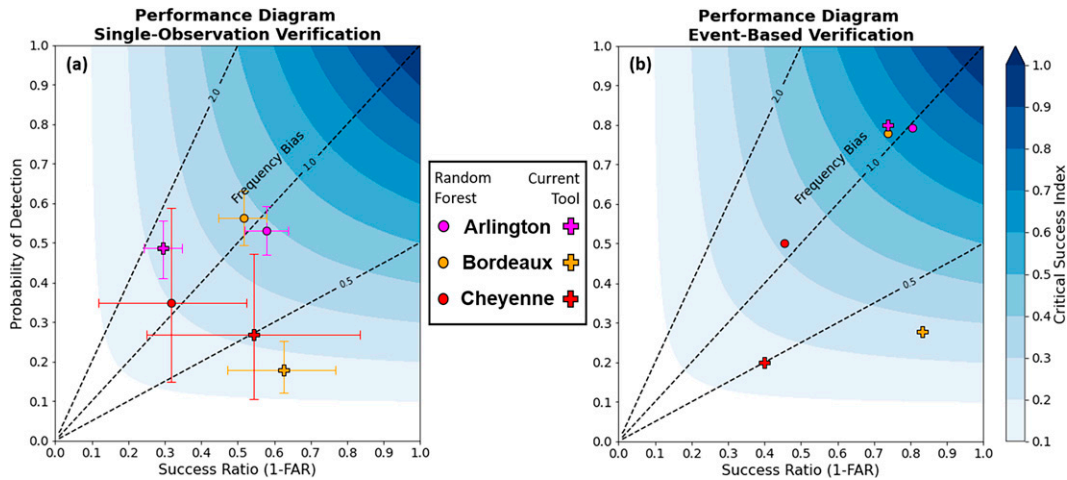
FIG. 5. Performance diagram for (a) single-observation verification and (b) event-based verification methods for the random forest models (circles) and currently used operational guidance (crosses) at three locations: Arlington (magenta), Bordeaux (orange), and Cheyenne (red). Success ratio $(1 - \mathrm{FAR})$ is shown on the x axis, and POD is shown on the y axis, with black dashed lines representing frequency bias. Blue shading shows the CSI, which is maximized in the upper-right-hand corner of the diagrams. Error bars around points on (a) represent 95% confidence interval of POD and success ratio scores using 1000 random bootstrapped samples with replacement over the testing dataset.

performance for each site. Although the skill scores for the initial Cheyenne RF model were less desirable, these case studies will still help evaluate which synoptic patterns the model could be most useful in predicting high winds.

b. RF model feature importance

Feature importances were computed using a single-pass permutation importance method (Pedregosa et al. 2011; McGovern et al. 2019) in order to analyze which predictors provide the most value to each RF model. Permutation importance was configured to calculate the loss in CSI score, as a predictor's data values are rearranged using the single-observation verification method on the deterministic RF model prediction. Rearranging a vital predictor's data values will result in a higher reduction in CSI scores, unless another highly correlated predictor is present. The box-and-whiskers plots in Fig. 7 shows the distribution of CSI score loss as the rearrangement of predictor data occurs over 100 repetitions for each predictor. Figures 7a–c reveal the most significant predictors for the RF models at Arlington, Bordeaux, and Cheyenne, respectively.

The top five predictors for the Arlington RF model include the 700–500-mb lapse rate (°C km^{-1}), 700-mb wind direction (°), 700-mb omega ($-\mu$b s^{-1}), and the CAG–CPR height gradient (m) at both 700- and 750-mb. Previous local research correlated wind gusts to the same CAG–CPR height gradients at the 700- and 850-mb level. These predictors have been used successfully among local forecasters, and inclusion of these predictors as top influencers for the RF model at Arlington makes meteorological sense. The 700–500-mb lapse rates by far show the highest importance, most likely because this level aids in the identification of mountain top inversion layers that

are critical for mountain waves or forcing winds through a gap channel. Lower levels would typically be more important for identifying mountain-top inversions; however, Arlington is located at an elevation of 2369 m (7772 ft), with an average surface pressure of roughly 760 mb. Additionally, Arlington high wind events are typically associated with winds out of the southwest to west, supporting the importance of 700-mb wind direction for the RF model. Enhanced gap flow and mountain wave activity are more likely based on local topography near Arlington with these typical wind directions, assuming other favorable meteorological parameters are present.

The top five predictors for the Bordeaux RF model include the Rawlins (RWL)–Douglas, Wyoming (DGW), MSLP gradient (mb), 750-mb omega ($-\mu$b s^{-1}), 650–850-mb vertical potential temperature difference (K), CAG–DGW MSLP gradient (mb), and the SFC-700-mb lapse rate (°C km^{-1}). While the RWL–DGW MSLP gradient is not positioned directly overhead of the wind gap upstream of Bordeaux, the orientation crossing the Laramie Range resembles the orientation of the gap path. These top predictors are all different from those selected in previous local research when developing a logistic regression model for Bordeaux. The ARL–BRX MSLP gradient, 850-mb CAG–CPR height gradient, and the 800-mb wind speed were used to develop the current operational model used locally. These predictors' (or most similar) importances rank 8th, 17th, and 25th out of 27 predictors. Using a 95% confidence interval of mean CSI score reduction from 1000 bootstrapped samples with replacement, the notion that the ARL–BRX MSLP gradient, 850-mb CAG–CPR height gradient, and the 800-mb wind speed are outside the top five RF model predictors is statistically significant. This continues to show support for the development of the RF model, as predictors
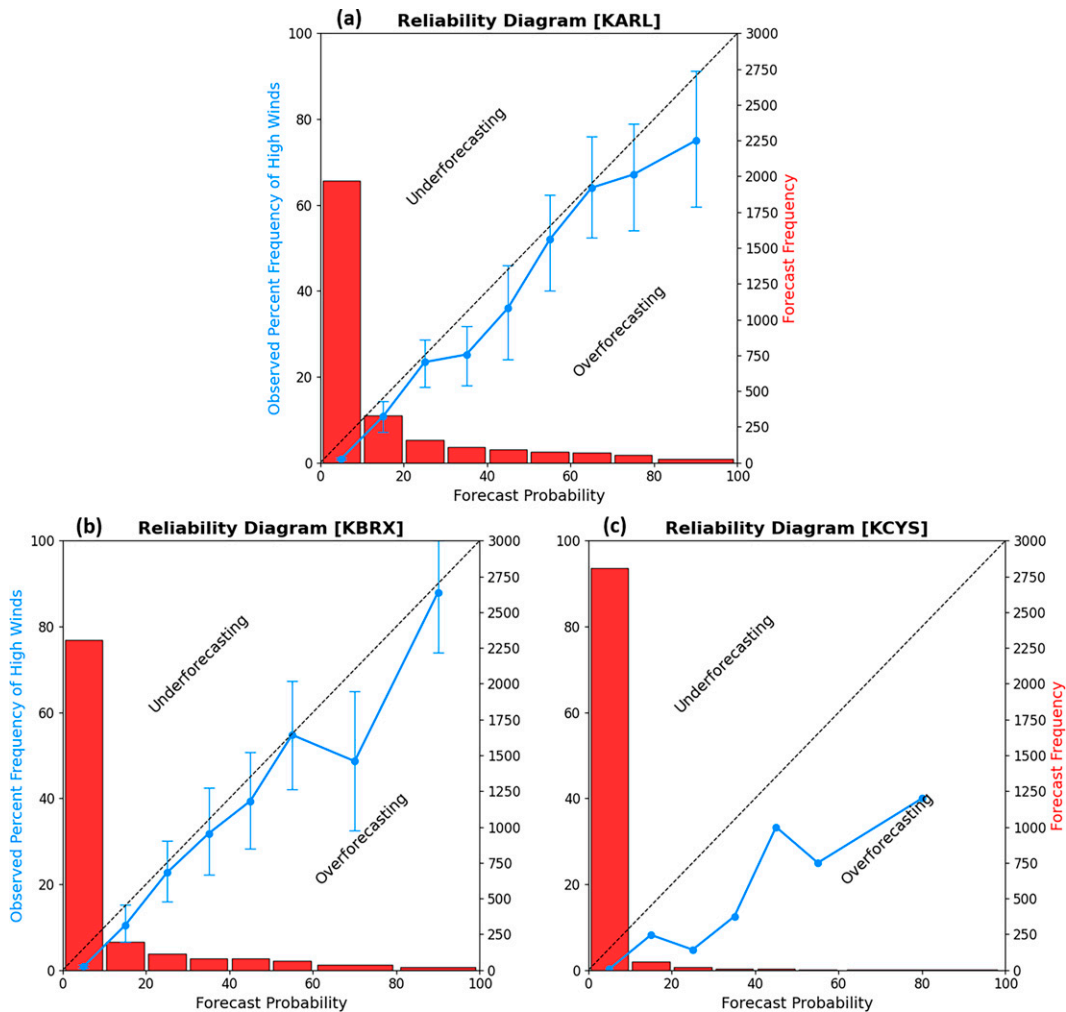
FIG. 6. Reliability diagram for random forest models at (a) Arlington, (b) Bordeaux, and (c) Cheyenne. The line plot (blue) shows the observed percent frequency of high winds for defined forecast probability bin ranges. Error bars (blue) around the line plots in (a) and (b) represent the 95% confidence interval using 1000 random bootstrapped samples with replacement over the testing dataset. Black dashes indicate the 1:1 line, where values above (below) this line indicate RF model high wind probabilities were underforecast (overforecast). The bar plot (red) shows the RF forecast high wind probability frequency based on defined probability bin ranges. Forecast probability bin ranges are represented by the width of the bar plots. Note the error bars were not included on (c) as a result of too few forecast probabilities in the higher probability bin ranges.

previously omitted from the logistic regression model show value in predicting high winds for Bordeaux.

