USING ARTIFICIAL INTELLIGENCE TO IMPROVE REAL-TIME DECISION-MAKING FOR HIGH-IMPACT WEATHER

Amy McGovern, Kimberly L. Elmore, David John Gagne II, Sue Ellen Haupt, Christopher D. Karstens, Ryan Lagerquist, Travis Smith, and John K. Williams

Modern artificial intelligence (AI) techniques can aid forecasters on a wide variety of high-impact weather phenomena.

Weather significantly impacts society for better and for worse. For example, severe weather hazards caused over $7.9 billion of property damage in 2015 (National Oceanic and Atmospheric Administration/National Centers for Environmental Information 2016; CoreLogic 2016). The National Academies of Sciences, Engineering, and Medicine (2016) cites improving forecasting of such events as a critical priority, and the European Centre for Medium-Range Weather Forecasts (ECMWF) recently announced goals for 2025 (ECMWF 2016) that stress the importance of improving these forecasts. On the positive side, improvements in forecasting solar power, which increasingly impacts the electrical grid, are expected to save utility companies $455 million by 2040 (Haupt et al. 2016). Additional savings can be found through improved forecasting in other areas of computational sustainability. Computational sustainability is a new and growing interdisciplinary research area focusing on computational solutions for questions of Earth sustainability.

In recent years, operational numerical weather prediction (NWP) models have significantly increased in resolution (e.g., Weygandt et al. 2009). At the same time, the number and quality of observational systems has grown, and new systems, such as Geostationary Operational Environmental Satellite R series (GOES-R), will generate high-quality data at fine spatial and temporal resolutions. These data contain valuable information, but their variety and volume can be overwhelming to forecasters, and this can hinder decision-making if not handled properly (Karstens et al. 2015, 2016). This data deluge is commonly termed "big data." Artificial intelligence (AI) and related data science methods have been developed to work with big data across a variety of disciplines.

Applying AI techniques in conjunction with a physical understanding of the environment can substantially improve prediction skill for multiple types of high-impact weather. This approach expands on traditional model output statistics (MOS) techniques (Glahn and Lowry 1972), which derive probabilistic, categorical, and deterministic forecasts from NWP model output. Because of their simplicity and longevity, forecasters have gained trust in MOS techniques. AI techniques provide a number of advantages, including easily generalizing spatially and temporally, handling large numbers of predictor variables, integrating physical understanding into the

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 09/24/21 06:09 PM UTC

models, and discovering additional knowledge from the data. In recent years, forecasters and researchers have begun to adopt AI techniques much more widely, as they demonstrate their power in a wide variety of applications, including postmodel bias correction, handling large datasets, reducing cognitive overload, and discovering new knowledge in large datasets. With the growth in applications for data science techniques outside of atmospheric science as well, AI techniques promise to continue to enhance prediction and understanding of many weather-related phenomena. The primary goals of this paper are to introduce modern AI techniques to a broad audience and to demonstrate their utility in predicting a wide variety of high-impact weather phenomena.

The rest of this paper is organized as follows: We first review related work and provide a brief overview of some AI techniques highlighted in this paper, followed by demonstrations of how we have applied AI techniques to multiple high-impact weather applications. We discuss the benefits of AI and automation to both researchers and forecasters and conclude by discussing how AI techniques can be further used to help meteorologists and decision-makers.

RELATED WORK. Statistical models for postprocessing NWP model output have evolved within two general frameworks. "Perfect prog" models fit relationships between observed or analyzed variables and observations of a weather feature, such as temperature or precipitation (Klein et al. 1959). The models are then applied to NWP forecasts, thus implicitly assuming that the NWP model is perfect. In contrast, MOS fits a statistical model between NWP output at a given time horizon and subsequent observations at that time (Glahn and Lowry 1972), often using linear regression. Because MOS fits use the NWP output directly, they can correct for systematic biases in a model. When NWP model configurations are updated, MOS must be retrained after a sufficient number of new model forecasts are collected. Perfect-prog models are generally less accurate than a well-tuned MOS model, but they are less sensitive to model configuration changes and tend to be more robust over time. AI techniques can be used in both frameworks.

Haupt et al. (2008) provide an overview of AI techniques applied to the environmental sciences, including artificial neural networks (ANNs), decision trees, genetic algorithms (Allen et al. 2007), fuzzy logic, and principal component analysis (Elmore and Richman 2001). Baldwin et al. (2005) used hierarchical clustering to classify precipitation areas, Gagne et al. (2009) used *k*-means clustering to segment a radar image, Lakshmanan et al. (2010, 2014) used *k*-means clustering to segment a map of radar-echo classifications, and Miller et al. (2013) used clustering to identify storm tracks.

ANNs are interconnected networks of weighted nonlinear functions. When connected and trained in multiple layers, ANNs can represent any nonlinear function. They also provide the foundation for deep learning methods. ANNs have been used in a wide variety of meteorology applications since the late 1980s (Key et al. 1989), including cloud classification (Bankert 1994), tornado prediction and detection (Marzban and Stumpf 1996; Lakshmanan et al. 2005), damaging winds (Marzban and Stumpf 1998), hail size (Marzban and Witt 2001; Manzato 2013), precipitation classification (Anagnostou 2004; Lakshmanan et al. 2014), tracking storms (Lakshmanan et al. 2000), and radar quality control (Lakshmanan et al. 2007; Newman et al. 2013).

Support vector machines (SVMs) have also been used to detect and predict tornadoes (Trafalis et al. 2003; Adrianto et al. 2009). SVMs learn a linear model in a nonlinear space by transforming the data to the nonlinear space using kernels. Both ANNs and SVMs are flexible and powerful but produce models that are often difficult to interpret in terms of underlying physical concepts that the model has identified. For the ANNs, it is difficult to interpret the weights through the nonlinear functions. For the SVMs, the data transformation makes it difficult to identify the

AFFILIATIONS: McGovern—School of Computer Science, University of Oklahoma, Norman, Oklahoma; Elmore, and Smith—Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, and National Severe Storms Laboratory, Norman, Oklahoma; Gagne and Haupt—National Center for Atmospheric Research, Boulder, Colorado; Karstens*—Cooperative Institute for Mesoscale Meteorological Studies, University of Oklahoma, Norman, Oklahoma; Lagerquist—School of Meteorology, University of Oklahoma, Norman, Oklahoma; Williams*—National Center for Atmospheric Research, Boulder, Colorado
*** CURRENT AFFILIATIONS:** Karstens—NOAA/National Weather Service/Storm Prediction Center, Norman, Oklahoma; Williams—The Weather Company, An IBM Business, Andover, Massachusetts
CORRESPONDING AUTHOR: Amy McGovern, amcgovern@ou.edu

The abstract for this article can be found in this issue, following the table of contents.
DOI:10.1175/BAMS-D-16-0123.1

In final form 8 February 2017
©2017 American Meteorological Society
For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy.

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 09/24/21 06:09 PM UTC

most important features of the data or what the model has identified.

One of the simplest and most well-known statistical learning methods, linear regression, has been used in weather prediction since at least the early 1950s (Malone 1955). Kitzmiller et al. (1995) used regression to forecast the probability of severe weather, Billet et al. (1997) used it to forecast maximum hail size and large-hail probability, and Mecikalski et al. (2015) used logistic regression to forecast the probability of convective initiation, to name just a few recent examples. In linear regression, a set of weights is chosen to combine input features x_i so as to best predict an output variable $\hat{y}$, for example, to minimize the summed squared prediction error. The weights can be trained using matrix inversion or other optimization schemes, ranging from basic gradient descent through genetic algorithms if the matrix is poorly conditioned. Although linear regression learns quickly even on large datasets, it works best with problems that require a linear model and have a limited feature set. If features are redundant or not predictive, they can make learning more challenging. Ridge regression (Hoerl and Kennard 1988) penalizes the sum of squared weights in order to simplify models and improve their generalization. The lasso method penalizes the sum of the weights' absolute values, which tends to remove irrelevant variables (Tibshirani 1996). Elastic nets (Zou and Hastie 2005) combine both penalties.

Decision-tree-based methods are popular in data science for handling big data. They are able to identify and learn with only the most relevant variables, enabling users to provide many possible predictive features without worrying whether extraneous variables will overwhelm the training process. Decision trees are also human readable, which can provide insight into what relationships the model has identified related to the event being forecasted. Decision-tree-based methods have proven quite powerful in a wide variety of weather applications (Williams et al. 2008a,b; Gagne et al. 2009; McGovern et al. 2014; Williams 2014; McGovern et al. 2015; Clark et al. 2015; Elmore and Grams 2016).

Although the first objective decision-tree learning method was not developed until the mid-1980s (Quinlan 1986, 1993), subjective (human derived) decision trees have been used in meteorology since at least the mid-1960s (Chisholm et al. 1968). A decision tree splits data recursively by identifying the most relevant question at each level of the data. The tree shown in Fig. 1 was automatically developed to predict whether hail will occur. At the root node, the data are split with the question "Is the mean radar reflectivity ≤ 43.4 dBZ?" The data are further refined down each of the yes and no branches until a prediction is made at a leaf node, which may contain a class label (e.g., hail: yes), probability p [e.g., p(hail) = 0.8; Provost and Domingos 2000], scalar prediction [hail size = 3.1 in. (~7.9 cm)], or a linear predictive function.

A powerful related method is random forests (RFs; Breiman 2001). An RF is an ensemble of decision trees, each of which is trained on a separate set of bootstrapped resampled training data and selects from a random subset of questions at each node. Since they are trained on different data and using different predictors, the individual trees in the forest are diverse, providing an "ensemble of experts" that performs better than any individual tree.

Gradient boosted regression trees (GBRT; Friedman 2002) construct an ensemble of decision trees trained using boosting (Schapire 2003). Whereas each tree in an RF is equally weighted and trained on equally weighted examples, a GBRT trains on differently weighted subsets of data, where the weights are determined by the error residuals of the previous training step.

We will demonstrate the use of both RFs and GBRTs in several of the high-impact weather domains described below. While both methods are similar in performance in some cases, because of the equal

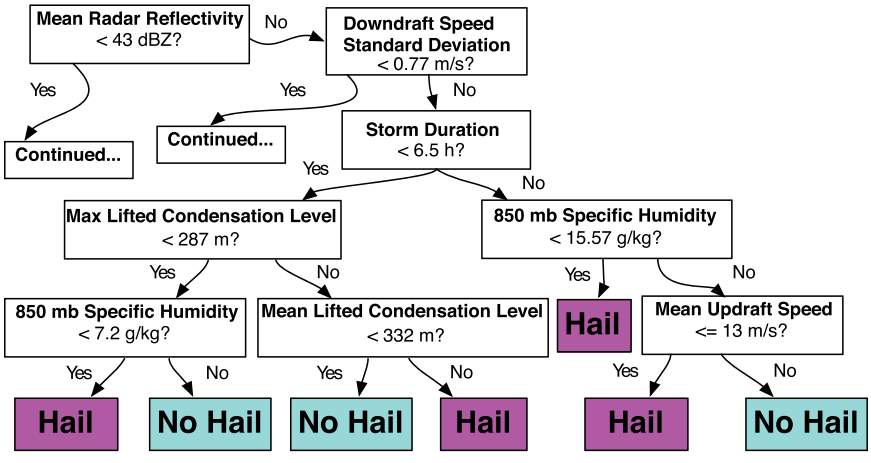


Fig. 1. An example of a decision tree for predicting if hail will occur. A version of this decision tree first appeared in Gagne (2016).

