Article

Influence of the Characteristics of Weather Information in a Thunderstorm-Related Power Outage Prediction System

Peter L. Watson *, Marika Koukoula and Emmanouil Anagnostou

Department of Civil & Environmental Engineering, University of Connecticut, Storrs, CT 06269, USA; MARIKA.KOUKOULA@uconn.edu (M.K.); emmanouil.anagnostou@uconn.edu (E.A.)
* Correspondence: peter.watson@uconn.edu

Abstract: Thunderstorms are one of the most damaging weather phenomena in the United States, but they are also one of the least predictable. This unpredictable nature can make it especially challenging for emergency responders, infrastructure managers, and power utilities to be able to prepare and react to these types of events when they occur. Predictive analytical methods could be used to help power utilities adapt to these types of storms, but there are uncertainties inherent in the predictability of convective storms that pose a challenge to the accurate prediction of storm-related outages. Describing the strength and localized effects of thunderstorms remains a major technical challenge for meteorologists and weather modelers, and any predictive system for storm impacts will be limited by the quality of the data used to create it. We investigate how the quality of thunderstorm simulations affects power outage models by conducting a comparative analysis, using two different numerical weather prediction systems with different levels of data assimilation. We find that limitations in the weather simulations propagate into the outage model in specific and quantifiable ways, which has implications on how convective storms should be represented to these types of data-driven impact models in the future.

Keywords: power outages; machine learning; thunderstorms; numerical weather prediction

1. Introduction

Weather-related power outages, and the severe weather events that cause them, pose a persistent threat to the functioning of the infrastructure and economy of the United States. These types of power outages affect millions of people and cost the U.S. economy tens of billions of dollars every year; moreover, the rate at which they occur appears to be increasing [1]. Anticipating the damages that storms can cause is a critical step in electrical utility managers' storm outage management process. They need reliable information before a storm to be able to stage repair crews and effectively prepare for the damages that the storm will cause [2]. As such, there has been a recent surge in research and development activity into methods to predict storm damages and weather-related power outages.

Arguably, the most destructive types of storms in the United States are thunderstorms, including the associated convective phenomena (tornadoes, microbursts, hail, etc). While hurricanes often receive special attention because they are larger and more dramatic, thunderstorms are more common and cause more damage to the electrical infrastructure every year than any other type of weather. Indeed, investigations of major outage events reported to the Department of Energy have found that convective storms are responsible for the majority of weather-related outage events, the greatest number of customer outages, and the most outage hours [3,4]. Additionally there is every indication that the severity of thunderstorms is going to increase in the future. Changes in the climatic patterns of thunderstorms can already be seen in a time series analysis [5], and long-term climate projections suggest that, because of climate change, thunderstorms are likely to become stronger, more frequent, and more damaging [6,7].

Despite the demonstrated risk that thunderstorms present to the electrical infrastructure, they have not received much attention in the recent research for modeling weather-related power outages. While there are some outage modeling approaches that are generalized to a range of types of weather [8–11], much of research in this field has been focused on other types of storms. The vast majority of the work has focused on tropical storms and hurricanes, which can have particularly dramatic impacts [12–16], but several mature modeling approaches, specifically for extratropical storms [17–19], have also been developed.

In the existing general outage models, thunderstorms are sometimes included in the analysis [9,10,20,21], but the weather characteristics of these storms are treated in a similar fashion to other, more structured types of weather. There are also some studies that infer a focus on thunderstorms by including information about lightning strikes [11,22,23], but do not have an explicit focus on thunderstorms because they also include other types of weather events in their analysis.

This lack of focus on thunderstorms may be a result of the technical difficulty associated with describing and simulating them. Convective storms are particularly challenging for established numerical weather prediction (NWP) models and meteorological forecasts. While the increased horizontal resolution of convective-allowing configurations can lead to improved simulations, even with state-of-the-art high-resolution NWP models, reliable deterministic forecasts of thunderstorms longer than several hours are elusive [24–26]. As Yano et al. describe, there may be limitations to modern NWP models' ability to simulate convective storms because of the wide-spread use of assumptions and parameterizations that are reasonable for synoptic-scale weather patterns but are much less applicable to more complex convective phenomena [26]. These potential limitations of NWP simulations are long standing, and multiple strategies for mitigating them have emerged. Assimilating radar or even lightning observations into initial conditions of simulations can be used to improve short-term predictions [27,28]; forecasting systems that leverage this type of data assimilation for rapidly-updating nowcasts are currently operational [29]. In addition, for forecasts longer than several hours, stochastic predictions from convective-allowing ensembles have shown an improved forecasting skill by being able to capture the range of potential outcomes, instead of one deterministic scenario [30–32].