The top five predictors for the Cheyenne model include the 650-mb omega ($-\mu$b s^{-1}), MSLP (mb), GJT–CYS MSLP gradient, 800-mb GJT–CYS height gradient (m), and the hour (time of day). While the distribution of CSI score reductions varies greatly for the hour predictor, the climatology of high wind occurrences strongly suggests that time of day is an important factor (Fig. 2). Typically, strong mixing around midday assists downward momentum transport to the surface. The SFC–700-mb lapse rate comes in as the sixthmost important predictor. Rare nighttime occurrences of high winds at Cheyenne have been associated with mountain waves or strong frontal passages.

Three types of predictors included within the RF models are considered circular variables. These predictors include wind direction (degrees ranging from 1° to 360°), month (standard numerical value ranging from 1 to 12), and hour (numerical value in UTC ranging from 0 to 23). Typical predictor wind directions during high wind events for these sites range between southwest (~225°) and northwest (~315°). Therefore, the discontinuity in this circular predictor between 359° and 1° is rarely a factor. Future work will explore breaking down wind speed and direction into zonal and meridional components of the wind before training the RF models. The month and hour predictors were included in the RF models to try to find the best possible analogs
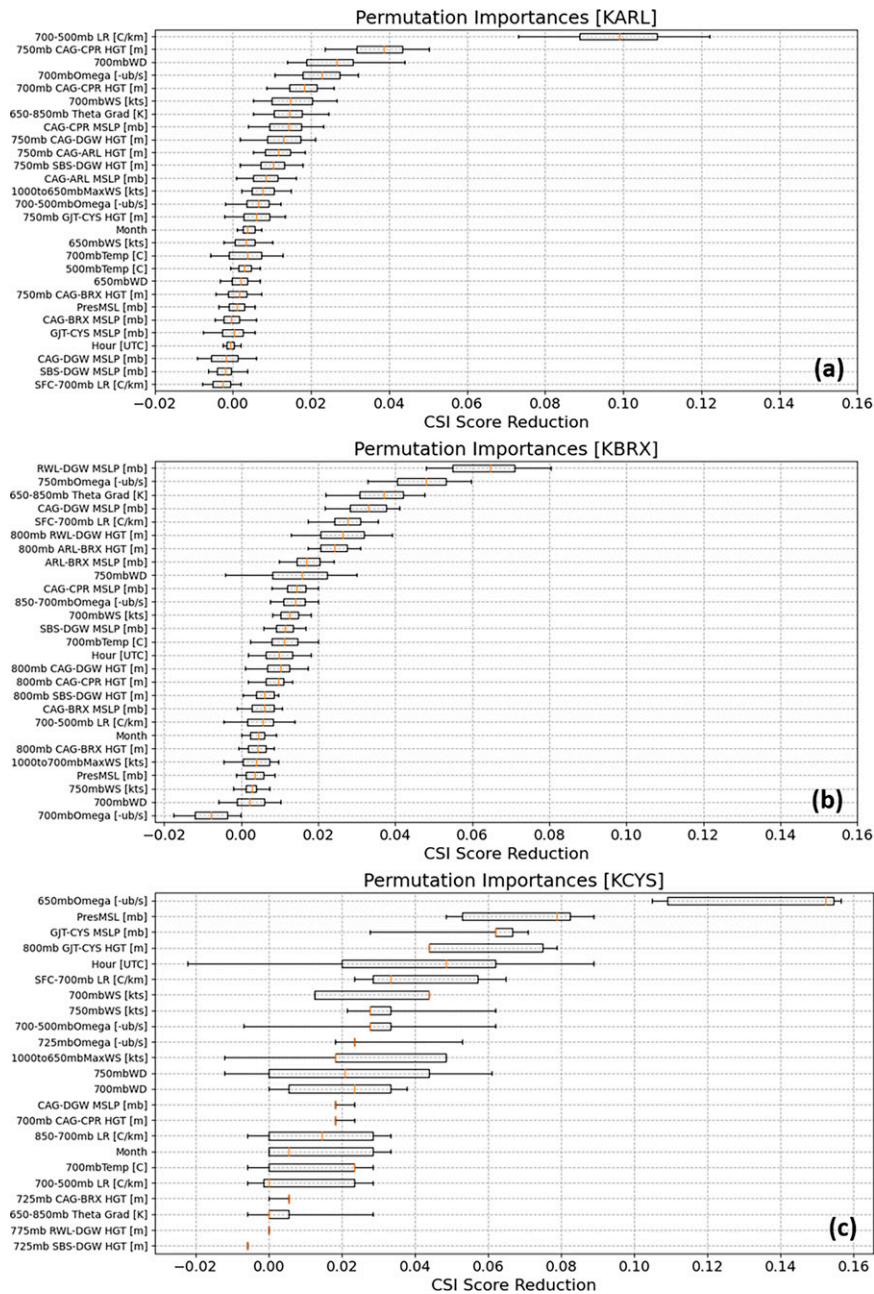
FIG. 7. Box-and-whisker plots for distribution of CSI score reductions when calculating permutation importances over 100 iterations at (a) Arlington, (b) Bordeaux, and (c) Cheyenne. Predictors are ordered along the *y* axis with the most important at the top of the plot. Boxes show the interquartile range between the 25th and 75th percentile, with whiskers representing the 5th and 95th percentiles. The orange line inside the boxes represents the median values.

beyond the meteorological predictor data. However, using the standard numerical value for months results in the point of discontinuity occurring right at the peak of high wind season (December and January), which could be a difficult pattern for the RF models to learn in the training process. Future work will look into rearranging the standard numerical value for months to begin outside of the climatological high wind season. Despite known discontinuities in these predictors, case studies beyond those presented in section 3c suggest these predictors are still valuable to the RF models. Higher high wind probabilities have been noted in numerous cases at Cheyenne around 1800 UTC, which is climatologically the most frequent time of occurrence of high winds at this location (Fig. 2).
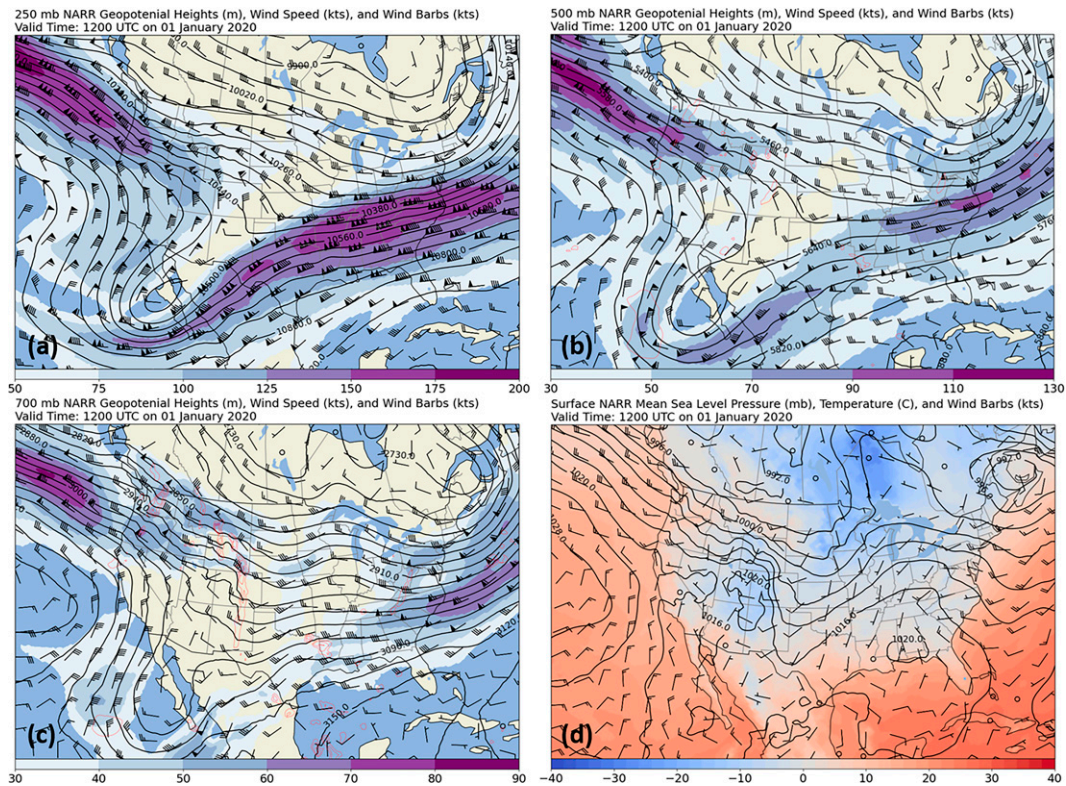
FIG. 8. NARR data from 1200 UTC 1 Jan 2020 at the (a) 250-mb level, (b) 500-mb level, (c) 700-mb level, and (d) surface. Black contours in (a)–(c) show geopotential height (m), while color fill shows wind speed (kt). Red dashed contours in (b) and (c) show omega at intervals of -5, -10, -15, and -30 $-\mu b\ s^{-1}$. Black contours in (d) show MSLP (mb), while color fill shows surface temperature (°C).