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 09/24/21 06:09 PM UTC

weighting of the individual trees in the forest, an RF will tend to regress to the mean predictions and thus not produce as sharp of a forecast. GBRTs can address this issue, but sometimes, postmodel correction is also needed. We typically use isotonic regression for postmodel correction (Niculescu-Mizil and Caruana 2005).

Both RFs and GBRTs provide the ability to measure the importance of each attribute in the dataset, which is called variable importance. After the trees are trained, each variable's data are permuted, and performance is measured with both the permuted and original data. The most important variables are those that cause the largest drop in performance. These importance estimates can be used to gain insight into the choices made by the forests, enabling physical interpretation of the models.

AI FOR HIGH-IMPACT WEATHER. This section presents some of our recent work in applying AI to a variety of high-impact weather applications. The diversity in applications is intentional, to demonstrate to the reader that AI can be used for multiple problems.

Storm duration. Predicting a storm's lifetime is important for forecasters as it helps to guide the creation of watches and warnings. This task requires knowledge of the current status of the storm as well as knowledge of the nearby environment. The training data for this task come from a preoperational product called ProbSevere (Cintineo et al. 2014). ProbSevere identifies and tracks storms in real time using composite reflectivity (maximum column reflectivity derived from multiple radars simultaneously) from the operational Multi-Radar Multi-Sensor (MRMS; Smith et al. 2016) system over the continental United States. ProbSevere also provides a small number of attributes that summarize information about the environment near the storm along with information on the current speed of the storm. The training labels are provided by running a post hoc storm-tracking program called best track (Lakshmanan et al. 2015). These labels were obtained by using data from the Multi-Year Reanalysis of Remotely Sensed Storms (MYRORSS; Ortega et al. 2012) project. The training and testing data were drawn from 9 April 2015 through 31 January 2016. Data are available for each storm cell on an approximate 2-min basis. To ensure there was no cross contamination between the training and the testing set, training was on all data except July, with testing on July and the day closest to the testing data dropped from training. For bias correction, we withheld an extra month of data (August). The training data were also subsampled for all storms lasting less than 7,200 s. Only 10% of this data were used for training. All storms lasting longer than 7,200 s in the training set were retained. This still yielded 2,872,680 samples for training. Testing data were evaluated independently on each day in July, enabling us to bootstrap the results for statistical analysis. Testing data were not subsampled.

We tested three machine-learning methods: GBRT, RF, and elastic nets. We also examined post-training bias corrected using isotonic regression. We examined multiple settings of the standard parameters to the RF and GBRT using a validation set (results not shown because of space). The best choices for the RF and GBRT were 100 trees and a maximum depth of 5. For the GBRTs, the Huber loss function was significantly better than the other loss functions. For elastic nets, we used an alpha of 0.05 and the L1 ratio of 0.9.

Figure 2 displays the predicted distributions versus the observed distributions. GBRT stands out as the best-performing method across the range

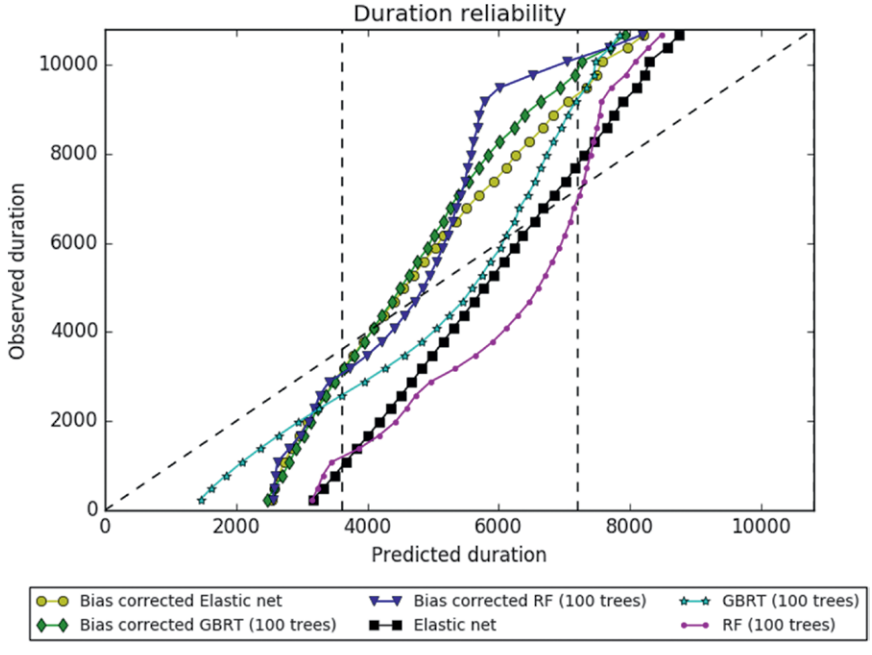


Fig. 2. Reliability diagram for predicting a storm's lifetime.

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 09/24/21 06:09 PM UTC

of predictions. While bias correction is able to improve the performance at the end points, it is not an overall improvement on the models and was left out of the real-time testing.

The best duration prediction method, GBRT, was implemented into a real-time system running in National Oceanic and Atmospheric Administration (NOAA)'s Hazardous Weather Testbed (HWT) called Probabilistic Hazard Information (PHI; Karstens et al. 2015) that uses ProbSevere to generate automated probabilistic forecasts for thunderstorm hazards. The forecasts were tested and evaluated with nine National Weather Service (NWS) forecasters in a 3-week human–machine-mix experiment during May and June of 2016, and the acceptance of the duration predictions was evaluated. As shown in Fig. 3, forecasters on average used the predicted ProbSevere duration in approximately 75% of all forecasts, while individual acceptance of these predictions varied from as low as approximately 25% to as much as 100%. These results imply that most forecasters trust these predictions or that the predictions are within an acceptable range at the time of warning decision. However, evidence (not shown) suggests that forecasters have a strong tendency to accept the default duration value so long as it is "good enough," and the default duration value during the experiment was assigned from our duration predictions. Therefore, forecasters may not be giving much thought to this predictive aspect of the forecast. Interestingly, research in optimizing decision-making suggests that "choice architects" should account for inaction bias by assigning the most likely best option to the available default (Milkman et al. 2008).

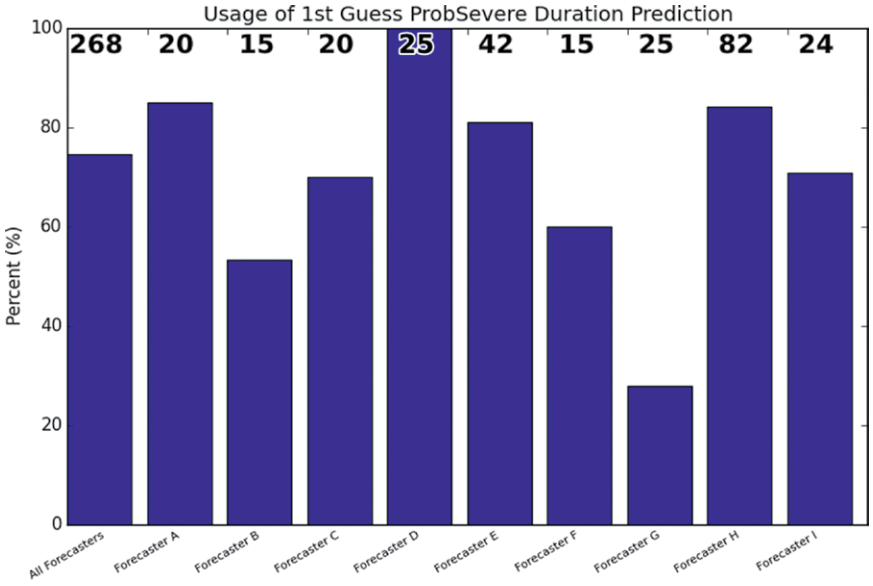


Fig. 3. Observed usage of the storm duration predictions by forecasters in the spring 2016 HWT experiment. Numbers on the chart are the counts in each bin.

Severe wind. Real-time prediction of severe wind, defined by the NWS as a gust ≥50 knots (kt; 25.7 m s^{-1}), is another important task for forecasters. This project uses AI techniques to predict the probability of severe wind within various buffer distances (0, 5, and 10 km around the storm cell) and time windows (0–15, 15–30, 30–45, 45–60, and 60–90 min into the future). We use two datasets to create predictors: quality-controlled radar images from MYRORSS and near-storm environment soundings from the Rapid Update Cycle (RUC) model (Benjamin et al. 2004). MYRORSS has a resolution of 1 km and 5 min, while the RUC has a resolution of 13 km (20 km for earlier times) and 1 h. To determine when and where severe winds occurred (verification data), we use surface observations from four datasets: the Meteorological Assimilation Data Ingest System (MADIS; McNitt et al. 2008), Oklahoma Mesoscale Network (Mesonet; McPherson et al. 2007), 1-min meteorological aerodrome reports (METARs; National Climatic Data Center 2006), and NWS local storm reports (Storm Prediction Center 2015).