Similar approaches and findings can be seen in the few studies in the literature that specifically focus on predicting thunderstorm-related power outages. In Alpay et al., the authors take a rapid-refresh nowcasting approach to modeling thunderstorm-related outages, using an LSTM neural network trained on data from a rapidly updating radar-ingesting weather model from NOAA [33]. The works of Shield and Kabir et al. both describe a thunderstorm outage prediction system trained on weather data from the National Digital Forecast Database for an area in Alabama [34,35]. Shields investigates the limitations of the model he develops and identifies that it has better skill at a synoptic scale, which illustrates the difficulty of forecasting with thunderstorms [34]. Kabir et al. take a more stochastic approach and develop a quantile regression model, which allows the communication of the significant uncertainties associated with predicting the impacts of thunderstorms [35].

While this previous work attempts to manage the known limitations of weather simulations of thunderstorms, how these limitations propagate from weather simulations into machine-learning based impact models remains poorly described. The problem of poor inputs for a computational algorithm has been recognized since the dawn of computation [36], but the effects in this context are not fully understood. In this paper, we attempt to shed light on this matter by analyzing the quality of the weather data from two different weather simulation systems with differing amounts of data assimilation, determine how outage models trained on these different sets of weather data differ in skill and accuracy, and what information the outage models learn from. This knowledge is critical to build an understanding of the limitations of the data used to build impact models

for thunderstorms and to suggest how improved representations of weather will improve the quality of the insights that can be derived from them.

2. Materials and Methods

This study involved the creation and comparison of two separate machine-learning models designed to predict thunderstorm-related power outages, using data from NWP-based weather simulations and a wide range of other data sources in a region covering three states: Connecticut, Massachusetts, and New Hampshire, and five distinct electrical utility service territories: Eversource Connecticut (CT), Eversource Western Massachusetts (WMA), Eversource Eastern Massachusetts (EMA), Eversource New Hampshire (NH), and AVANGRID United Illuminating (UI). For geographical details of the modeling domain, refer to Figure 1.
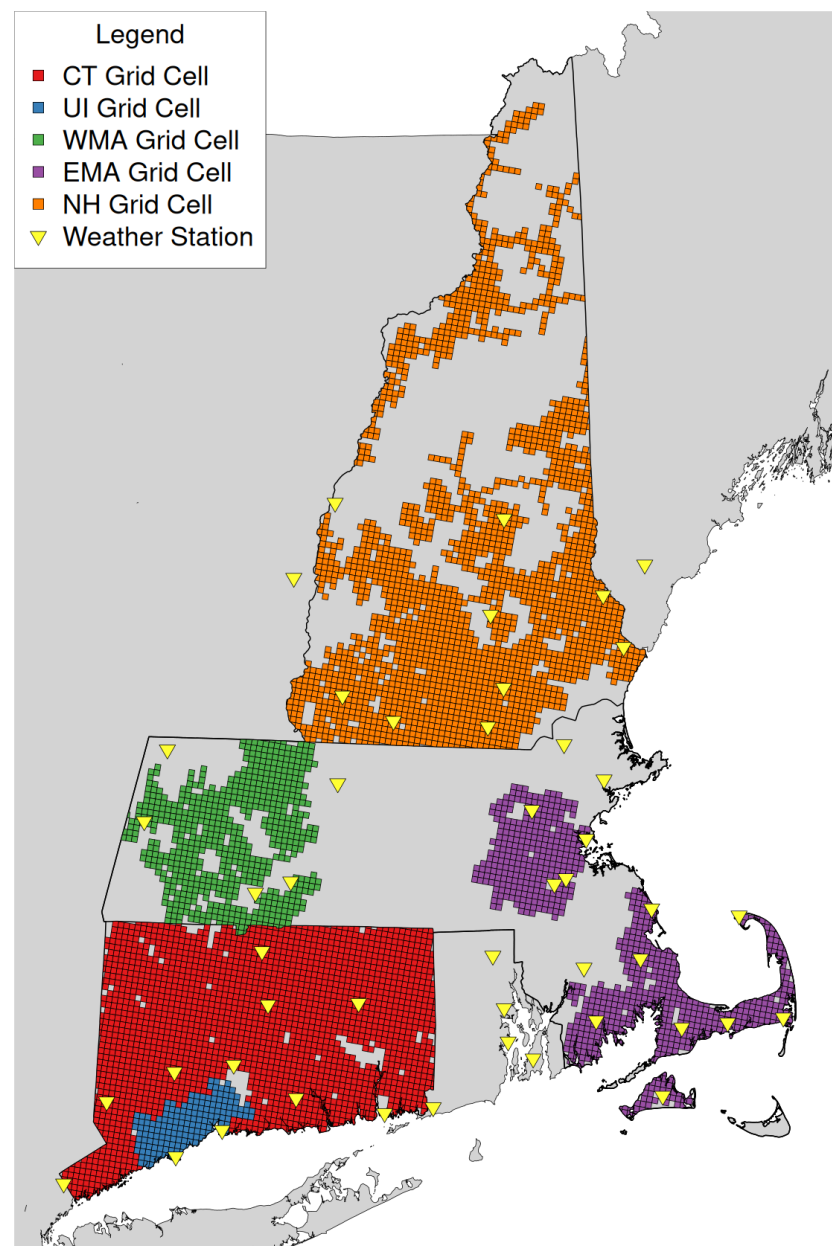