c. High wind event case studies

1) CASE 1: 1 JANUARY 2020

A widespread bora-type high wind event with localized extreme gap flows impacted a large portion of southeast Wyoming from 31 December 2019 to 1 January 2020. A significant mid- and upper-level jet was advancing across the northeast Pacific, with 90-kt winds observed at 500 mb over Quillayute, Washington (KUIL), at 0000 UTC 1 January 2020. Figure 8 shows a four-panel plot of NARR data from 1200 UTC 1 January 2020. The general pattern remained quasi-zonal through 1200 UTC, with strong northwesterly flow extending from the Pacific Northwest through the central and northern High Plains regions. Rapid surface cyclogenesis was occurring over southern Saskatchewan and Alberta, with pressures falling into the 990–995-mb range from eastern Montana and northeast Wyoming into the western Dakotas. A 1020-mb surface high was analyzed along the Colorado–Utah border, yielding a very strong (10–15 mb) GJT–CYS MSLP gradient by 1800 UTC. Maximum wind speeds in the surface–650-mb layer ranged from 50 to 60 kt across southeast Wyoming. Strong cold-air advection aloft, coupled with extreme subsidence (downward omega exceeding -15 $-\mu b\ s^{-1}$) in the right-exit region of the upper-level jet streak supported downward momentum transfer of stronger winds from aloft.

This was a classic pattern for strong winds in southeast Wyoming, which were forecast quite well by local meteorologists. A high wind watch was posted for the gap regions on the afternoon of 30 December 2019, roughly 24–36 h ahead of the expected onset of strong winds. These watches would eventually be upgraded to high wind warnings, with the city of Cheyenne added to the warning early in the morning on 1 January 2020.

Figure 9a shows the timeline of wind observations at Arlington compared to the RF model prediction and CAG–CPR height gradients. Height gradient thresholds for Arlington performed quite well for this event, with the 700-mb CAG–CPR gradient climbing to between 50 and 55 m around the time of the first observed 58+ mph wind gust. The strongest winds generally coincided with peak gradients in the low 70s (mph) between 0600 and 1200 UTC 1 January. The gradient remained elevated and even increased further to near 80 m at 1800 UTC, but observed wind gusts decreased below warning criteria during this time. The Arlington RF model predicted winds in the elevated category as early as 1500 UTC 31 December and successfully began predicting high winds at 2100 UTC, earlier than current guidance suggested. The peak wind gusts for this event coincided with the highest probabilities of high winds at Arlington. The RF model continued to predict high winds until 1800 UTC 1 January, which verified

(a)

Arlington (KARL) Forecast Tools vs. Observations	00z Dec 31	03z Dec 31	06z Dec 31	09z Dec 31	12z Dec 31	15z Dec 31	18z Dec 31	21z Dec 31	00z Jan 1	03z Jan 1	06z Jan 1	09z Jan 1	12z Jan 1	15z Jan 1	18z Jan 1	21z Jan 1	00z Jan 2	03z Jan 2	06z Jan 2	09z Jan 2	12z Jan 2
Gust [MPH]	45	46	42	41	37	38	56	60	63	61	63	65	61	57	54	52	46	47	38	37	32
Sustained Winds [MPH]	32	34	33	31	29	25	39	43	41	43	45	46	42	41	40	36	31	34	28	28	23
Random Forest Predicted Class	None	None	None	None	None	Elev	Elev	High Wind	High Wind	High Wind	High Wind	High Wind	High Wind	High Wind	High Wind	Elev	Elev	Elev	Elev	Elev	None
RF Probability of High Wind	1%	5%	3%	13%	2%	25%	37%	74%	71%	75%	93%	91%	87%	87%	92%	43%	37%	19%	22%	18%	2%
700 mb CAG-CPR Gradient [m]	24	28	28	32	23	43	47	51	49	54	70	72	68	70	78	60	54	40	48	50	28

(b)

Bordeaux (KBRX) Forecast Tools vs. Observations	00z Dec 31	03z Dec 31	06z Dec 31	09z Dec 31	12z Dec 31	15z Dec 31	18z Dec 31	21z Dec 31	00z Jan 1	03z Jan 1	06z Jan 1	09z Jan 1	12z Jan 1	15z Jan 1	18z Jan 1	21z Jan 1	00z Jan 2	03z Jan 2	06z Jan 2	09z Jan 2	12z Jan 2
Gust [MPH]	34	39	44	56	57	58	64	58	60	50	73	76	81	81	79	63	45	46	41	39	26
Sustained Winds [MPH]	27	30	37	41	43	47	49	44	44	38	52	53	58	60	55	44	33	35	32	28	20
Random Forest Predicted Class	None	None	None	None	None	High Wind	High Wind	High Wind	High Wind	High Wind	High Wind	High Wind	High Wind	High Wind	High Wind	High Wind	Elev	None	None	None	None
RF Probability of High Wind	1%	10%	10%	23%	14%	55%	53%	56%	55%	74%	84%	95%	95%	91%	54%	29%	16%	18%	8%	1%	2%
Logistic Regression Prob	0%	2%	2%	4%	1%	7%	3%	4%	6%	15%	42%	62%	63%	63%	88%	14%	16%	6%	11%	10%	2%

(c)

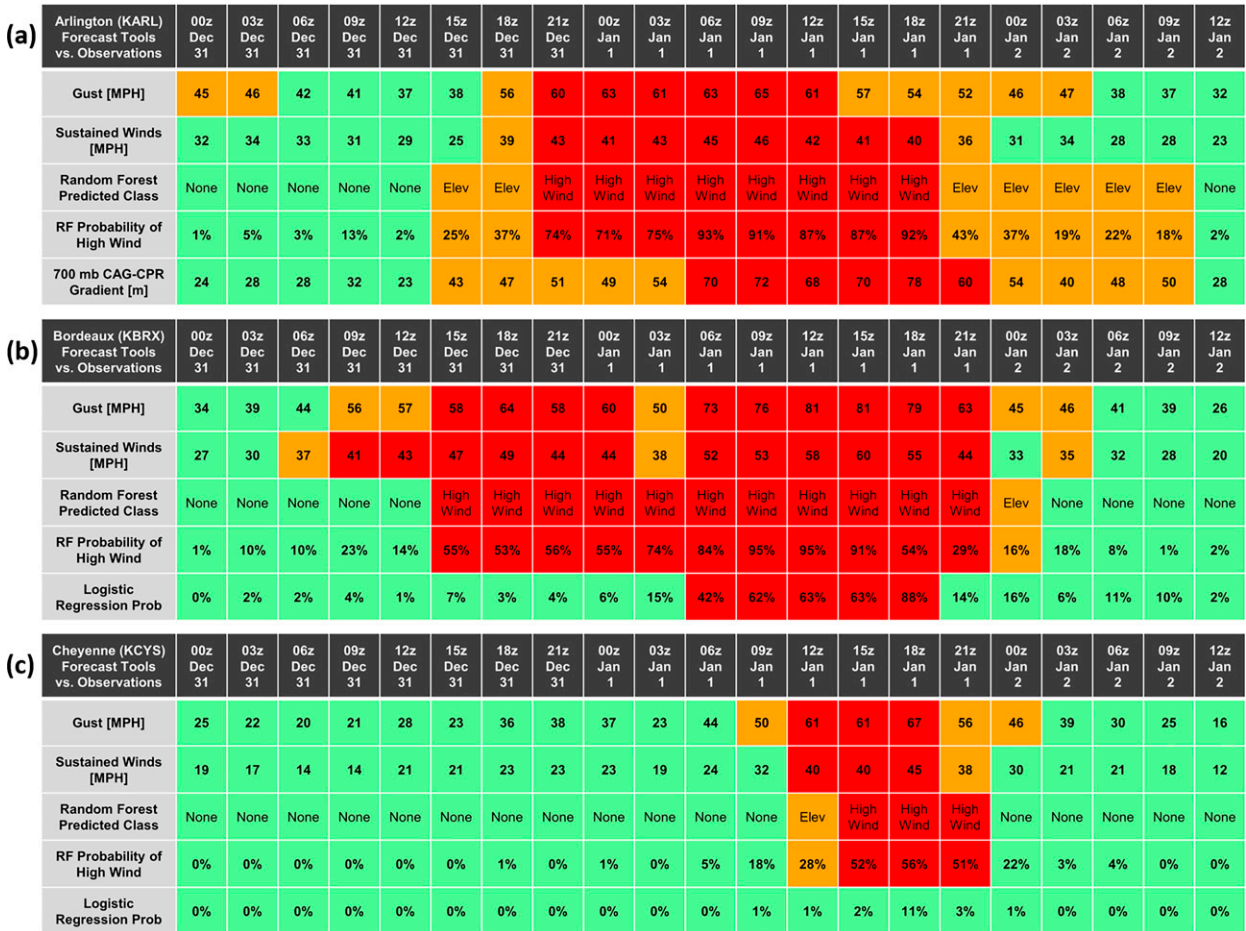| Cheyenne (KCYS) Forecast Tools vs. Observations	00z Dec 31	03z Dec 31	06z Dec 31	09z Dec 31	12z Dec 31	15z Dec 31	18z Dec 31	21z Dec 31	00z Jan 1	03z Jan 1	06z Jan 1	09z Jan 1	12z Jan 1	15z Jan 1	18z Jan 1	21z Jan 1	00z Jan 2	03z Jan 2	06z Jan 2	09z Jan 2	12z Jan 2
Gust [MPH]	25	22	20	21	28	23	36	38	37	23	44	50	61	61	67	56	46	39	30	25	16
Sustained Winds [MPH]	19	17	14	14	21	21	23	23	23	19	24	32	40	40	45	38	30	21	21	18	12
Random Forest Predicted Class	None	None	None	None	None	None	None	None	None	None	None	None	Elev	High Wind	High Wind	High Wind	None	None	None	None	None
RF Probability of High Wind	0%	0%	0%	0%	0%	0%	1%	0%	1%	0%	5%	18%	28%	52%	56%	51%	22%	3%	4%	0%	0%
Logistic Regression Prob	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	1%	2%	11%	3%	1%	0%	0%	0%	0%