Before training the models, four types of data processing are applied. First, storm cells are identified and tracked through time using both real-time (Lakshmanan and Smith 2010) and postevent (Lakshmanan et al. 2015) methods. Real-time tracking outlines the edge of each storm cell, and postevent tracking corrects deficiencies in real-time tracking, mainly false truncations. Data are processed for 804 days in the continental United States (all days from 2004 to 2011 with ≥30 NWS wind reports and available MYRORSS data). This results in nearly 20 million storm objects, where a "storm object" is one storm cell at one time step. Second, wind observations are causally linked to storm cells. For each wind observation *W*, storm objects are interpolated along their respective tracks to the same time as *W*. If the edge of the nearest storm object *S* is within the given

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 09/24/21 06:09 PM UTC

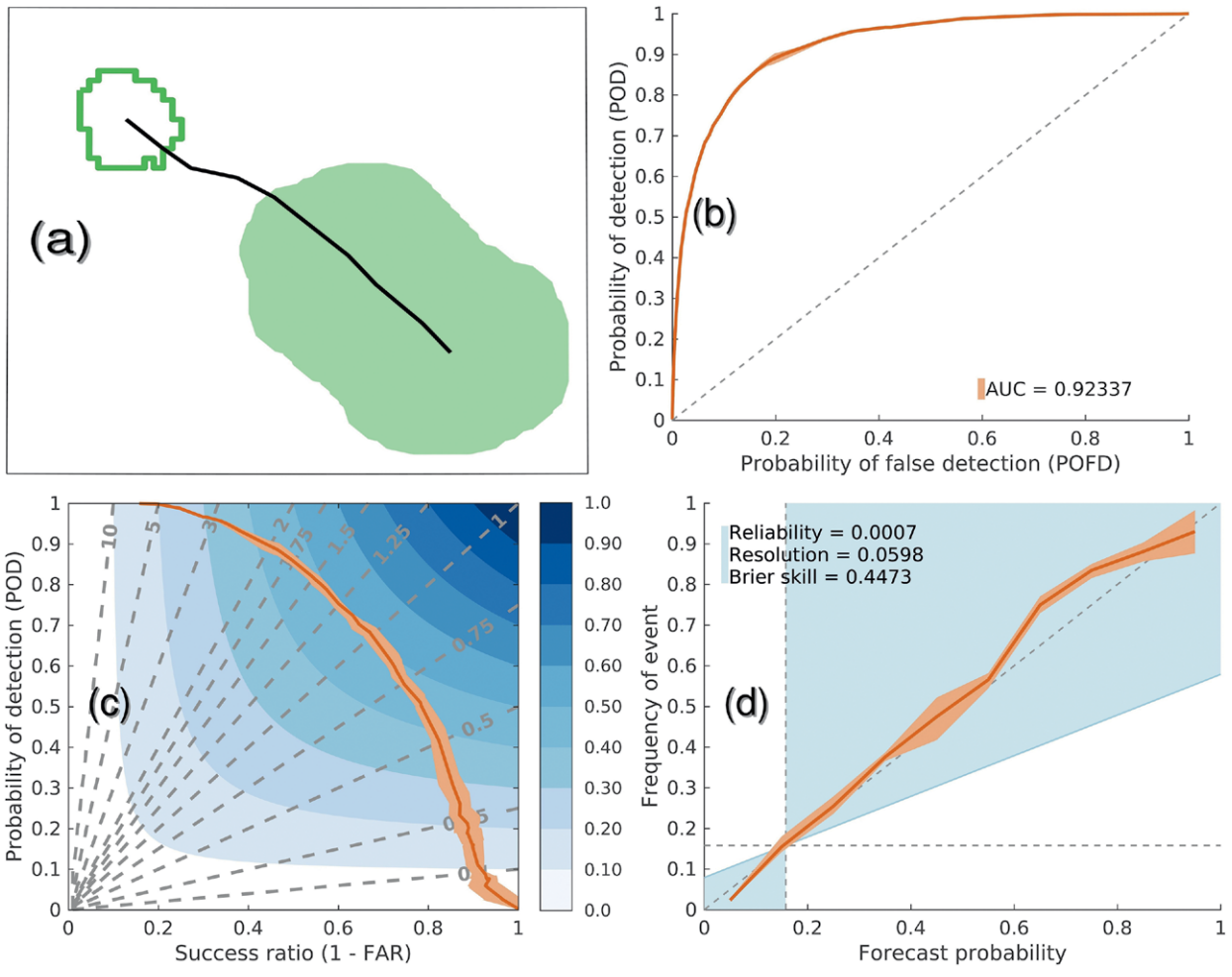


Fig. 4. (a) Labeling of storm object *S* (dark green polygon). Label is based on wind gusts in light green area, which is a 5-km buffer around storm objects occurring 30–45 min later in the same track. (b) ROC curve (Metz 1978). The gray line is the ROC curve for a random predictor, and AUC is the area under the curve. (c) Performance diagram (Roebber 2009). The gray lines are frequency bias, and the color fill is CSI. (d) Attributes diagram (Hsu and Murphy 1986). The orange line is the reliability curve, the diagonal gray line is a perfect reliability curve, the vertical gray line is climatology, and the horizontal gray line is the no-resolution line (reliability curve for a model that always predicts climatology). In all cases, the orange line is the mean, and the envelope is the 95% confidence interval, determined by bootstrapping.

buffer distance (0, 5, or 10 km), *W* is linked to *S* and all other storm objects in the same track.

Third, predictors are calculated for each storm object. There are four types of predictors: radar statistics (mean, standard deviation, skewness, kurtosis, and seven percentiles calculated for each of 12 variables, based only on pixels inside the storm object; the same statistics are calculated for gradient magnitudes of the 12 variables), storm motion (speed and direction), shape parameters (area, orientation, eccentricity, etc., of the storm object), and sounding indices (both dynamic and thermodynamic). Sounding indices are calculated from interpolated RUC data using the Sounding and Hodograph Analysis and Research Program in Python (SHARPpy) software (Halbert et al. 2015). There are a total of 431 predictors. The fourth step is to label each storm object *S*. If *S* is linked to a wind observation ≥50 kt (25.7 m s^{-1}) over the given buffer distance and time window, its label is "true" (Fig. 4a).

For each buffer distance and time window, a GBRT ensemble is trained. Then, isotonic regression (IR) is trained with independent data (no case within 24 h of a GBRT-training case) to bias correct the GBRTs. Next, the calibrated model (GBRT + IR) is tested on independent data. Results are shown in Fig. 4 for the median buffer distance (5 km) and lead time (30–45 min). The model shown in Fig. 4 is an ensemble of 500 GBRTs trained with the AdaBoost algorithm (Freund and Schapire 1997), resampling

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 09/24/21 06:09 PM UTC

factor of 0.15 (with replacement), learning rate of 0.1, 25 variables tested per branch node, and a minimum of 10 storm objects per leaf node. Results are based on 12,155 test cases. No premodel variable selection was done, because decision trees perform built-in variable selection. The area under the receiver operating characteristic (ROC) curve (AUC) is >0.9, which is generally considered excellent (Luna-Herrera et al. 2003; Muller et al. 2005; Mehdi et al. 2011), and the reliability curve (Fig. 4d) is very close to perfect ($x = y$). Furthermore, the maximum critical success index (CSI) occurs with a frequency bias of 1.0 (unbiased model), which suggests that bias need not be sacrificed for other performance metrics. These results are based on Lagerquist (2016).

Severe hail. Prediction of hail occurrence and size days to hours ahead is needed to guide the issuance of convective outlooks and watches. Convection-allowing model (CAM) ensembles provide information about storm intensity, location, and evolution but do not forecast maximum hail size at the surface directly. Machine-learning models have been developed to predict the probability of hail occurrence and the expected hail-size distribution given information about storms and their environment from CAM output. The machine-learning hail models have been run in real time on two CAM ensemble systems and have been validated against the HAILCAST diagnostic (Adams-Selin and Ziegler 2016) and storm surrogate variables, such as updraft helicity (Sobash et al. 2016).

A storm-centered method is used for producing machine-learning hail forecasts. First, potential hailstorms are identified from the hourly maximum column total graupel field in the 2014 and 2015 Center for Analysis and Prediction of Storms (CAPS) CAM ensemble using the enhanced watershed feature identification technique. Observed hailstorms are identified from the maximum expected size of hail (MESH) field (Witt et al. 1998) in the NOAA National Severe Storms Laboratory (NSSL) Multi-Radar Multi-Sensor mosaic (Smith et al. 2016). Both forecast and observed storms are tracked through time and then matched based on proximity in space and time. Statistics describing the storm and environmental variables from within the bounds of each forecast storm are extracted and are used as input into the machine-learning models. A gamma distribution is fit to the distribution of MESH within an observed hailstorm, and the parameters of the gamma distribution are used as target labels for the machine-learning models.

An RF classification model predicts whether hail will occur based on whether an observed storm was matched with a given forecast storm and the hail-size distribution parameters given that hail occurred. An RF regression model estimates both the shape and scale parameters of the gamma distribution simultaneously to preserve the correlations among the parameters in the predictions. Gridded hail-size forecasts are produced by sampling hail sizes from the predicted distribution and applying them in rank order onto the column total graupel field. Potential hailstorms with less than 50% chance of hail occurrence are removed from the grid.

Verification results and a single forecast case are shown in Fig. 5 for the machine-learning hail forecasts and other storm surrogate probability forecasts, including HAILCAST, column total graupel, and updraft helicity. The RF used for this experiment was trained on CAPS ensemble forecasts from May to June 2014 and evaluated on CAPS ensemble forecasts for the same period in 2015. These results were based on analysis from Gagne (2016). The performance diagram (Roebber 2009) in Fig. 5a shows that for a given probability threshold, the machine-learning models tend to have fewer false alarms, lower frequency bias, and higher accuracy than other methods. The attributes diagram (Hsu and Murphy 1986) in Fig. 5b indicates the probabilities from the machine-learning models and updraft helicity are generally reliable, while other methods tend to produce probabilities that are overconfident. The case study in Fig. 5c shows that the RF model performed best at capturing the area where 50-mm hail occurred. The other two methods had both lower probabilities and enhanced probabilities in areas where 50-mm hail did not occur.

Precipitation classification. The Meteorological Phenomena Identification Near the Ground (mPING; Elmore et al. 2014) project has collected over 1.1 million observations since its launch on 19 December 2012. The mPING project uses crowdsourced observations of precipitation type (ptype) submitted anonymously through a smartphone app. Various other weather conditions can also be reported, such as floods, visibility restrictions, wind damage, hail, and tornadoes. The ptype observations have been used to help characterize the sensitivity of various ptype algorithms to model errors (Reeves et al. 2014) and to verify current NWP model performance and, in the process, find an outright error within the postprocessing of the RAP model (Elmore et al. 2015).