Figure 1. The location of the outage model grid cells by service territory as well as the location of the airport weather stations used in the meteorological analysis.

2.1. Data

The outage models developed in this analysis use data describing 372 thunderstorm events that occurred in the utility service territories from 2016 to 2020, as well as a range of environmental characteristics, such as vegetation and drought status, as well as proprietary outage and infrastructure data provided by the power utilities, aggregated to the grid cells of the weather simulations. We included as many thunderstorm events that could be observed in weather station reports from each utility service territory, and aggregated the data to the RTMA grid cells of each service territory for each thunderstorm event. For details about the amount of data used from each territory, see Table 1.

Table 1. The amount of data available for training the thunderstorm-related outage models.

	CT	WMA	EMA	NH	UI	Total
Number of Storms	74	82	69	91	56	372
Territory Grid Cells	2019	638	820	2128	169	5774
Total Entries	149,406	52,316	56,580	193,648	9464	461,414

2.1.1. Weather

The core of the analysis centers around datasets produced by two separate NWP gridded weather simulation systems: a hybrid NOAA analysis system, and a WRF 2 km simulation system. The NOAA analysis dataset is a combination of data from the Real-Time Mesoscale Analysis (RTMA) [37] and Stage IV Quantitative Precipitation Estimates (Stage IV) [38]. RTMA is a weather analysis product that produces a gridded estimate of weather conditions by statistically downscaling a 1 h short-term forecast and adjusting it with weather station observations. It produces a high-resolution, near real-time estimate of temperature, humidity, dew point, wind speed and direction, wind gusts, and surface pressure for the entire United States. The RTMA data were sourced from the archive hosted on the Google Earth Engine [39]. Stage IV is a Quantitative Precipitation Estimate (QPE) dataset created by the National Weather Service and the National Centers for Environmental Prediction (NWS, NCEP), using a blend of NEXRAD radar and the NWS River Forecast Center precipitation processing system [40]. It takes gridded precipitation estimates derived from radar scans, adjusts the values based on rain gauge data, and aggregates the data to produce gridded hourly estimates of precipitation for the continental United States. It is popular for analytical purposes and is often used as a reference to analyze the accuracy of satellite and other precipitation estimates [38]. By using a blend of RTMA and Stage IV, we are able to have a reasonable estimate of the average hourly weather conditions in each grid cell during each storm used in this analysis. For the sake of brevity, this dataset will sometimes be referred to as the "RTMA" system.