FIG. 9. Maximum wind gust (mph) and maximum 1-h sustained winds (mph) observed at (a) Arlington, (b) Bordeaux, and (c) Cheyenne compared to RF model predicted classification and high wind probability over 3-h periods from 0000 UTC 31 Dec 2019 until 1200 UTC 2 Jan 2020. Current operational tools displayed include 700-mb CAG–CPR height gradient (m) in (a) and logistic regression model probabilities in (b) and (c). Values are color coded based on the wind classification they would be associated with as follows: high wind (red), elevated (orange), and none (green).

as 1-h average sustained winds remained around 40 mph until 1800 UTC despite gusts dropping below 58 mph. This was an extremely successful forecast, given the high accuracy of both the start and end times of the event along with the timing of the highest probabilities. The RF model appears to be an improvement on current tools, considering that the correlation study was based entirely on wind gusts and not sustained winds in any capacity.

Figure 9 also shows the timeline of wind observations at (Fig. 9b) Bordeaux and (Fig. 9c) Cheyenne compared to the RF model output (deterministic prediction and high wind probabilities) and logistic regression model probabilities. The classifications for the logistic regression models are based on the optimal probabilities of 12% and 26% for Cheyenne and Bordeaux, respectively. Since the logistic regression models were trained using only a binary classification, only none and high wind classifications are indicated. The peak probability from the Cheyenne logistic regression model approached 11% when the highest wind gusts were observed, but still fell

short of the optimal value required to predict high winds. The model did not show any notable increase until at least 3 h after the first wind gusts to 58+ mph. The Cheyenne RF model correctly predicted a sharp increase in probabilities between 0900 and 1500 UTC 1 January, with warning-criteria winds first observed around 1200 UTC. A high wind classification was predicted between 1500 and 2100 UTC, about 3 h delayed from the actual onset of warning-criteria winds. However, RF classification probabilities at 1200 UTC were similar with a 36% (not shown) and a 28% probability of elevated and high wind classifications, respectively. This means that the RF model was strongly suggesting potential for high winds, even though RF model probabilities favored the lower classification. It seems prudent that forecasters are provided with as much probabilistic information as possible to understand how the model is arriving at its final decision. The maximum RF high wind probability was 56%, and coincided well with the peak sustained winds of 45 mph and gusts to 67 mph at Cheyenne around 1800 UTC. The proposed RF model for
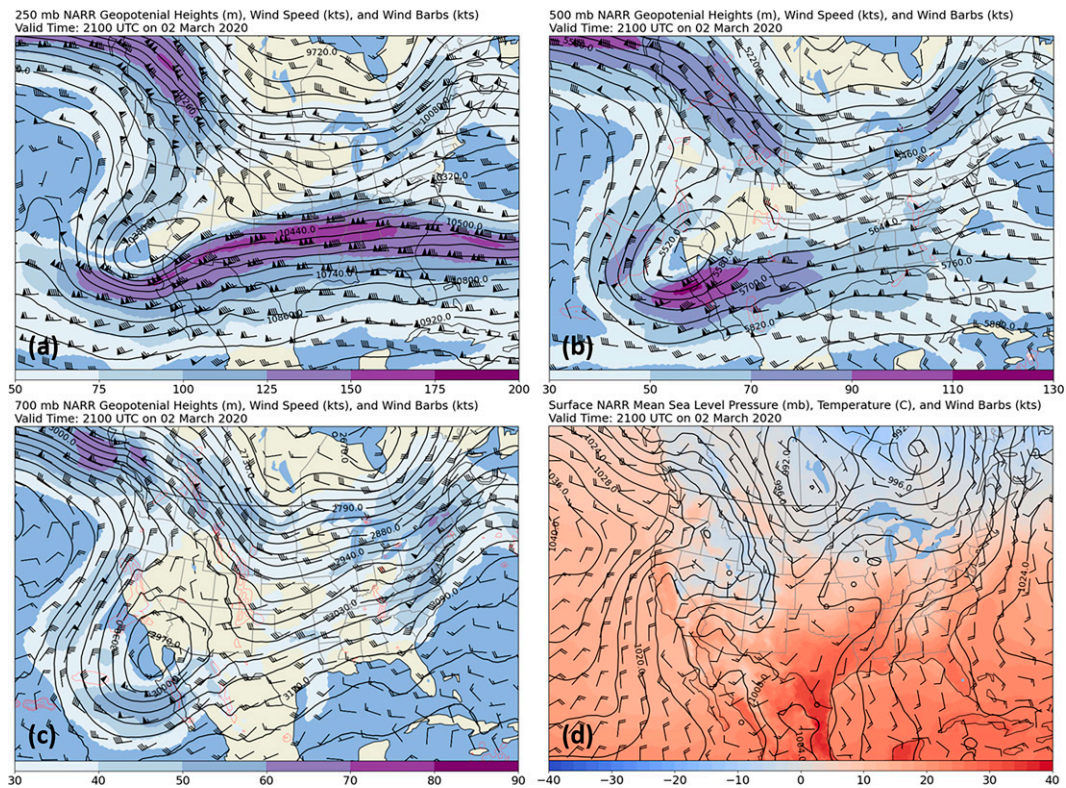
FIG. 10. As in Fig. 8, but for 2100 UTC 2 Mar 2020.

Cheyenne was a significant improvement in this particular case, successfully predicting high winds and for a longer duration than the logistic regression guidance implied.

The Bordeaux logistic regression model did not show probabilities exceeding optimal thresholds until 0600 UTC 1 January, significantly later than the initial onset of 58+ mph gusts around 1500 UTC 31 December. This extreme event featured peak wind gusts up to 81 mph between 1200 and 1500 UTC 1 January, occurring 3–6 h earlier than the logistic regression model predicted the maximum probability of 88%. The proposed RF model for Bordeaux also struggled to capture the early onset of high wind observations, with the none classification favored as 1-h sustained winds climbed above warning criteria at 0900 UTC 31 December. The RF model began predicting high winds at 1500 UTC. This was 6 h after the initial onset of warning-criteria winds, but was still 12–15 h earlier than was implied by the logistic regression guidance. The Bordeaux RF model continued to accurately predict high winds through the remainder of the event ending at 2100 UTC 1 January. The strongest winds coincided well with the peak high wind probability (95%) from the RF model.

All three RF models performed exceptionally well relative to the observations, with all high wind predictions occurring within 3–6 h of the actual onset of warning-criteria winds. The existing logistic regression and correlation guidance for each site did key in to the potential for strong winds, but was generally too late with the overall timing. This appears to be one of the greatest values of the new guidance, as timing information is critical to transportation management partners to display on dynamic message boards on area highways. One major shortcoming is the models' inability to predict actual wind speeds. A RF regressor could prove useful to gain insight into the actual intensity of a given event, especially for long duration events or when gusts exceed 75 mph. A preliminary analysis into a simple RF regressor for this case predicted peak wind gusts reasonably well (within 5 mph) and almost perfectly pinned down the timing of the 80+ mph gusts at Bordeaux. These experimental results strongly suggest that future analysis and development of RF regressor models will be necessary to complement the RF classifier models.

2) CASE 2: 2–5 MARCH 2020

During the 72-h period between 1200 UTC 2 March 2020 and 1200 UTC 5 March 2020, three shortwave troughs passed through northeast Wyoming and eastern Montana, while much of the area remained under northwest flow. Figure 10 shows a four-panel plot of NARR data from 2100 UTC 2 March 2020, with the approach of the first of three shortwaves. CAG-CPR height gradients at 850 and 700 mb increased to 80 m by 2100 UTC 2 March and remained above 60 m through 1800 UTC 3 March. These gradients spiked up above 60 m again during the day on 4 March.