Given that the skill of NWP ptype forecasts have been characterized with mPING observations, a compelling next step is to use the mPING observations to build a new, hopefully improved, ptype algorithm. As

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 09/24/21 06:09 PM UTC

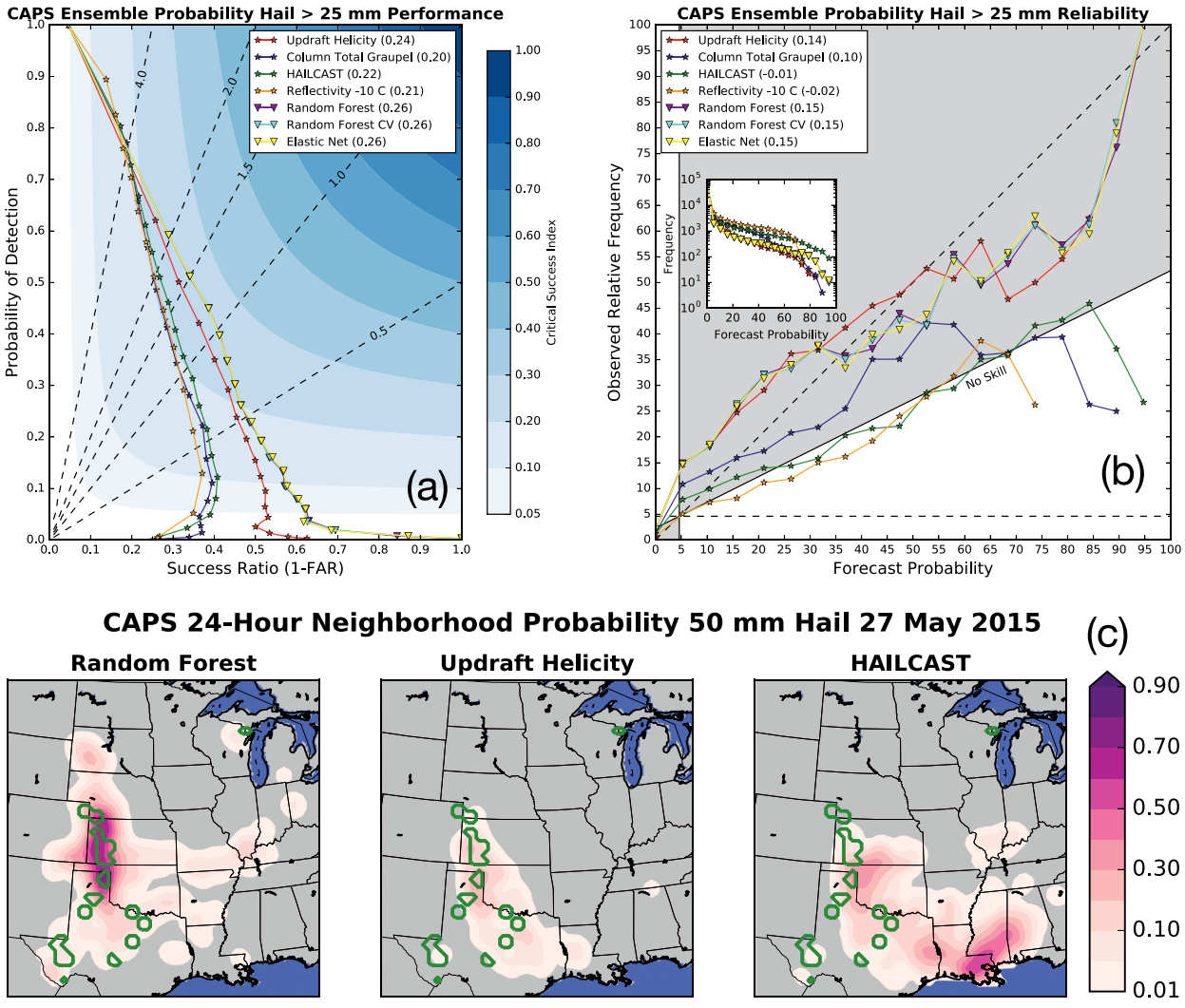


Fig. 5. (a) Performance diagram comparing different hail forecasting methods. (b) Attributes diagram indicating the reliability of different forecasting methods. (c) Forecast case showing the probability of 50-mm-or-larger hail in filled contours and observed 50-mm hail in green contours.

a first attempt, the wet-bulb temperature T_w profiles from 5,000 m AGL to the surface created by each NWP model are characterized as one of four different types that are identical to the four types described in Schuur et al. (2012): type 1 is all T_w below freezing (273.16 K); type 2 has one freezing level such that T_w at the surface is above freezing; type 3 has three freezing levels with an elevated warm layer, an elevated cold layer, and T_w at the surface above freezing; and type 4 is the "classic" elevated-warm-layer profile with T_w at the surface below freezing. Multiple predictors are computed for each profile type, including area above and below freezing for each layer, height of the various freezing levels, wind shear [both bulk and in latitudinal and meridional (u and v, respectively) directions] in warm and cold layers and across the entire depth of the profile, area of relative humidity (RH) above and below 0.8 for each layer along with the mean RH in each layer, and minimum T_w in the cold surface layers. Each profile type has a different set of predictors, though some predictors are common across all profile types. Overall, type 1 profiles have 28 predictors, type 2 profiles have 23, type 3 profiles have 49, and type 4 profiles have 38.

Because each profile type has a different set of predictors, each has its own RF. Training data consist of 80% of the available hours of data selected randomly. The remaining 20% of the hours are used for testing. Hours, instead of individual observations, are chosen so as to lessen cross contamination of testing data with training data. Thus, a training profile and a testing profile cannot come from the same hour.

These data are not balanced, in that there are far more snow and rain reports than ice pellets and

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 09/24/21 06:09 PM UTC

freezing rain. Sampling weights and maximum tree size are adjusted by trial and error such that the bias for each of the four classes generated by each random forest is close to 1. No other adjustments are made.

Applying RFs in this way results in marked improvement in the ptype prediction form NWP models. Figure 6 is an example of this improvement for the Rapid Refresh (RAP) model over the cold season of 2014/15 with confidence intervals for each score. The right set of bars shows the output of the RAP postprocess ptype algorithm while the left set of bars displays the results of an RF ptype algorithm. Scores for the RF algorithm come from a smaller number of cases (the test data) than the scores for the RAP, which use the entire available dataset. There is not much room for improvement in predicting rain and snow, but the improvement for freezing rain and ice pellets is quite dramatic. In addition, the RF ptype output is unbiased, unlike the postprocessed ptype output. RFs can also provide probabilistic information about the ptype, which will likely be useful to operational forecasters and those involved in maintaining infrastructure systems. Clearly, if sufficient data are available, an RF approach to forecast ptype can lead to significant improvement to the most troublesome winter precipitation types.

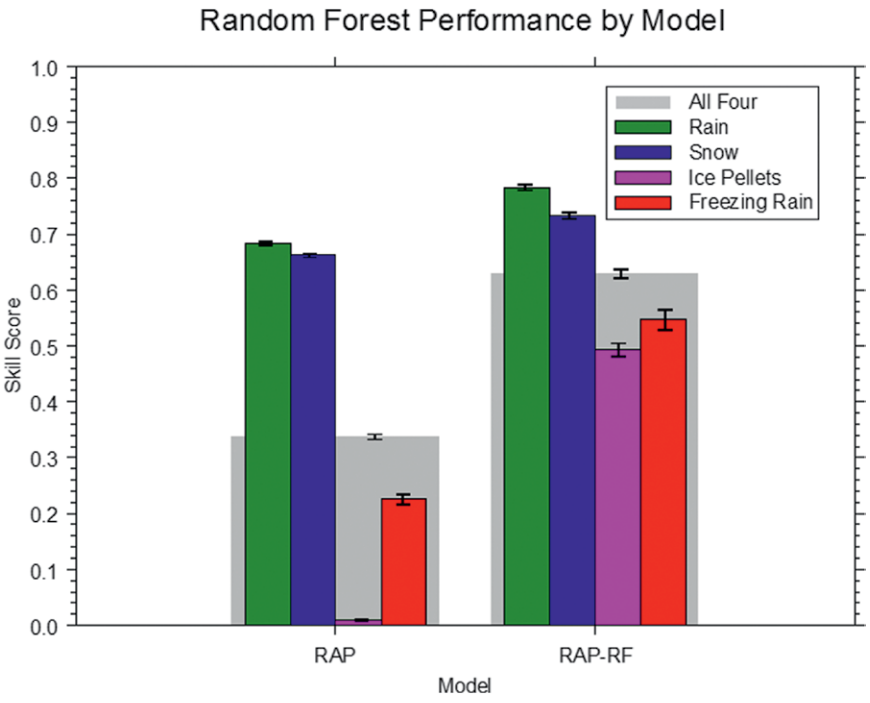


Fig. 6. Scores for the RAP postprocess ptype algorithm (left set of bars) and the RF ptype algorithm (right set of bars) based on mPING data for winter 2014/15. Score values are oriented such that larger positive numbers are better. The colored bars show the Pierce skill score for rain (green), snow (blue), ice pellets (magenta), and freezing rain (red). The gray bar is the Gerrity score for all four types taken together ordered as snow, rain, ice pellets, and freezing rain (Elmore et al. 2015).

Variable importance is examined for each forest for each model. No variable stands out as much more important than another. At the most extreme, the most important variable is roughly twice as important as the least important variable. No variable in particular stands out; because of this characteristic, variable selection is deemed unnecessary.

Renewable energy. Forecasting for renewable energy resources is another example of high-impact weather forecasts. In this case, forecasting enables using clean, locally available, but highly variable renewable resources to produce energy in place of fossil fuel energy sources. Because the wind, water, and solar resources are highly variable, forecasting is needed to blend renewable power with other energy sources to assure reliable, efficient, and economic deployment. Utilities require forecasts on various scales. Here, we describe two shorter-range scales: the nowcast, for the next 3–6 h, and the day-ahead forecast (which can extend to 72 h to cover weekends). The nowcast is necessary to blend renewable energy into the grid in order to meet the electric load in real time. The day-ahead forecast is used for planning unit allocation and trading energy with other utilities. We specifically discuss how AI is used for forecasting for wind and solar energy, with more detailed descriptions of additional prediction methods being provided by Ahlstrom et al. (2013), Orwig et al. (2014), Tuohy et al. (2015), and Haupt et al. (2016).

The nowcast typically leverages observations from the wind or solar plant or remotely sensed data. The goal is to improve upon a persistence forecast at the location of the plant. Statistical learning and AI methods capture changes or deviations from persistence. One statistical learning method for wind speed nowcasts is the Markov-switching vector autoregressive

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 09/24/21 06:09 PM UTC

model (Hering et al. 2015). Solar power nowcasting has leveraged various statistical learning methods. Hassanzadeh et al. (2010) and Yang et al. (2012) used autoregressive integrated moving average (ARIMA) models to predict solar irradiance and power, demonstrating lower errors than other time series models. ANNs are commonly used for nonlinear solar predictions (Mellit 2008) and have shown skill over other baseline techniques (Marquez and Coimbra 2011; Wang et al. 2012; Chu et al. 2013). Support vector machines have also shown skill over linear regression in postprocessing NWP model output (Sharma et al. 2011).

Solar models typically predict the clearness index, the ratio of the global horizontal irradiance (GHI) that reaches the surface of Earth to that at the top of the atmosphere. The clearness index ranges between 0 and 1 and depicts the depletion of solar energy via absorption and scattering by clouds and aerosols on its path through the atmosphere. It also removes the effects of the seasonal cycles and partially accounts for diurnal effects. One can explicitly compute the GHI at the top of the atmosphere given the solar angle and location information.

Some recent work has sought to identify regimes and forecast solar irradiance changes specific to those regimes through both implicit and explicit methods. The implicit method employs a regression tree approach (Quinlan 1996) with an embedded nearest neighbor scheme to forecast both deterministic irradiance and its variability (McCandless et al. 2015). Explicit regime identification using *k*-means clustering and training ANNs for each cluster was shown to improve over training a single ANN on the entire training dataset (McCandless et al. 2016b,a). These approaches to statistical forecasting outperformed a "smart persistence" approach that includes the change in solar angle. When compared to other nowcasting products, the statistical forecasting approach outperformed all others for the first hour (Haupt et al. 2016), as demonstrated in Fig. 7.