We compare this hybrid NOAA analysis dataset with another weather dataset developed from a configuration of the Weather Research and Forecasting Model (WRF), similar to one that was used in several outage predictions models [17,18], but with an increased horizon resolution to potentially help resolve convection. This model is initialized with the North American Mesoscale Forecast System analysis [41], which has 2 km horizontal grid spacing with one 6 km external domain. For configuration details, please see Table 2. These WRF simulations use a different projection than the RTMA system, so the results were resampled with bilinear interpolation to match the spatial characteristics of the RTMA analysis product.

Table 2. Details of the WRF simulation configuration.

Horizontal Resolution	2 km
Vertical Levels	51
Horizontal Grid Scheme	Arakawa C Grid
Nesting	One 6km Nested Domain
Microphysics Option	Thompson Graupel Scheme [42]
Longwave Radiation Option	RRTM Scheme [43]
Shortwave Radiation Option	Goddard Shortwave Scheme [44]
Surface-Layer Option	Revised MM5 Scheme [45]
Land-Surface Option	Noah Land-Surface Model [46]
Planetary Boundary Layer	Yonsei Scheme [47]

For outage modeling purposes, 24 h time series of a common set of weather variables generated from both weather simulation systems were processed to generate descriptive data features for each thunderstorm in this analysis. The weather variables considered are dew point temperature, specific humidity, air temperature, surface pressure, wind speed, wind gust speed, wind direction, and hourly precipitation rate. Established weather parameters that directly describe convective potential, such as CAPE and CIN, were unfortunately not available for this study because they are not published in RTMA, which is primarily a surface analysis product. For each of the included variables, the mean, max, minimum, standard deviation, 4 h mean during peak winds, and total were calculated for each storm, except for wind direction for which we took the median value. The median was taken to limit its sensitivity to outliers. Several additional features were calculated: the number of hours of winds above various wind speeds, calculated using various thresholds applied to wind speeds and gusts; typical wind direction by taking the mean of the median wind direction of included storms; and the difference between the typical wind direction and the median wind direction for that storm. To preserve its characteristics, all computation and analysis of wind direction was performed via the `circular` library in R [48]. Additionally, we included an additional set of features describing the time series of wind stress exerted on the trees by taking the product of the leaf area index (see below) and the square of the wind speed. For details, please see Appendix A, which contains a detailed table of all data features used for modeling.

2.1.2. Infrastructure and Outage Data

Proprietary data of the infrastructure and historical outages are made available for this study for the five utility service territories. Using `rgdal` and `rgeos` [49,50] for the area within each outage model grid cell, we calculated the length of overhead power lines, the number of utility poles, the number of fuses and cutouts, and the number of circuit reclosers.

The historic outage data describes the time and location of where damage occurred to the power distribution grid for a period of five years (2016 to 2020). Based on this information, we were able to calculate the number of damage locations within each outage model grid cell associated with each storm. A damage location is a physical location where repair crews are dispatched to repair damage after a storm. In the vast majority of cases, this meant counting the damage locations that were identified in the 24 h storm period, but in several cases, additional "nested" storm-related outages were recognized by utility operators after the storm period, so a longer window was sometimes used. These damage data were extracted from the utility outage management system, which is a software tool used by most large utilities to identify outages and dispatch repair crews.

2.1.3. Environmental Data

Because weather-related power outages are the result of interactions between the weather, the infrastructure and the environment, a range of environmental information was considered for this analysis. We processed the environmental data in several different ways depending on spatial resolution. When working with datasets with a resolution higher than the 2.5 km RTMA grid, for each grid cell, the raster data were sampled from a 60 m buffer around the overhead lines in that cell, and we calculated the representative percentage for the categorical data, or the average and standard deviation for the numerical data. We applied this process to a range of datasets, including the following: categorical land cover from the 2016 National Land Cover Database (NLCD) [51], 2016 NLCD Tree Canopy Coverage [52], vegetation height estimates from the Global Ecosystem Dynamics Investigation (GEDI) lidar instrument on the International Space Station [53], USGS 3DEP DEM elevation [54], and several other datasets, which required special processing. For example, we sampled the soils dataset developed by Watson et al. [18] from the USDA SSURGO database [55] to describe the soil characteristics (density, porosity, hydraulic conductivity, composition, and saturation). Additionally, because that previous work suggests that systemic biases caused by differences in the elevations of the weather predictions and the infrastructure may be present, we used the difference between those two elevations as an additional feature, `elvDiff`.