The Arlington and Bordeaux RF models predicted this long-duration event extremely well. Figure 11 shows RF probabilities compared to observed wind gusts and high wind
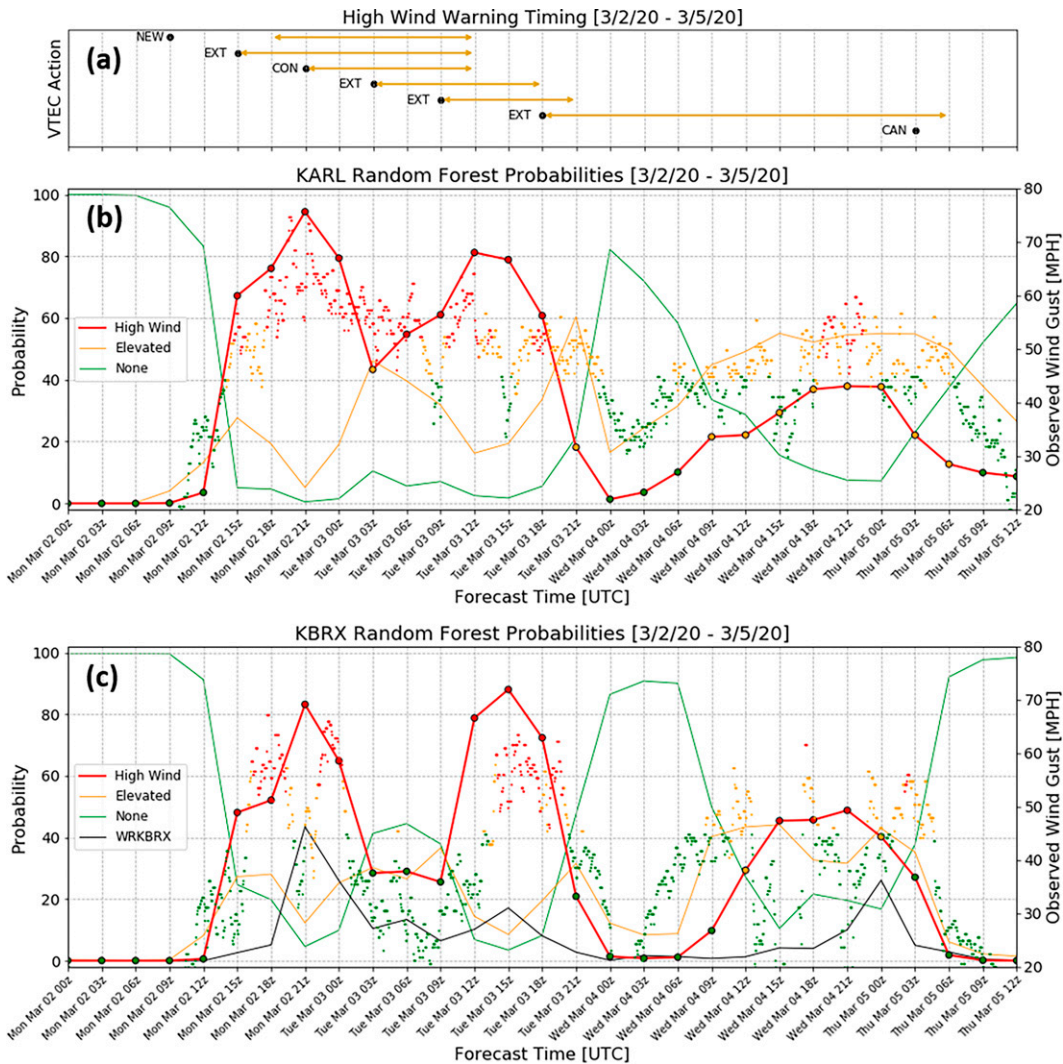
FIG. 11. Time series of (a) high wind warning timing and RF model probability for high wind (red), elevated (orange), and none (green) classifications at (b) Arlington and (c) Bordeaux between 0000 UTC 2 Mar 2020 and 1200 UTC 5 Mar 2020. Logistic regression model probabilities (black) are provided for Bordeaux (WRKBRX) in (c). Scatter points located on high wind probability line (red) are color coded to indicate the RF model predicted classification. Smaller, more numerous, scatter points represent observed wind gusts color coded to indicate observed wind classification (wind gust ≥ 58 mph and/or 1-h sustained wind ≥ 40 mph). For example, an observed 50-mph wind gust may be color coded red if observed sustained winds (not shown) at the time meet high wind criteria. Black dots and orange arrows shown in (a) indicate issuance time and duration of high wind warning issued by NWS Cheyenne, respectively, along with VTEC action including new events (NEW), event continued (CON), event extended in time (EXT), and event cancelled (CAN).

warning valid times. The arrows on Fig. 11a show the start and end time of the high wind warning. Black scatter points are annotated for each valid time event code (VTEC) action on the high wind warnings including new events (NEW), event continued (CON), event extended in time (EXT), and event cancelled (CAN). A CON would indicate that the forecaster believes the duration and areal coverage of the warning accurately represents the hazard, and no changes are necessary at the time of the forecast update. An EXT means the forecaster believes the threat may extend beyond the time

frame covered by the warning, and therefore adjusts the start or end time of the warning.

At Arlington, high winds were correctly predicted between 1500 UTC 2 March and 1800 UTC 3 March for all but one time step, as wind gusts varied between 55 and 75 mph. The initial high wind warning issued at 0900 UTC 2 March for Arlington went into effect at 1800 UTC 2 March. Shortly thereafter, this warning needed to be started earlier, as wind gusts approaching 60 mph were being observed by 1500 UTC 2 March. Additionally, the initial expiration time of 1200 UTC
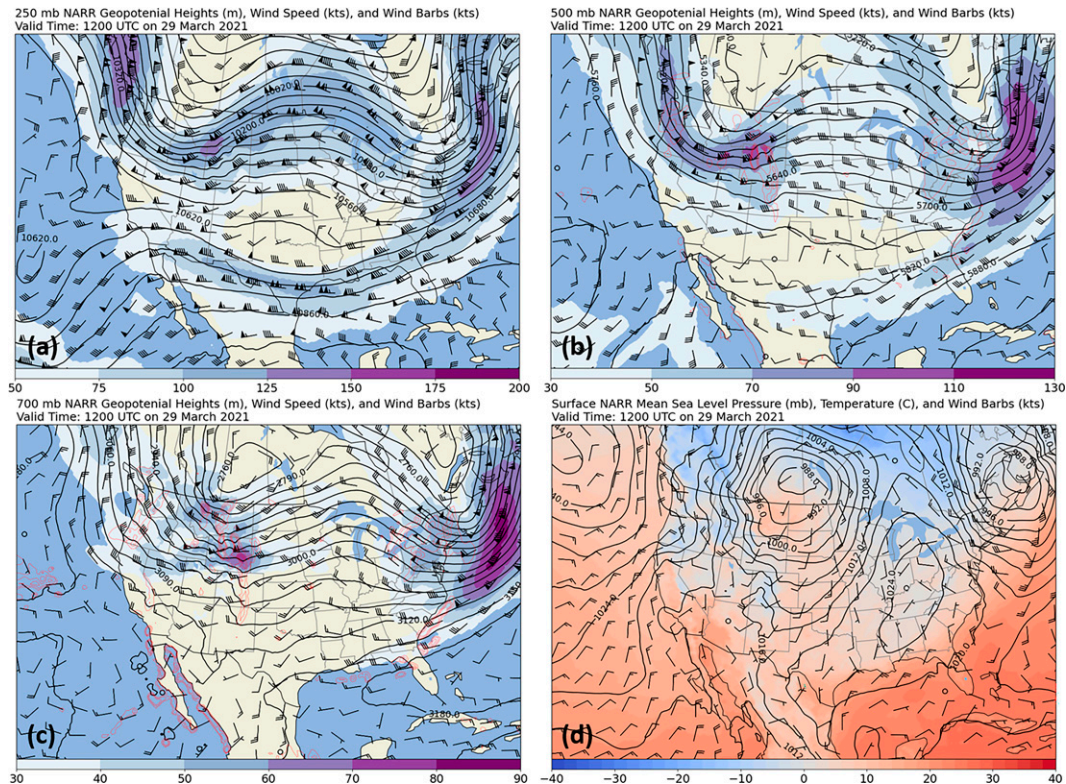
FIG. 12. As in Fig. 8, but for 1200 UTC 29 Mar 2020.

3 March needed to be extended multiple times, with high winds continuing to be observed like RF model probabilities suggested. After the second time the warning needed to be extended beyond the initial expiration, forecasters extended the warning through 0600 UTC 5 March. A 9-h decrease in winds (still gusting around 35–45 mph) was accurately depicted by the RF model before increasing again through 0600 UTC 5 March.

At Bordeaux, the RF model accurately depicted the select times where high winds occurred during the 3-day stretch. The RF model correctly predicted high winds between 1500 and 2100 UTC each day from 2 to 4 March, while also accurately predicting a lull in winds during the overnight hours. Just like for Arlington, the high wind warning needed to be started earlier compared to the timing from the initial issuance, as wind gusts approached 60 mph by 1500 UTC 2 March. Additionally, multiple extensions to the high wind warning ending time were needed by forecasters to fully capture this event. This uncertainty of timing of the high wind warning by forecasters could be attributed to the poor performance of the previously developed logistic regression model for Bordeaux (black line in Fig. 11c), which only briefly reached the optimal threshold of 26% during this event.