Day-ahead forecasting approaches use AI models to postprocess and correct NWP model output toward observations. Common methods of postprocessing include ANNs and blended optimization methods. The Dynamic Integrated Forecast (DICast) system (Myers et al. 2011; Mahoney et al. 2012) first applies a dynamic MOS approach followed by optimized blending. This system has improved forecasts of wind and solar power by at least 15% (Mahoney et al. 2012; Haupt et al. 2016).

For true decision support, utilities and grid operators do not want only wind speed or GHI forecasts; they actually require power predictions. Although manufacturers of wind turbines and solar panels provide average power curves, these are not perfectly representative of actual power produced at a site because of variation in terrain elevation, turbulence, and other factors. Thus, training an AI method to convert from wind or GHI to power can produce better power predictions for a specific site (Parks et al. 2011) and does not require the detailed metadata needed to apply alternative methods for solar irradiance (Haupt and Kosovic 2016). The National Center for Atmospheric Research (NCAR) has successfully applied the cubist regression tree approach to both wind (Kosovic et al. 2015) and solar (Haupt and Kosovic 2016).

Finally, many utilities request probabilistic

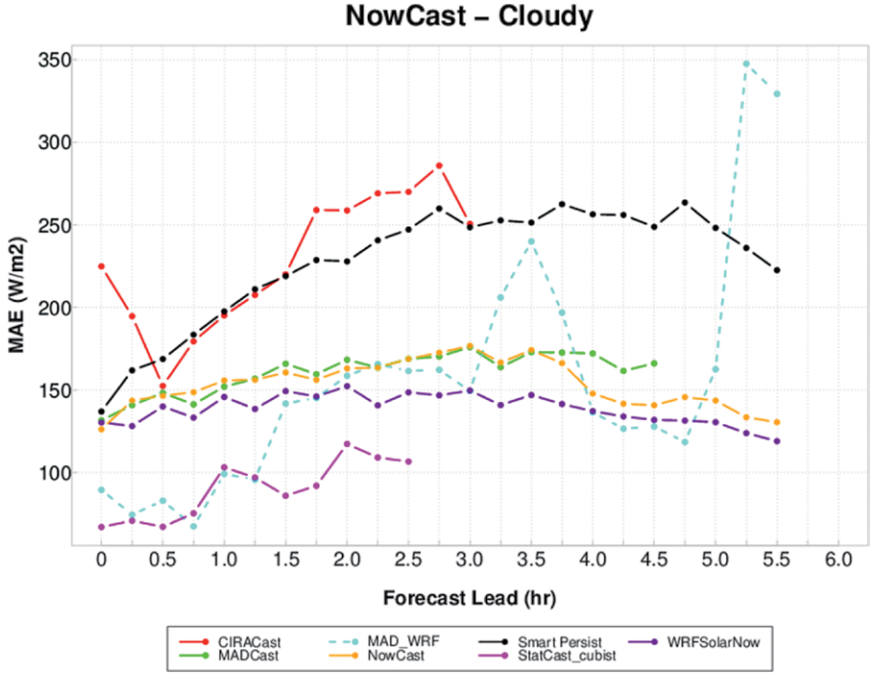


Fig. 7. Mean absolute error (MAE; W m^{-2}) calculated over a 15-month period for all nowcast components aggregated over all sites (New York, Colorado, and California) for cloudy conditions. Note that StatCast performs the best, on average, for these difficult-to-forecast conditions.

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 09/24/21 06:09 PM UTC

predictions to estimate the forecast uncertainty and to plan their reserve requirements. Although NWP model ensembles traditionally provide probabilistic forecasts, the analog ensemble approach (AnEn; Delle Monache et al. 2013) has successfully produced probabilistic forecasts based on a single high-quality forecast from a consistent prediction system. The AnEn searches through historical forecasts for those most similar to the current forecast. Observations associated with each historical forecast form a probability density function that defines the forecast uncertainty. The AnEn mean can correct for systematic biases. This approach has proven to be at least as reliable as some of the best dynamical ensembles for wind speed (Delle Monache et al. 2013; Haupt and Delle Monache 2014), wind power (Kosovic et al. 2015), and solar power (Alessandrini et al. 2015).

AI methods are directly providing decision support for utilities and grid operators around the world and are enabling increases in the deployment of the variable renewable energy resources. All of the methods described in this section have been operationalized and used by utilities. In this way, enabling higher capacities of renewable energy can lead to energy security, lower the use of water in energy production, and lower the emissions of carbon dioxide and other pollutants, thus providing the world with a clean source of sustainable energy.

Aviation turbulence. Although much of the severe weather of concern to humans occurs near the surface, conditions far above the ground may be equally hazardous. Commercial aviation is impacted by various weather threats, including airframe icing by supercooled liquid water, engine flameouts in areas of high ice water content, hail, lightning, and atmospheric turbulence. Turbulence is one of the most significant en route aviation hazards from an operational standpoint. Flying through turbulent eddies causes an aircraft to bounce from side to side and up and down, making passengers and crew uncomfortable and occasionally injuring them or damaging the aircraft. Turbulence is created by wind shear in regions of low stability, which may result from jet streams and fronts, mountain-wave or convectively induced gravity wave breaking, or the updrafts and downdrafts of thunderstorms. Because it is often a small-scale and fundamentally stochastic phenomenon, turbulence is difficult to forecast or even nowcast. Moreover, NWP models are not generally tuned to accurately forecast aviation-scale turbulence, and output variables such as subgrid turbulent kinetic energy (TKE) are not skillful in predicting aircraft observations of turbulence (Sharman 2016).

AI has become a key tool for observing, nowcasting, and forecasting aviation turbulence. For observing turbulence in clouds and storms, a fuzzy logic algorithm was developed to carefully quality control ground-based Doppler radar spectrum width measurements, allowing them to be scaled and combined into an estimate of the turbulence eddy dissipation rate (EDR). Fuzzy logic is a tool for building expert systems that mimic human reasoning, smoothly combining various sources of evidence to form a final assessment (Williams 2009). For the turbulence detection algorithm, the likelihood of radar spectrum width contamination is scored as a "confidence" between 0 and 1 for each of several diagnostic quantities derived from the radar signal or its spatial context and then these are combined in a geometric average to obtain an overall assessment. The spectrum widths are scaled to EDR based on distance from the radar, and a confidence-weighted average is performed to obtain the final EDR estimate (Williams and Meymaris 2016).

Aviation turbulence forecasting utilizes diagnostics, or indices, computed from NWP model-resolved wind shear, stability, and various other functions of the modeled variables (Sharman 2016). Although none of these explicitly represents aircraft-scale turbulence, the form of the turbulent energy cascade means that they may be related to it and thus may be transformed and weighted to form a good estimate. The Graphical Turbulence Guidance (GTG) algorithm (Sharman et al. 2006) evaluates each diagnostic against aircraft observations of turbulence, rescales it using a piecewise linear function, and uses weights based on the resulting skill scores to compute a weighted-mean consensus. More recent versions of GTG incorporate lognormal remapping functions. A weakness of this approach is that it does not take into account the linear and nonlinear dependencies between the diagnostics, many of which are highly correlated.

Decision-tree-based techniques offer the ability to incorporate features not proportional or even monotonically related to turbulence severity. Williams (2014) used RFs to combine both NWP diagnostics and features derived from satellite and radar products to create turbulence nowcasts. Predictors included NWP-derived turbulence diagnostics and thermodynamic variables such as convective available potential energy (CAPE) and convective inhibition (CIN); distances to relevant reflectivity, echo top, lightning, and in-cloud turbulence objects; and disc statistics over various radii from both the radar and satellite

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 09/24/21 06:09 PM UTC

imagery. Several hundred candidate predictors were whittled down first through the RF's variable importance analyses and then through forward and backward selection, where an RF is trained and evaluated on independent datasets and the predictor variables producing the best discrimination skill are preserved. The RF is then calibrated to produce either EDR or turbulence probability, and the resulting algorithm is run at every point in a predefined grid to produce a map suitable for use by pilots, dispatchers, or air traffic controllers. Although the benefits of an AI approach are particularly clear for fusing multiple data sources for turbulence nowcasting, Table 1 indicates that logistic regression, *k*-nearest neighbor, and especially RF exceed GTG's skill even in the case when only NWP model data are used as predictors. A similar approach has been used to forecast convection (Mecikalski et al. 2015; Ahijevych et al. 2016). A downside of this approach is the need for significant feature engineering, that is, calculating many different features and then testing which are relevant. This requirement is somewhat mitigated by McGovern et al. (2014), who used spatiotemporal relational random forests guided by a schema identifying possibly relevant relationships between an aircraft location and various storm-related objects. In the future, convolutional neural networks operating in a deep learning framework may reduce the need for feature engineering even further. The use of AI for turbulence prediction will continue to make flights safer and more comfortable.

DISCUSSION. Application of modern AI techniques to high-impact weather forecasting is improving our ability to sift through the deluge of big data to extract insights and accurate, timely guidance for human weather forecasters and decision-makers. AI techniques build on traditional methods, such as MOS, by providing more flexible and powerful models capable of identifying complex relationships between a huge number of modeled and observed weather features or derived quantities. In addition, AI methods extend easily to directly predicting impacts of high-impact weather, such as power generated by variable sources such as solar or wind, energy consumption in an area, or airport arrival capacity.

This paper raises the interesting question of the role of automated guidance in forecasts. While we have demonstrated that AI/data science techniques can be used to significantly improve forecasts in a variety of high-impact weather domains, it is not simply a matter of bringing these techniques to operations. The forecasters must be able to trust the forecast produced by such techniques, as has been demonstrated in the HWT/PHI experiments (Karstens et al. 2016).

For forecasts of standard weather variables, such as temperature and precipitation, the NWS currently operates with a human-in-the-loop paradigm in which forecasters subjectively blend and adjust multiple sources. Local offices add predictive value in situations where local effects have a larger impact on the forecast. At the NWS Weather Prediction Center, which issues temperature and precipitation forecasts over the entire United States, the human forecasts now perform significantly worse than downscaled, bias-corrected ensemble forecasts for temperature and precipitation (Novak et al. 2014). Official NWS track forecasts of hurricanes, a major form of high-impact weather, also perform worse than weighted ensemble consensus forecasts (Cangialosi and Franklin 2015). There are also issues with spatial discontinuities in forecasts and warnings between the domains of different forecast offices (Gilbert et al. 2015). Private weather firms, including The Weather Company, operate in a human-over-the-loop paradigm in which an optimal blend of bias-corrected model output is generated as needed by users, and human forecasters can add filters and qualifiers to account for observed short-term biases or data quality issues (Williams et al. 2016). This approach scales easily and only requires a small team of meteorologists to oversee a mostly automated system. The downside of a heavily

Table 1. High-altitude turbulence forecast skill scores for years 2010 and 2011, evaluated using pilot reports and automated aircraft turbulence reports. AI methods were trained on 40,000 random samples from 2011 with 30% turbulence cases and evaluated on all of 2010, and vice versa. The *k*-nearest neighbors method used 100 analogs. TSS is the true skill score.