As seen in other outage modeling work [15,16], high-resolution data from the Individual Tree Species Parameter Maps (developed to support the USDA National Insect and Disease Risk Map) were used to calculate information about the density of the forest and the presence of various tree species [56]. However, because these data contain information about 264 individual tree species, we aggregated the basal area and stand density index of the species data by wood type (hardwood or softwood). Additionally, we were able to calculate the mean and standard deviation of the basal area (BA), stand density index (SDI), quadratic mean diameter (DQ), total frequency (TF), and trees per acre (TPA) for all trees, and generate statistics for the area around the infrastructure as described in the previous paragraph.

Data at the courser resolutions were handled more simply by sampling the data using the centroid of the grid cell. This included data describing the climatological leaf area index generated by Cerrai et al. [9], and a collection of drought indices published by the West Wide Drought Tracker [57]. While drought data was used in outage modeling in the past [12,15], we included more information, including the 1, 3 and 12 month Standardized Precipitation Index (SPI) of the month of the storm, as well as 12 month SPI from 1 to 5 years before the storm occurred. This information was included to capture not only the immediate drought conditions, but also any lingering effects of long-term drought stress on the vegetation.

2.2. Outage Modeling

To generate a robust outage prediction system based on the 131 data features, generated via the processes described in the previous section, additional steps were taken to confirm each variable's importance for the modeling outage, tune the model's hyperparameters, and test the system's performance via cross-validation. All modeling processes were coded in R [58], using a range of support libraries.

Variable importance for modeling was initially confirmed via a Boruta variable selection process. This process involves calculating the variable importance in a random forest model, and comparing each variable's importance against the importance of a randomized variable with the same distribution of values. Over many iterations, this process can confirm the importance of each variable in a dataset in comparison to random noise [59]. This was implemented via the `Boruta` R library [60].

Based on experience and the previous literature [9,10,18], we chose the Bayesian Additive Regression Tree (BART) model for this analysis [61], implemented via the `BART` R library [62]. While this is a quantile regression algorithm, we simplified outputs to

deterministic predictions for each storm by taking the mean of the outputs of the model. The hyperparameters used by the BART algorithm (sparse parameters `a` and `b`, shrinkage parameter `k`, the number of trees, the number of posterior draws, and the number of iterations used to initialize the Monte-Carlo Markov Chains) were tuned for this dataset via differential evolution [63] implemented via the `DEoptim` library [64]. It was used to find the optimal configuration of the BART algorithm based on the mean root mean square logarithmic error (RMSLE) of a fixed 5-fold cross-validation of the RTMA system dataset. To maintain comparability, these optimized hyperparameter values were consistently applied to all models and experiments in this analysis. RMSLE was chosen because it is less sensitive to extreme errors.

2.3. Analysis

To understand the differences between the hybrid NOAA analysis dataset, the WRF simulation dataset, and the outage prediction models built on them, we evaluated each weather simulation's ability to represent the local weather conditions by comparing its predictions against weather station observations. Then, to understand the different qualities of the two outage models, as well as evaluate the importance of individual and groups of variables in the outage models, we compared the cross-validation results, using traditional and spatial error metrics.

More specifically, to evaluate the two-gridded weather simulations, data were collected from METAR and SPECI reports via the Integrated Surface Data archive maintained by the National Centers for Environmental Information [65]. Any data flagged with quality issues were removed, and all observations reported were averaged for every hour to produce a 24 h time series. Any station or variable with more than two hours of missing data were removed from the analysis. Then, the same summary statistics used to generate the outage model features (mean, minimum, maximum, standard deviation, total, 4 h mean during peak winds) were calculated based on the weather station observations. Any mean or maximum gust values reported as zero by the weather stations were also removed from consideration.

For this analysis, all weather stations in the proximity of the outage prediction service territories were considered, with the exception of Northern New Hampshire. We removed that area from consideration because it is dominated by the White Mountains, and the complex topography would cause biased results. See Figure 1 for the detailed weather station location information used in this analysis. While additional data cleaning steps are common when this process is used for weather model evaluation, we determined that this would not be appropriate because the localized differences between the weather station observations and gridded NWP data are of interest.