The Cheyenne RF model predicted only one time step as high wind, which ended up being a false alarm (not shown) as high winds were not observed at Cheyenne for this event. However, a forecaster could see this lone high wind prediction as a potential outlier or low confidence prediction as all

surrounding predictions were for the none classification along with very low high wind probabilities. The logistic regression model for Cheyenne only peaked around 2% at this time, which did not suggest high winds. Possible reasoning for this one high wind prediction by the RF model is the simultaneous peak in height and pressure gradient predictors at this time. These predictors were in the higher percentiles of their respective distributions from the training dataset, while predictors such as wind speeds and omega were much more modest.

With the RF guidance available, forecasters could have provided earlier lead time and better time duration for high wind warnings at the time of initial issuance (under the assumption that NARR data were fairly similar to GFS prediction). Multiple EXT actions suggest the timing of the high wind threat was not accurately represented by current operational tools and traditional forecast analysis. High wind probabilities for the RF models at Arlington and Bordeaux remained high beyond the initial expiration time, 1200 UTC 3 March. Despite wind gusts not constantly meeting high wind criteria after 1800 UTC 3 March for Arlington or Bordeaux, the RF model's prediction would have indicated that the potential for high winds would continue for a longer duration.

3) CASE 3: 28–29 MARCH 2021 (REAL-TIME GFS DATA)

This final case study was the first examined using GFS data as input into the RF models, similar to how forecasters would analyze the data in real-time operations. Southeast Wyoming was in the right-exit region of an approaching 300-mb jet
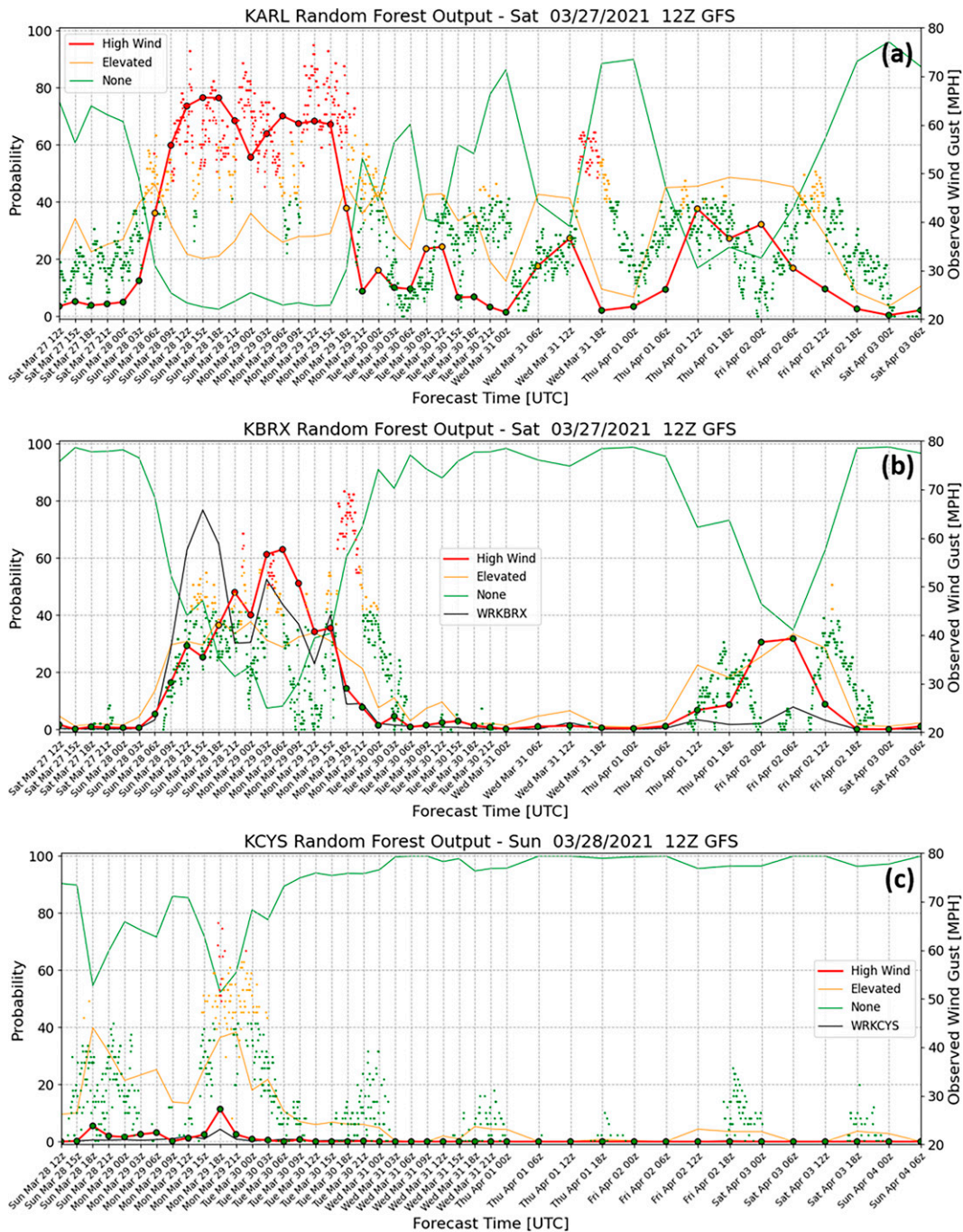
FIG. 13. As in Fig. 11, but at (a) Arlington, (b) Bordeaux, and (c) Cheyenne starting at 1200 UTC 27 Mar 2021 in (a) and (b) and 1200 UTC 28 Mar 2021 in (c) without arrows to represent high wind warning valid times. Logistic regression model probabilities (black) are provided for Bordeaux (WRKBRX) and Cheyenne (WRKCYS) in (b) and (c), respectively.

streak, with 50 kt of 700-mb flow during the morning of 28 March as height and MSLP gradients tightened. These conditions remained favorable until the passage of a surface cold front late in the afternoon of 29 March. Figure 12 shows a four-panel plot of NARR data from 1200 UTC 29 March

2021 as the system approached southeast Wyoming. Near Cheyenne, low- and midlevel lapse rates steepened significantly around 1800 UTC 29 March immediately ahead of the cold front, which provided a short window of opportunity for high winds.
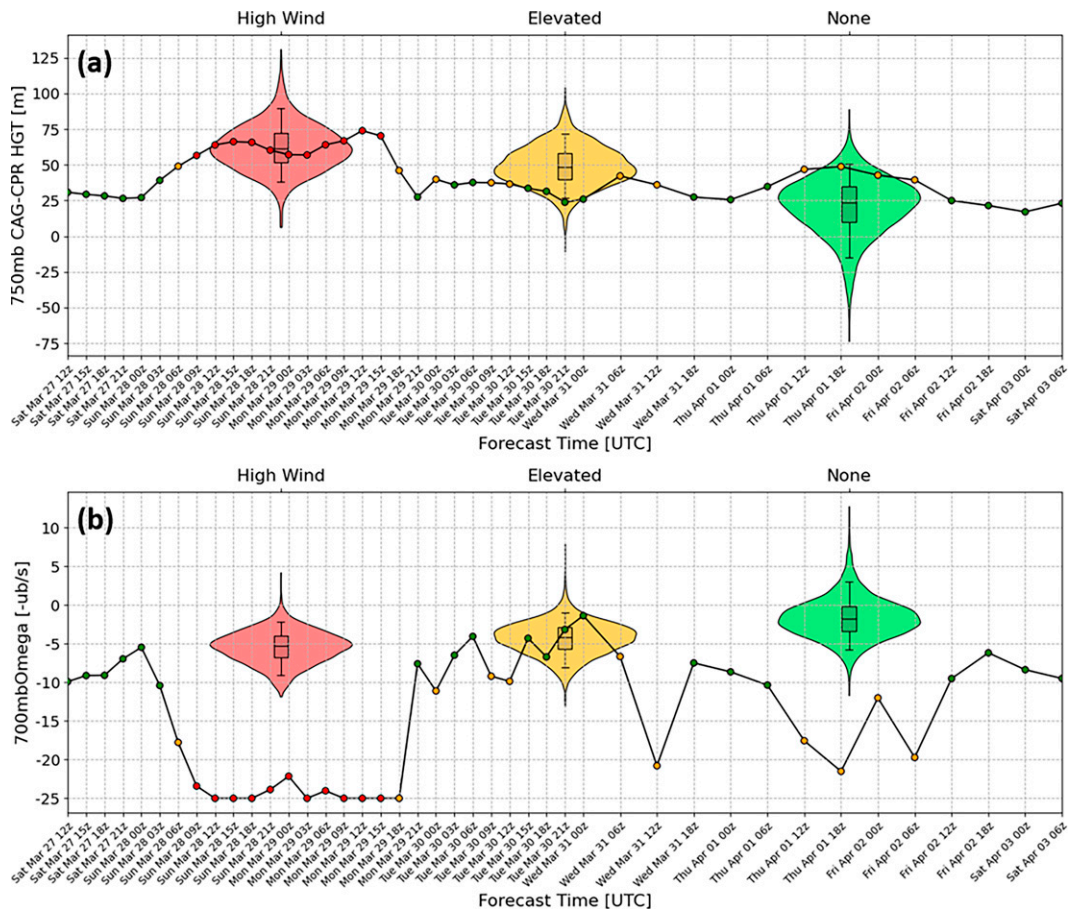
FIG. 14. Time series of GFS (a) 750-mb CAG–CPR height gradients (m) and (b) 700-mb omega ($-\mu$b s^{-1}) at Arlington between 1200 UTC 27 Mar 2021 and 0600 UTC 3 Apr 2021. Scatter points located on black lines are color coded to indicate the RF model predicted classification. Violin plots with imbedded boxplots highlight the predictor's distribution from the NARR dataset for each wind classification. Boxes show the interquartile range between the 25th and 75th percentile with whiskers representing the 5th and 95th percentiles.