Method	Year	ROC AUC	Max CSI	Max TSS
GTG weighted mean	2010	0.791	0.137	0.443
	2011	0.775	0.132	0.418
Logistic regression	2010	0.822	0.162	0.496
	2011	0.805	0.149	0.461
k-nearest neighbors	2010	0.832	0.167	0.514
	2011	0.818	0.163	0.482
Random forest	2010	0.849	0.179	0.541
	2011	0.830	0.169	0.499

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 09/24/21 06:09 PM UTC

automated approach is that forecasters may become disengaged from the forecast process (Pliske et al. 2004) and struggle to take appropriate corrective action when automation fails (Skitka et al. 1999; Pagano et al. 2016).

By studying the error characteristics of different machine-learning methods in high-impact weather situations, researchers and forecasters can identify when the automated guidance should be trusted and when it is more likely to struggle. The methods presented in this paper are able to blend physical knowledge with automated corrections to produce critical products in this age of information overload.

ACKNOWLEDGMENTS. This material is based upon work supported by the National Science Foundation under Grant SHARP NSF AGS-126-1776. Funding was provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA–University of Oklahoma Cooperative Agreement NA11OAR4320072, U.S. Department of Commerce. This work was supported by the NEXRAD Product Improvement Program, by NOAA/Office of Oceanic and Atmospheric Research. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of NOAA, the U.S. Department of Commerce, or the University of Oklahoma. NCAR is sponsored by the National Science Foundation. The authors thank Tara Jensen for producing Fig. 7.

REFERENCES

Adams-Selin, R. D., and C. L. Ziegler, 2016: Forecasting hail using a one-dimensional hail growth model within WRF. *Mon. Wea. Rev.*, **144**, 4919–4939, doi:10.1175/MWR-D-16-0027.1.

Adrianto, I., T. Trafalis, and V. Lakshmanan, 2009: Support vector machines for spatiotemporal tornado prediction. *Int. J. Gen. Syst.*, **38**, 759–776, doi:10.1080/03081070601068629.

Ahijevych, D., J. O. Pinto, J. K. Williams, and M. Steiner, 2016: Probabilistic forecasts of mesoscale convective system initiation using the random forest data mining technique. *Wea. Forecasting*, **31**, 581–599, doi:10.1175/WAF-D-15-0113.1.

Ahlstrom, M., and Coauthors, 2013: Knowledge is power: Efficiently integrating wind energy and wind forecasts. *IEEE Power Energy Mag.*, **11**, 45–52, doi:10.1109/MPE.2013.2277999.

Alessandrini, S., L. D. Monache, S. Sperati, and G. Cervone, 2015: An analog ensemble for short-term probabilistic solar power forecast. *Appl. Energy*, **157**, 95–110, doi:10.1016/j.apenergy.2015.08.011.

Allen, C. T., S. E. Haupt, and G. S. Young, 2007: Source characterization with a genetic algorithm–coupled dispersion–backward model incorporating SCIPUFF. *J. Appl. Meteor. Climatol.*, **46**, 273–287, doi:10.1175/JAM2459.1.

Anagnostou, E., 2004: A convective/stratiform precipitation classification algorithm for volume scanning weather radar observations. *Meteor. Appl.*, **11**, 291–300, doi:10.1017/S1350482704001409.

Baldwin, M., J. Kain, and S. Lakshmivarahan, 2005: Development of an automated classification procedure for rainfall systems. *Mon. Wea. Rev.*, **133**, 844–862, doi:10.1175/MWR2892.1.

Bankert, R., 1994: Cloud classification of AVHRR imagery in maritime regions using a probabilistic neural network. *J. Appl. Meteor.*, **33**, 909–918, doi:10.1175/1520-0450(1994)033<0909:CCOAII>2.0.CO;2.

Benjamin, S., and Coauthors, 2004: An hourly assimilation–forecast cycle: The RUC. *Mon. Wea. Rev.*, **132**, 495–518, doi:10.1175/1520-0493(2004)132<0495:AHACTR>2.0.CO;2.

Billet, J., M. DeLisi, B. Smith, and C. Gates, 1997: Use of regression techniques to predict hail size and the probability of large hail. *Wea. Forecasting*, **12**, 154–164, doi:10.1175/1520-0434(1997)012<0154:UORTTP>2.0.CO;2.

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

Cangialosi, J. P., and J. L. Franklin, 2015: 2014 National Hurricane Center forecast verification report. National Hurricane Center Tech. Rep., 82 pp. [Available online at www.nhc.noaa.gov/verification/pdfs/Verification_2014.pdf.]

Chisholm, D., J. Ball, K. Veigas, and P. Luty, 1968: The diagnosis of upper-level humidity. *J. Appl. Meteor.*, **7**, 613–619, doi:10.1175/1520-0450(1968)007<0613:TDOULH>2.0.CO;2.

Chu, Y., H. Pedro, and C. Coimbra, 2013: Hybrid intra-hour DNI forecasts with sky image processing enhanced by stochastic learning. *Sol. Energy*, **98**, 592–603, doi:10.1016/j.solener.2013.10.020.

Cintineo, J. L., M. J. Pavolonis, J. M. Sieglaff, and D. T. Lindsey, 2014: An empirical model for assessing the severe weather potential of developing convection. *Wea. Forecasting*, **29**, 639–653, doi:10.1175/WAF-D-13-00113.1.

Clark, A. J., A. MacKenzie, A. McGovern, V. Lakshmanan, and R. Brown, 2015: An automated, multi-parameter dryline identification algorithm. *Wea. Forecasting*, **30**, 1781–1794, doi:10.1175/WAF-D-15-0070.1.

CoreLogic, 2016: CoreLogic pegs total damage from Texas spring hail storms at nearly $700m. Insurance Journal. [Available online at www.insurancejournal.com/news/southcentral/2016/07/12/419831.htm.]

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 09/24/21 06:09 PM UTC

Delle Monache, L., T. Eckel, D. Rife, B. Nagarajan, and K. Searight, 2013: Probabilistic weather predictions with an analog ensemble. *Mon. Wea. Rev.*, **141**, 3498–3516, doi:10.1175/MWR-D-12-00281.1.

ECMWF, 2016: ECMWF strategy 2016–2025: The strength of a common goal. ECMWF Tech. Rep., 32 pp. [Available online at www.ecmwf.int/sites/default/files/ECMWF_Strategy_2016-2025.pdf.]

Elmore, K. L., and M. B. Richman, 2001: Euclidean distance as a similarity metric for principal component analysis. *Mon. Wea. Rev.*, **129**, 540–549, doi:10.1175/1520-0493(2001)129<0540:EDAASM>2.0.CO;2.

——, and H. Grams, 2016: Using mPING data to generate random forests for precipitation type forecasts. *14th Conf. on Artificial and Computational Intelligence and Its Applications to the Environmental Sciences*, New Orleans, LA, Amer. Meteor. Soc., 4.2. [Available online at https://ams.confex.com/ams/96Annual/webprogram/Paper289684.html.]

——, Z. L. Flamig, V. Lakshmanan, B. T. Kaney, V. Farmer, H. D. Reeves, and L. P. Rothfusz, 2014: mPING: Crowd-sourcing weather reports for research. *Bull. Amer. Meteor. Soc.*, **95**, 1335–1342, doi:10.1175/BAMS-D-13-00014.1.

——, ——, ——, ——, ——, ——, and ——, 2015: Verifying forecast precipitation type with mPING. *Wea. Forecasting*, **30**, 656–657, doi:10.1175/WAF-D-14-00068.1.

Freund, Y., and R. Schapire, 1997: A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.*, **55**, 119–139, doi:10.1006/jcss.1997.1504.

Friedman, J. H., 2002: Stochastic gradient boosting. *Comput. Stat. Data Anal.*, **38**, 367–378, doi:10.1016/S0167-9473(01)00065-2.

Gagne, D. J., II, 2016: Coupling data science techniques and numerical weather prediction models for high-impact weather prediction. Ph.D. dissertation, University of Oklahoma, 185 pp. [Available online at http://hdl.handle.net/11244/44917.]

——, A. McGovern, and J. Brotzge, 2009: Classification of convective areas using decision trees. *J. Atmos. Oceanic Technol.*, **26**, 1341–1353, doi:10.1175/2008JTECHA1205.1.

Gilbert, K. K., J. P. Craven, D. R. Novak, T. M. Hamill, J. Sieveking, D. P. Ruth, and S. J. Lord, 2015: An introduction to the national blend of global models project. *Special Symp. on Model Postprocessing and Downscaling*, Phoenix, AZ, Amer. Meteor. Soc., 3.1. [Available online at https://ams.confex.com/ams/95Annual/webprogram/Paper267282.html.]

Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203–1211, doi:10.1175/1520-0450(1972)011<1203:TUOMOS>2.0.CO;2.

Halbert, K., W. Blumberg, and P. Marsh, 2015: SHARPpy: Fueling the Python cult. *Fifth Symp. on Advances in Modeling and Analysis Using Python*, Phoenix, AZ, Amer. Meteor. Soc., 402. [Available online at https://ams.confex.com/ams/95Annual/webprogram/Paper270233.html.]

Hassanzadeh, M., M. Etezadi-Amoli, and M. Fadali, 2010: Practical approach for sub-hourly and hourly prediction of PV power output. *North American Power Symp.*, Arlington, TX, Institute of Electrical and Electronics Engineers, 5618944, doi:10.1109/NAPS.2010.5618944.

Haupt, S. E., and L. Delle Monache, 2014: Understanding ensemble prediction: How probabilistic wind power prediction can help in optimizing operations. WindTech International. [Available online at www.windtech-international.com/editorial-features/understanding-ensemble-prediction.]

——, and B. Kosovic, 2016: Variable generation power forecasting as a big data problem. *IEEE Trans. Sustainable Energy*, **8**, 725–732, doi:10.1109/TSTE.2016.2604679.

——, A. Pasini, and C. Marzban, Eds., 2008: *Artificial Intelligence Methods in the Environmental Sciences*. Springer, 424 pp.

——, and Coauthors, 2016: The Sun4Cast solar power forecasting system: The result of the public-private-academic partnership to advance solar power forecasting. NCAR Tech. Note NCAR/TN-526+STR, 307 pp., doi:10.5065/D6N58JR2.

Hering, A. S., K. Kazor, and W. Kleiber, 2015: A Markov-switching vector autoregressive stochastic wind generator for varying spatial and temporal scales. *Resources*, **4**, 70–92, doi:10.3390/resources4010070.

Hoerl, A., and R. Kennard, 1988: Ridge regression. *Encyclopedia of Statistical Sciences*, Vol. 8, Wiley, 129–136.

Hsu, W.-R., and A. H. Murphy, 1986: The attributes diagram: A geometrical framework for assessing the quality of probability forecasts. *Int. J. Forecast.*, **2**, 285–293, doi:10.1016/0169-2070(86)90048-8.

Karstens, C. D., and Coauthors, 2015: Evaluation of a probabilistic forecasting methodology for severe convective weather in the 2014 hazardous weather testbed. *Wea. Forecasting*, **30**, 1551–1570, doi:10.1175/WAF-D-14-00163.1.