Figures 13a and 13b show a time series of probabilities for each classification from the 1200 UTC GFS run on 27 March 2021. This was the latest data available at the time of high wind warning issuance (2100 UTC 27 March) for Arlington and Bordeaux, while a high wind watch was issued at this same time for Cheyenne. Overall, the models performed reasonably well for Arlington and Bordeaux as they highlighted an extended period of elevated to high winds. The predicted timing of high winds was impressive at Arlington, with the model suggesting 0900 UTC 28 March–1500 UTC 29 March. The high wind event began just before, and ended just after this period. The prediction of high winds at Arlington was fairly consistent with multiple GFS model cycles beginning as far as five days prior to the event (not shown). The Bordeaux RF model more accurately predicted the onset of high winds compared to the logistic regression model (black line in Fig. 13b), which was 9 h too early. Despite the RF model at Cheyenne predicting a none classification when high winds were observed, it was encouraging to see a peak in high wind probability at the time of high winds (1800 UTC 29 March). Optimal probabilities

could be determined for high wind prediction, as discussed for previous guidance in section 1a.

One point to remember is that these RF models are trained on the maximum wind gusts and maximum 1-h sustained winds over a 3-h period centered on these forecast times. Therefore, temporary decreases in observed wind speed without model acknowledgment can be expected at times. This can be seen in Fig. 13a for Arlington around 0600 UTC 29 March, where maximum wind gusts exceeded 60 mph within the 3-h period centered on 0600 UTC, but also wind gusts as low as 35 mph were observed in this period.

One caveat with this system is that the training dataset used the NARR, which is a reanalysis system and not an operational forecast model. For real-time implementation, GFS forecast data were used. While this might not be the best practice for machine learning due to difference between the GFS and NARR, real-time evaluation of GFS forecast data compared to the distribution of NARR data from the training dataset is provided to forecasters.

TABLE A1. List of predictors used for RF models at Arlington, Bordeaux, and Cheyenne.

	Arlington	Bordeaux	Cheyenne
Height gradients	750-mb CAG–ARL height gradient (m) 750-mb CAG–BRX height gradient (m) 750-mb CAG–CPR height gradient (m) 700-mb CAG–CPR height gradient (m) 750-mb CAG–DGW height gradient (m) 750-mb GJT–CYS height gradient (m) 750-mb SBS–DGW height gradient (m)	800-mb ARL–BRX height gradient (m) 800-mb CAG–BRX height gradient (m) 800-mb CAG–CPR height gradient (m) 800-mb CAG–DGW height gradient (m) 800-mb RWL–DGW height gradient (m) 800-mb SBS–DGW height gradient (m)	725-mb CAG–BRX height gradient (m) 700-mb CAG–CPR height gradient (m) 800-mb GJT–CYS height gradient (m) 775-mb RWL–DGW height gradient (m) 725-mb SBS–DGW height gradient (m)
Pressure gradients	CAG–ARL MSLP gradient (mb) CAG–BRX MSLP gradient (mb) CAG–CPR MSLP gradient (mb) CAG–DGW MSLP gradient (mb) GJT–CYS MSLP gradient (mb) SBS–DGW MSLP gradient (mb)	ARL–BRX MSLP gradient (mb) CAG–BRX MSLP gradient (mb) CAG–CPR MSLP gradient (mb) CAG–DGW MSLP gradient (mb) RWL–DGW MSLP gradient (mb) SBS–DGW MSLP gradient (mb)	CAG–DGW MSLP gradient (mb) GJT–CYS MSLP gradient (mb)
Wind speed and direction	700-mb wind speed (kt) 700-mb wind direction (°) 650-mb wind speed (kt) 650-mb wind direction (°) 1000–650-mb max wind speed (kt)	750-mb wind speed (kt) 750-mb wind direction (°) 700-mb wind speed (kt) 700-mb wind direction (°) 1000–700-mb max wind speed (kt)	750-mb wind speed (kt) 750-mb wind direction (°) 700-mb wind speed (kt) 700-mb wind direction (°) 1000–650-mb max wind speed (kt)
Omega	700-mb omega ($-\mu$b s^{-1}) 700–500-mb mean omega ($-\mu$b s^{-1})	750-mb omega ($-\mu$b s^{-1}) 700-mb omega ($-\mu$b s^{-1}) 850–700-mb mean omega ($-\mu$b s^{-1})	650-mb omega ($-\mu$b s^{-1}) 725-mb omega ($-\mu$b s^{-1}) 700–500-mb mean omega ($-\mu$b s^{-1})
Lapse rates and vertical profiles	700–500-mb lapse rate (°C km^{-1}) SFC–700-mb lapse rate (°C km^{-1}) 650–850-mb potential temperature difference (K)	700–500-mb lapse rate (°C km^{-1}) SFC–700-mb lapse rate (°C km^{-1}) 650–850-mb potential temperature difference (K)	700–500-mb lapse rate (°C km^{-1}) 850–700-mb lapse rate (°C km^{-1}) SFC–700-mb lapse rate (°C km^{-1}) 650–850-mb potential temperature difference (K)
General	500-mb temperature (°C) 700-mb temperature (°C) MSLP (mb) Month Hour	700-mb temperature (°C) MSLP (mb) Month Hour	700-mb temperature (°C) MSLP (mb) Month Hour

Figure 14a shows a time series of GFS forecast data for 750-mb CAG–CPR height gradient, with violin plots representing the distribution of this predictor in the NARR dataset for each of the three classifications. This provides the forecaster with an additional tool to assess the value for each predictor. Forecast data for all lead times in Fig. 14a remain within the distribution of NARR data, which was observed for almost every predictor in this case. The one exception was the 700-mb omega values at Arlington, which can be seen in Fig. 14b. GFS forecast data during the high wind event reaches $-25 -\mu$b s^{-1}, while the distribution of 700-mb omega for any classification does not appear to drop below $-13 -\mu$b s^{-1}. Arlington is situated at the base of the northern slopes of the Snowy Range, where the elevation gradient might be better resolved in the 13-km GFS grid spacing compared to the 32-km NARR grid

spacing. Therefore, terrain effects on the omega fields might not be captured well by this reanalysis system. Regardless, no evidence has been found that this large difference between NARR and GFS 700-mb omega values at Arlington influences the RF model's predictions. Additional analysis on differences between NARR and GFS data for these sites could be conducted in the future and potentially factored back into the training process for the RF models.

4. Conclusions

Random forest (RF) models were developed to predict high winds (wind gusts ≥ 58 mph and/or 1-h sustained winds ≥ 40 mph) for three locations across southeast Wyoming. In addition to Cheyenne, locations selected included two areas

where strong crosswinds frequently result in truck blow overs along Interstate 25 (Bordeaux) and Interstate 80 (Arlington) during the winter months. Model variables from the North American Regional Reanalysis (NARR) were selected based on correlations to high winds and critical success index (CSI) scores using univariate thresholds.

The RF models were evaluated against current operational tools using two methods: 1) single-observation verification and 2) event-based verification. Single-observation verification only compares the prediction and observation at a single time, while event-based verification groups together multiple times based on occurrence and duration of high wind events. The RF models trained for Arlington and Bordeaux performed the best, with CSI scores for event-based verification greater than 0.6 over the testing dataset, which covers two winters. The RF model for Cheyenne had a CSI score of 0.313 for event-based verification, as both the POD and FAR hovered around 0.5. These three RF models all showed improved skill at predicting high wind events over current operational tools. While the Cheyenne RF model did not perform as well, there are a few possible explanations, including 1) less frequent occurrence of high winds and 2) predictors not being as strongly correlated with high winds. Regardless, the output still presents high wind probabilities for operational meteorologists to leverage in the forecast process and was a significant improvement over current operational tools. The only exception was the Cheyenne RF model for single-observation verification. Future improvements will be explored, including identifying more skillful predictors and different methods for RF model configuration.

Three case studies presented further explored the usefulness of newly developed RF models to predict occurrences of high winds and displayed their improved forecast skill over current operational tools. The final case analyzed the output when GFS forecast data were used as inputs into the RF models. While these RF classification models predict defined ranges of wind speed and gust including the local high wind criteria, additional information about wind gust magnitude could be useful to forecasters and core partners. Future work will explore RF regressor models to aid forecasters in not only predicting the occurrence of high winds, but also predicting the strength and most likely timing of the strongest winds.

These results have provided motivation for expansion of RF models to additional sites across southeast Wyoming where high winds and truck blow overs are frequent during the winter months. The authors hope this work could also be a useful blueprint for improving high wind prediction at additional locations in complex mountainous terrain around the world.

APPENDIX

List of RF Model Predictors

Table A1 presents a full list of predictors used for RF models at Arlington, Bordeaux, and Cheyenne.

REFERENCES

Alpers, W., A. Ivanov, and J. Horstmann, 2009: Observations of bora events over the Adriatic Sea and Black Sea by spaceborne synthetic aperture radar. *Mon. Wea. Rev.*, **137**, 1150–1161, https://doi.org/10.1175/2008MWR2563.1.

Breiman, L., 2001: Random forests. *Mach. Learn.*, **45**, 5–32, https://doi.org/10.1023/A:1010933404324.

Brinkmann, W. A. R., 1974: Strong downslope winds at Boulder, Colorado. *Mon. Wea. Rev.*, **102**, 592–602, https://doi.org/10.1175/1520-0493(1974)102<0592:SDWABC>2.0.CO;2.

Bröcker, J., and L. A. Smith, 2007: Increasing the reliability of reliability diagrams. *Wea. Forecasting*, **22**, 651–661, https://doi.org/10.1175/WAF993.1.

Carr, N., B. Haines, and V. Meola, 2020: Using a random forest model to forecast the development of dense fog at TAF sites. Poster, *45th National Weather Association Annual Meeting*, online, National Weather Association, Abstract NWAS20, https://nwas.org/annual-meeting-events/past-nwa-annual-meetings/.

Colle, B. A., and C. F. Mass, 2000: High-resolution observations and numerical simulations of easterly gap flow through the Strait of Juan de Fuca on 9–10 December 1995. *Mon. Wea. Rev.*, **128**, 2398–2422, https://doi.org/10.1175/1520-0493(2000)128<2398:HROANS>2.0.CO;2.

Collins, E. M., Z. J. Lebo, B. Geerts, R. Capella, and R. Cox, 2020: Verification of wind forecasts from the High-Resolution Rapid Refresh. *10th Conf. on Transition of Research to Operations*, Boston, MA, Amer. Meteor. Soc., 8A.5, https://ams.confex.com/ams/2020Annual/meetingapp.cgi/Paper/370705.

Cotton, W. R., J. F. Weaver, and B. A. Beitler, 1995: An unusual summertime downslope wind event in Fort Collins, Colorado on 3 July 1993. *Wea. Forecasting*, **10**, 786–797, https://doi.org/10.1175/1520-0434(1995)010<0786:AUSDWE>2.0.CO;2.

Dang, T., 2020: Using machine learning to help communicate winter weather impacts. Oral presentation, *45th National Weather Association Annual Meeting*, online, National Weather Association, https://nwas.org/annual-meeting-events/past-nwa-annual-meetings/.

Defant, F., 1951: Local winds. *Compendium of Meteorology*, T. F. Malone, Ed., Amer. Meteor. Soc., 655–672.

Drechsel, S., and G. J. Mayr, 2008: Objective forecasting of foehn wind for a subgrid-scale Alpine valley. *Wea. Forecasting*, **23**, 205–218, https://doi.org/10.1175/2007WAF2006021.1.

Finch, Z., and D. T. Lindsey, 2016: A high-wind prediction model for Bordeaux, Wyoming. *23rd Conf. on Probability and Statistics in the Atmospheric Sciences*, New Orleans, LA, Amer. Meteor. Soc., 226, https://ams.confex.com/ams/96Annual/webprogram/Paper281470.html.

Fudeyasu, H., T. Kuwagata, Y. Ohashi, S. Suzuki, Y. Kiyohara, and Y. Hozumi, 2008: Numerical study of the local downslope wind "Hirodo-Kaze" in Japan. *Mon. Wea. Rev.*, **136**, 27–40, https://doi.org/10.1175/2007MWR2049.1.

Hammer, C. L., 2015: Downslope windstorms: Improving forecast skill for Cheyenne, Wyoming. Poster, *40th National Weather Association Annual Meeting*, Oklahoma City, OK, National Weather Association, CP-16, https://nwas.org/annual-meeting-events/past-nwa-annual-meetings/.

Herman, G. R., and R. S. Schumacher, 2018a: Dendrology in numerical weather prediction: What random forests and logistic regression tell us about forecasting extreme precipitation. *Mon. Wea. Rev.*, **146**, 1785–1812, https://doi.org/10.1175/MWR-D-17-0307.1.

——, and ——, 2018b: Money doesn't grow on trees, but forecasts do: Forecasting extreme precipitation with random forests. *Mon. Wea. Rev.*, **146**, 1571–1600, https://doi.org/10.1175/MWR-D-17-0250.1.

Hill, A. J., G. R. Herman, and R. S. Schumacher, 2020: Forecasting severe weather with random forests. *Mon. Wea. Rev.*, **148**, 2135–2161, https://doi.org/10.1175/MWR-D-19-0344.1.

Horel, J., and Coauthors, 2002: MesoWest: Cooperative mesonets in the western United States. *Bull. Amer. Meteor. Soc.*, **83**, 211–226, https://doi.org/10.1175/1520-0477(2002)083<0211:MCMITW>2.3.CO;2.

Klink, K., 1999: Climatological mean and interannual variance of United States surface wind speed, direction and velocity. *Int. J. Climatol.*, **19**, 471–488, https://doi.org/10.1002/(SICI)1097-0088(199904)19:5<471::AID-JOC367>3.0.CO;2-X.

Lindsey, D. T., and Coauthors, 2011: A high wind statistical prediction model for the northern front range of Colorado. *Electron. J. Oper. Meteor.*, **12**, 2011-EJ03, https://nwafiles.nwas.org/ej/pdf/2011-EJ3.pdf.

Martner, B. E., and J. D. Marwitz, 1982: Wind characteristics in southern Wyoming. *J. Appl. Meteor.*, **21**, 1815–1827, https://doi.org/10.1175/1520-0450(1982)021<1815:WCISW>2.0.CO;2.

McGovern, A., K. L. Elmore, D. J. Gagne, S. E. Haupt, C. D. Karstens, R. Lagerquist, T. Smith, and J. K. Williams, 2017: Using artificial intelligence to improve real-time decision making for high-impact weather. *Bull. Amer. Meteor. Soc.*, **98**, 2073–2090, https://doi.org/10.1175/BAMS-D-16-0123.1.

——, R. Lagerquist, D. J. Gagne II, G. E. Jergensen, K. L. Elmore, C. R. Homeyer, and T. Smith, 2019: Making the black box more transparent: Understanding the physical implication of machine learning. *Bull. Amer. Meteor. Soc.*, **100**, 2175–2199, https://doi.org/10.1175/BAMS-D-18-0195.1.

Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360, https://doi.org/10.1175/BAMS-87-3-343.

Overland, J. E., and B. A. Walter, 1981: Gap winds in the Strait of Juan de Fuca. *Mon. Wea. Rev.*, **109**, 2221–2233, https://doi.org/10.1175/1520-0493(1981)109<2221:GWITSO>2.0.CO;2.

Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, **12**, 2825–2830, https://doi.org/10.48550/arXiv.1201.0490.

Reed, R. J., 1981: A case study of a bora-like windstorm in western Washington. *Mon. Wea. Rev.*, **109**, 2383–2393, https://doi.org/10.1175/1520-0493(1981)109<2383:ACSOAB>2.0.CO;2.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608, https://doi.org/10.1175/2008WAF2222159.1.

Sandmael, T., K. L. Elmore, and B. R. Smith, 2020: A new machine learning–based tornado detection algorithm for the WSR-99D network. *22nd Conf. on Artificial Intelligence for Environmental Science*, Boston, MA, Amer. Meteor. Soc., 364, https://ams.confex.com/ams/2020Annual/meetingapp.cgi/Paper/363586.

Schulz, B., and S. Lerch, 2022: Machine learning methods for postprocessing ensemble forecasts of wind gusts: A systematic comparison. *Mon. Wea. Rev.*, **150**, 235–257, https://doi.org/10.1175/MWR-D-21-0150.1.

Schumacher, R. S., A. J. Hill, M. Klein, J. A. Nelson, M. J. Erickson, S. M. Trojniak, and G. R. Herman, 2021: From random forests to flood forecasts: A research to operations success story. *Bull. Amer. Meteor. Soc.*, **102**, E1742–E1755, https://doi.org/10.1175/BAMS-D-20-0186.1.

Smith, R. B., 1987: Aerial observations of the Yugoslavian bora. *J. Atmos. Sci.*, **44**, 269–297, https://doi.org/10.1175/1520-0469(1987)044<0269:AOOTYB>2.0.CO;2.

Sprenger, M., S. Schemm, R. Oechslin, and J. Jenkner, 2017: Nowcasting foehn wind events using the AdaBoost machine learning algorithm. *Wea. Forecasting*, **32**, 1079–1099, https://doi.org/10.1175/WAF-D-16-0208.1.

Vosper, S. B., 2004: Inversion effects on mountain lee waves. *Quart. J. Roy. Meteor. Soc.*, **130**, 1723–1748, https://doi.org/10.1256/qj.03.63.

Wilson, T., and J. Cote, 2020: Using a convolutional neural network to assist in situational awareness. Oral presentation, *45th National Weather Association Annual Meeting*, online, National Weather Association, https://nwas.org/annual-meeting-events/past-nwa-annual-meetings/.