——, and Coauthors, 2016: Forecaster decision-making with automated probabilistic guidance in the 2015 Hazardous Weather Testbed probabilistic hazard information experiment. *Fourth Symp. on Building a Weather-Ready Nation: Enhancing Our Nation's*

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 09/24/21 06:09 PM UTC

Readiness, Responsiveness, and Resilience to High Impact Weather Events, New Orleans, LA, Amer. Meteor. Soc., 4.2. [Available online at https://ams.confex.com/ams/96Annual/webprogram/Paper286854.html.]

Key, J., J. Maslanik, and A. Schweiger, 1989: Classification of merged AVHRR and SMMR Arctic data with neural networks. *Photogramm. Eng. Remote Sensing*, **55**, 1331–1338.

Kitzmiller, D., W. McGovern, and R. Saffle, 1995: The WSR-88D severe weather potential algorithm. *Wea. Forecasting*, **10**, 141–159, doi:10.1175/1520-0434(1995)010<0141:TWSWPA>2.0.CO;2.

Klein, W. H., B. M. Lewis, and I. Enger, 1959: Objective prediction of five-day mean temperatures during winter. *J. Meteor.*, **16**, 672–682, doi:10.1175/1520-0469(1959)016<0672:OPOFDM>2.0.CO;2.

Kosovic, B., and Coauthors, 2015: Comprehensive forecasting system for variable renewable energy. *International Conf. on Energy and Meteorology*, Boulder, CO, 23 pp. [Available online at www.wemcouncil.org/wp/wp-content/uploads/2015/07/1830_Branko Kosovic.pdf.]

Lagerquist, R., 2016: Using machine learning to predict damaging straight-line convective winds. M.S. thesis, School of Meteorology, University of Oklahoma, 251 pp. [Available online at http://hdl.handle.net/11244/44921.]

Lakshmanan, V., and T. Smith, 2010: Evaluating a storm tracking algorithm. *26th Conf. on Interactive Information Processing Systems*, Atlanta, GA, Amer. Meteor. Soc., 8.2. [Available online at https://ams.confex.com/ams/90annual/techprogram/paper_162556.htm.]

——, R. Rabin, and V. DeBrunner, 2000: Identifying and tracking storms in satellite images. *Second Artificial Intelligence Conf.*, Long Beach, CA, Amer. Meteor. Soc., 90–95. [Available online at https://ams.confex.com/ams/annual2000/techprogram/paper_339.htm.]

——, G. Stumpf, and A. Witt, 2005: A neural network for detecting and diagnosing tornadic circulations using the mesocyclone detection and near storm environment algorithms. *21st Int. Conf. on Information Processing Systems*, San Diego, CA, Amer. Meteor. Soc., J5.2. [Available online at https://ams.confex.com/ams/Annual2005/webprogram/Paper82772.html.]

——, A. Fritz, T. Smith, K. Hondl, and G. J. Stumpf, 2007: An automated technique to quality control radar reflectivity data. *J. Appl. Meteor. Climatol.*, **46**, 288–305, doi:10.1175/JAM2460.1.

——, J. Zhang, and K. Howard, 2010: A technique to censor biological echoes in radar reflectivity data. *J. Appl. Meteor. Climatol.*, **49**, 453–462, doi:10.1175/2009JAMC2255.1.

——, C. Karstens, J. Krause, and L. Tang, 2014: Quality control of weather radar data using polarimetric variables. *J. Atmos. Oceanic Technol.*, **31**, 1234–1249, doi:10.1175/JTECH-D-13-00073.1.

——, B. Herzog, and D. Kingfield, 2015: A method for extracting postevent storm tracks. *J. Appl. Meteor. Climatol.*, **54**, 451–462, doi:10.1175/JAMC-D-14-0132.1.

Luna-Herrera, J., G. Martinez-Cabrera, R. Parra-Maldonado, J. Enciso-Moreno, J. Torres-Lopez, F. Quesada-Pascual, R. Delgadillo-Polanco, and S. Franzblau, 2003: Use of receiver operating characteristic curves to assess the performance of a microdilution assay for determination of drug susceptibility of clinical isolates of *Mycobacterium tuberculosis*. *Eur. J. Clin. Microbiol. Infect. Dis.*, **22**, 21–27, doi:10.1007/s10096-002-0855-5.

Mahoney, W., and Coauthors, 2012: A wind power forecasting system to optimize grid integration. *IEEE Trans. Sustainable Energy*, **3**, 670–682, doi:10.1109/TSTE.2012.2201758.

Malone, T., 1955: Application of statistical methods in weather prediction. *Proc. Natl. Acad. Sci. USA*, **41**, 806–815, doi:10.1073/pnas.41.11.806.

Manzato, A., 2013: Hail in northeast Italy: A neural network ensemble forecast using sounding-derived indices. *Wea. Forecasting*, **28**, 3–28, doi:10.1175/WAF-D-12-00034.1.

Marquez, R., and C. Coimbra, 2011: Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database. *Sol. Energy*, **85**, 746–756, doi:10.1016/j.solener.2011.01.007.

Marzban, C., and G. Stumpf, 1996: A neural network for tornado prediction based on Doppler radar–derived attributes. *J. Appl. Meteor.*, **35**, 617–626, doi:10.1175/1520-0450(1996)035<0617:ANNFTP>2.0.CO;2.

——, and ——, 1998: A neural network for damaging wind prediction. *Wea. Forecasting*, **13**, 151–163, doi:10.1175/1520-0434(1998)013<0151:ANNFDW>2.0.CO;2.

——, and A. Witt, 2001: A Bayesian neural network for severe-hail size prediction. *Wea. Forecasting*, **16**, 600–610, doi:10.1175/1520-0434(2001)016<0600:ABNNFS>2.0.CO;2.

McCandless, T. C., S. E. Haupt, and G. S. Young, 2015: A model tree approach to forecasting solar irradiance variability. *Sol. Energy*, **120**, 514–524, doi:10.1016/j.solener.2015.07.020.

——, ——, and ——, 2016a: A regime-dependent artificial neural network technique for short-range solar irradiance forecasting. *Renewable Energy*, **89**, 351–359, doi:10.1016/j.renene.2015.12.030.

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 09/24/21 06:09 PM UTC

——, G. S. Young, S. E. Haupt, and L. M. Hinkelman, 2016b: Regime-dependent short-range solar irradiance forecasting. *J. Appl. Meteor. Climatol.*, **55**, 1599–1613, doi:10.1175/JAMC-D-15-0354.1.

McGovern, A., D. J. Gagne II, J. K. Williams, R. A. Brown, and J. B. Basara, 2014: Enhancing understanding and improving prediction of severe weather through spatiotemporal relational learning. *Mach. Learn.*, **95**, 27–50, doi:10.1007/s10994-013-5343-x.

——, ——, J. Basara, T. M. Hamill, and D. Margolin, 2015: Solar energy prediction: An international contest to initiate interdisciplinary research on compelling meteorological problems. *Bull. Amer. Meteor. Soc.*, **96**, 1388–1395, doi:10.1175/BAMS-D-14-00006.1.

McNitt, J., J. Facundo, and J. O'Sullivan, 2008: Meteorological Assimilation Data Ingest System transition project risk reduction activity. *24th Conf. on Interactive Information Processing Systems*, New Orleans, LA, Amer. Meteor. Soc., 7C.1. [Available online at https://ams.confex.com/ams/88Annual/techprogram/paper_134617.htm.]

McPherson, R., and Coauthors, 2007: Statewide monitoring of the mesoscale environment: A technical update on the Oklahoma Mesonet. *J. Atmos. Oceanic Technol.*, **24**, 301–321, doi:10.1175/JTECH1976.1.

Mecikalski, J. R., J. K. Williams, C. P. Jewett, D. Ahijevych, A. LeRoy, and J. R. Walker, 2015: Probabilistic 0–1-h convective initiation nowcasts that combine geostationary satellite observations and numerical weather prediction model data. *J. Appl. Meteor. Climatol.*, **54**, 1039–1059, doi:10.1175/JAMC-D-14-0129.1.

Mehdi, T., N. Bashardoost, and M. Ahmadi, 2011: Kernel smoothing for ROC curve and estimation for thyroid stimulating hormone. *Int. J. Public Health Res. Spec. Issue*, 239–242.

Mellit, A., 2008: Artificial intelligence technique for modeling and forecasting of solar radiation data: A review. *Int. J. Artif. Intell. Soft Comput.*, **1**, 52–76, doi:10.1504/IJAISC.2008.021264.

Metz, C., 1978: Basic principles of ROC analysis. *Semin. Nucl. Med.*, **8**, 283–298, doi:10.1016/S0001-2998(78)80014-2.

Milkman, K., D. Chugh, and M. H. Bazerman, 2008: How can decision making be improved? Harvard Business School Tech. Rep. 08-102, 13 pp. [Available online at www.hbs.edu/faculty/Publication%20Files/08-102.pdf.]

Miller, M. L., V. Lakshmanan, and T. M. Smith, 2013: An automated method for depicting mesocyclone paths and intensities. *Wea. Forecasting*, **28**, 570–585, doi:10.1175/WAF-D-12-00065.1.

Muller, M., G. Tomlinson, T. Marrie, P. Tang, A. McGeer, D. Low, A. Detsky, and W. Gold, 2005: Can routine laboratory tests discriminate between severe acute respiratory syndrome and other causes of community-acquired pneumonia? *Clin. Infect. Dis.*, **40**, 1079–1086, doi:10.1086/428577.

Myers, W., G. Wiener, S. Linden, and S. E. Haupt, 2011: A consensus forecasting approach for improved turbine hub height wind speed predictions. *Proc. of Wind Power*, Anaheim, CA, American Wind Energy Association, 136.

National Academies of Sciences, Engineering, and Medicine, 2016: *Next Generation Earth System Prediction: Strategies for Subseasonal to Seasonal Forecasts*. National Academies Press, 350 pp., doi:10.17226/21873.

National Climatic Data Center, 2006: Data documentation for data set 6406 (DSI-6406): ASOS surface 1-minute, page 2 data. National Climatic Data Center Tech. Rep., 5 pp. [Available online at ftp://ftp.ncdc.noaa.gov/pub/data/asos-onemin/td6406.txt.]

National Oceanic and Atmospheric Administration/National Centers for Environmental Information, 2016: Billion-dollar weather and climate disasters: Table of events. National Oceanic and Atmospheric Administration/National Centers for Environmental Information. [Available online at www.ncdc.noaa.gov/billions/events.]

Newman, J., V. Lakshmanan, P. L. Heinselman, M. B. Richman, and T. M. Smith, 2013: Range-correcting azimuthal shear in Doppler radar data. *Wea. Forecasting*, **28**, 194–211, doi:10.1175/WAF-D-11-00154.1.

Niculescu-Mizil, A., and R. Caruana, 2005: Obtaining calibrated probabilities from boosting. *Proc. 21st Conf. on Uncertainty in Artificial Intelligence*, Edinburgh, Scotland, Association for Uncertainty in Artificial Intelligence, 413–420.

Novak, D. R., C. Bailey, K. F. Brill, P. Burke, W. A. Hogsett, R. Rausch, and M. Schichtel, 2014: Precipitation and temperature forecast performance at the Weather Prediction Center. *Wea. Forecasting*, **29**, 489–504, doi:10.1175/WAF-D-13-00066.1.

Ortega, K., T. Smith, J. Zhang, C. Langston, Y. Qi, S. Stevens, and J. Tate, 2012: The Multi-Year Reanalysis of Remotely Sensed Storms (MYRORSS) project. *26th Conf. on Severe Local Storms*, Nashville, TN, Amer. Meteor. Soc., 205. [Available online at https://ams.confex.com/ams/37RADAR/webprogram/Paper275486.html.]

Orwig, K., and Coauthors, 2014: Recent trends in variable generation forecasting and its value to the power system. *IEEE Trans. Renewable Energy*, **6**, 924–933, doi:10.1109/TSTE.2014.2366118.

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 09/24/21 06:09 PM UTC

Pagano, T. C., F. Pappenberger, A. W. Wood, M.-H. Ramos, A. Persson, and B. Anderson, 2016: Automation and human expertise in operational river forecasting. *Wiley Interdiscip. Rev.: Water*, **3**, 692–705, doi:10.1002/wat2.1163.

Parks, K., Y.-H. Wan, G. Wiener, and Y. Liu, 2011: Wind energy forecasting: A collaboration of the National Center for Atmospheric Research (NCAR) and Xcel Energy. National Renewable Energy Laboratory Tech. Rep., 35 pp. [Available online at www.nrel.gov/docs/fy12osti/52233.pdf.]

Pliske, R. M., B. Crandall, and G. Klein, 2004: Competence in weather forecasting. *Psychological Investigations of Competence in Decision Making*, K. Smith, J. Shanteau, and P. Johnson, Eds., Cambridge University Press, 40–70.

Provost, F., and P. Domingos, 2000: Well-trained PETs: Improving probability estimation trees. University of Washington; Stern School of Business Center for Digital Economy Research Working Paper 00-04-IS, 26 pp. [Available online at http://pages.stern.nyu.edu/~fprovost/Papers/pet-wp.pdf.]

Quinlan, J. R., 1986: Induction of decision trees. *Mach. Learn.*, **1**, 81–106, doi:10.1007/BF00116251.

——, 1993: *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 302 pp.

——, 1996: Improved use of continuous attributes in C4.5. *J. Artif. Intell. Res.*, **4**, 77–90, doi:10.1613/jair.279.

Reeves, H. D., K. L. Elmore, A. Ryzhkov, T. Schuur, and J. Krause, 2014: Sources of uncertainty in precipitation-type forecasting. *Wea. Forecasting*, **29**, 936–953, doi:10.1175/WAF-D-14-00007.1.

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

Schapire, R. E., 2003: The boosting approach to machine learning: An overview. *Nonlinear Estimation and Classification*, D. D. Denison et al., Eds., Springer, 149–171.

Schuur, T. J., H.-S. Park, A. V. Ryzhkov, and H. D. Reeves, 2012: Classification of precipitation types during transitional winter weather using the RUC model and polarimetric radar retrievals. *J. Appl. Meteor. Climatol.*, **51**, 763–779, doi:10.1175/JAMC-D-11-091.1.

Sharma, N., P. Sharma, D. Irwin, and P. Shenoy, 2011: Predicting solar generation from weather forecasts using machine learning. *Proc. Second IEEE Int. Conf. on Smart Grid Communications*, Brussels, Belgium, Institute of Electrical and Electronics Engineers, 32–37.

Sharman, R., 2016: Nature of aviation turbulence. *Aviation Turbulence: Processes, Detection, Prediction*, R. Sharman and T. Lane, Eds., Springer, 3–30, doi:10.1007/978-3-319-23630-8_1.

——, C. Tebaldi, G. Wiener, and J. Wolff, 2006: An integrated approach to mid- and upper-level turbulence forecasting. *Wea. Forecasting*, **21**, 268–287, doi:10.1175/WAF924.1.

Skitka, L. J., K. L. Mosier, and M. Burdick, 1999: Does automation bias decision-making? *Int. J. Hum. Comput. Stud.*, **51**, 991–1006, doi:10.1006/ijhc.1999.0252.

Smith, T. M., and Coauthors, 2016: Multi-Radar Multi-Sensor (MRMS) severe weather and aviation products: Initial operating capabilities. *Bull. Amer. Meteor. Soc.*, **97**, 1617–1630, doi:10.1175/BAMS-D-14-00173.1.

Sobash, R. A., C. S. Schwartz, G. S. Romine, K. Fossell, and M. L. Weisman, 2016: Severe weather prediction using storm surrogates from an ensemble forecasting system. *Wea. Forecasting*, **31**, 255–271, doi:10.1175/WAF-D-15-0138.1.

Storm Prediction Center, 2015: Severe weather event summaries. National Oceanic and Atmospheric Administration/National Weather Service Storm Prediction Center, accessed 14 June 2016. [Available online at www.spc.noaa.gov/climo/online.]

Tibshirani, R., 1996: Regression selection and shrinkage via the lasso. *J. Roy. Stat. Soc.*, **58B**, 267–288.

Trafalis, T. B., H. Ince, and M. B. Richman, 2003: Tornado detection with support vector machines. *Int. Conf. on Computational Science 2003*, Saint Petersburg, Russia, and Melbourne, Australia, International Conference on Computational Science, 289–298.

Tuohy, A., and Coauthors, 2015: Solar forecasting: Methods, challenges, and performance. *IEEE Power Energy Mag.*, **13**, 50–59, doi:10.1109/MPE.2015.2461351.

Wang, F., Z. Mi, S. Su, and H. Zhao, 2012: Short-term solar irradiance forecasting model based on artificial neural network using statistical feature parameters. *Energies*, **5**, 1355–1370, doi:10.3390/en5051355.

Weygandt, S., T. Smirnova, S. Benjamin, K. Brundage, S. Sahm, C. Alexander, and B. Schwartz, 2009: The High Resolution Rapid Refresh (HRRR): An hourly updated convection resolving model utilizing radar reflectivity assimilation from the RUC/RR. *23rd Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction*, Omaha, NE, Amer. Meteor. Soc., 15A.6. [Available online at https://ams.confex.com/ams/23WAF19NWP/techprogram/paper_154317.htm.]

Williams, J. K., 2009: Introduction to fuzzy logic. *Artificial Intelligence Methods in the Environmental Sciences*, S. E. Haupt, A. Pasini, and C. Marzban, Eds., Springer, 127–151.

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 09/24/21 06:09 PM UTC

——, 2014: Using random forests to diagnose aviation turbulence. *Mach. Learn.*, **95**, 51–70, doi:10.1007/s10994-013-5346-7.

——, and G. Meymaris, 2016: Remote turbulence detection using ground-based Doppler weather radar. *Aviation Turbulence: Processes, Detection, Prediction*, R. Sharman and T. Lane, Eds., Springer, 149–177, doi:10.1007/978-3-319-23630-8_7.

——, D. Ahijevych, S. Dettling, and M. Steiner, 2008a: Combining observations and model data for short-term storm forecasting. *Remote Sensing Applications for Aviation Weather Hazard Detection and Decision Support*, W. F. Feltz and J. J. Murray, Eds., International Society for Optical Engineering (SPIE Proceedings, Vol. 7088), 708805, doi:10.1117/12.795737.

——, R. Sharman, J. Craig, and G. Blackburn, 2008b: Remote detection and diagnosis of thunderstorm turbulence. *Remote Sensing Applications for Aviation Weather Hazard Detection and Decision Support*, W. F. Feltz and J. J. Murray, Eds., International Society for Optical Engineering (SPIE Proceedings, Vol. 7088), 708804, doi:10.1117/12.795570.

——, P. P. Neilley, J. P. Koval, and J. McDonald, 2016: Adaptable regression method for ensemble consensus forecasting. *30th AAAI Conf. on Artificial Intelligence*, Phoenix, AZ, Association for the Advancement of Artificial Intelligence, 3915–3921. [Available online at www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12492.]

Witt, A., M. Eilts, G. Stumph, J. Johnson, E. Mitchell, and K. Thomas, 1998: An enhanced hail detection algorithm for the WSR-88D. *Wea. Forecasting*, **13**, 286–303, doi:10.1175/1520-0434(1998)013<0286:AEHDAF>2.0.CO;2.

Yang, D., P. Jirutitijaroen, and W. M. Walsh, 2012: Hourly solar irradiance time series forecasting using cloud cover index. *Sol. Energy*, **86**, 3531–3543, doi:10.1016/j.solener.2012.07.029.

Zou, H., and T. Hastie, 2005: Regularization and variable selection via the elastic net. *J. Roy. Stat. Soc.*, **67B**, 301–320, doi:10.1111/j.1467-9868.2005.00503.x.

NEW FROM AMS BOOKS!

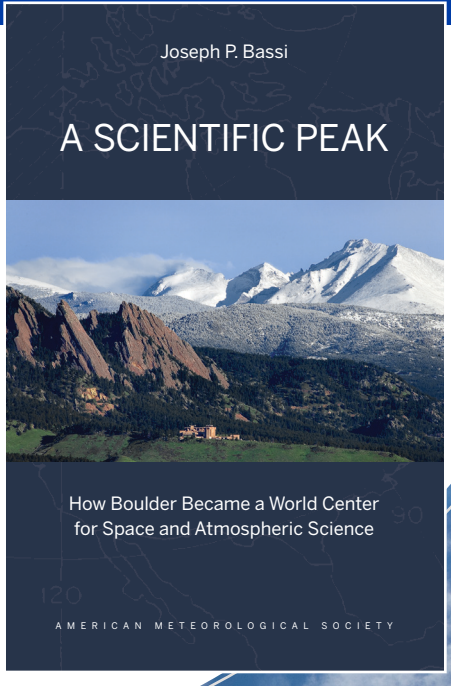
A Scientific Peak: How Boulder Became a World Center for Space and Atmospheric Science

Joseph P. Bassi

Once a Wild West city tucked between the Rocky Mountains and the Great Plains, Boulder is now home to some of the biggest names in science, including NCAR, NOAA, and NIST.

Why did big science come to Boulder? How did Boulder become the research mecca it is today?

A Scientific Peak is a fascinating history that introduces us to a wide variety of characters, such as Walter Orr Roberts, and the serendipitous brew of politics, passion, and sheer luck that, during the post-WWII and Cold War eras, transformed this "scientific Siberia" into one of America's smartest cities.



© 2015, 264 pages, paperback
print ISBN: 978-1-935704-85-0 eISBN: 978-1-940033-89-1
List price: $35 **AMS Member price:** $25

AMS BOOKS
RESEARCH APPLICATIONS HISTORY
bookstore.ametsoc.org

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 09/24/21 06:09 PM UTC