The outage model performance was evaluated using leave-one-date-out cross-validation. This validation process simulates the operational predictability of the outages caused by each weather event by iteratively isolating the information of each storm event, and testing the model's ability to predict it. More specifically, for each storm date and time present in the database of storms, we reserved the data from that date and time, trained the outage model on the remaining data, and tested that trained model on the reserved data. This way, we had a comprehensive evaluation of all storms in our database, but prevented any spatial or temporal correlations in the weather data from influencing the model performance. While 372 thunderstorm events were considered in this analysis, because of overlapping times, each outage model was only trained and tested 226 times for this cross-validation. To evaluate the overall cross-validation results, we calculated the median absolute percent error (MdAPE), mean absolute percent error (MAPE), centered root mean squared error (CRMSE), correlation coefficient (R^2), and the Nash–Suttcliffe efficiency (NSE) [66]. For definitions of these error metrics, please see Appendix C.

Because the spatial predictability of thunderstorm outages is also of interest, we also applied the fraction skill score (FSS) to evaluate the spatial skill of the outage models. FSS uses a threshold, or a series of thresholds, to generate binary rasters of predictions and

actual values, and compares the two within a series of neighborhoods [67]. A skillful model is able to predict a similar fraction of values above the threshold as the actual in a small area. This metric is becoming a widely accepted method to evaluate the spatial skill of precipitation forecasts, especially in the U.S. [68]. Under ideal conditions, an FSS value greater than 0.5 indicates a "useful" skill, but depending on the conditions of the baseline performance ($FSS_{uniform}$), it is subject to change as defined by the following equation:

$$FSS_{uniform} = 0.5 + FSS_{random}/2 \tag{1}$$

where FSS_{random} is the total of the derived binary raster, divided by the number of cells in the domain [67]. For precipitation, the threshold tends to increase with smaller domains and as the prevalence of precipitation increases [69]. For this analysis, we calculated the FSS for each storm by service territory for a range of scales (3×3 to 21×21 cells), and outage thresholds between upscaled outage predictions and actual outages via the `validation` library [70]. Upscaling the predicted and actual values for the FSS calculation was important because the resolution of our model and the frequency of actual damages is such that the actual values are extremely zero-inflated and very sparse (96.3% zeros, and mean of 0.048 damages per grid cell). The outage model predictions however, tend to be small (median of 0.0292 and 0.0314 for RTMA and WRF systems respectively) and are more evenly distributed. This difference in spatial distribution was minimized by applying boxcar smoothing to a small 3×3 neighborhood on both the actual and predicted outages for each event and territory via the `SpatialVx` library [71]. While this process effectively degrades the precision of the analysis, it generates more continuously distributed values that are more comparable, while not affecting the total number of damage locations for each event.

To measure the variable importance of each outage model, we applied the variable permutation technique described by Fisher et al. [72] via the `DALEX` library in R [73]. This technique is model agnostic and uses a loss function to measure model performance as the input variables are perturbed. This allows for a quantitative understanding of each variable's influence on the model performance. Doing this evaluation via cross-validation would be prohibitively complex and computationally expensive, so to evaluate the variable importance within the outage models, all available data were used to train the models before variable importance was measured. In addition, because there is a significant random component in this analysis, we calculated this variable importance over ten iterations for both outage models, and calculated the confidence intervals. The loss metric used to evaluate variable importance, root mean squared logarithmic error (RMSLE), was chosen because it is robust to the inclusion of zeros and is less sensitive to rare cases of extreme errors, which can be present because of the statistical distribution of actual outages as described above. However, because it is a logarithmic error metric, differences in RMSLE can often appear small, despite being significant.

3. Results

3.1. Weather Analysis

As demonstrated in Figure 2, the NOAA analysis dataset represents almost all weather parameters used in the outage models more accurately than the WRF simulation dataset. Very significant differences are seen between the quality of the precipitation parameters, as well as several wind and gust features. Both systems are able to represent parameters associated with synoptic scale processes, such as temperature, humidity, and surface pressure dynamics, much more accurately than mesoscale and microscale processes, such as wind and precipitation. Some surface pressure parameters appear to be poorly captured, but this is likely due to differences in elevation between the NWP data and weather station data, which are not accounted for in this evaluation. In general, these results are quite consistent with what we would expect from the state of the art of NWP of a deterministic 24 h simulation of thunderstorms. For detailed metrics, see Appendix B.
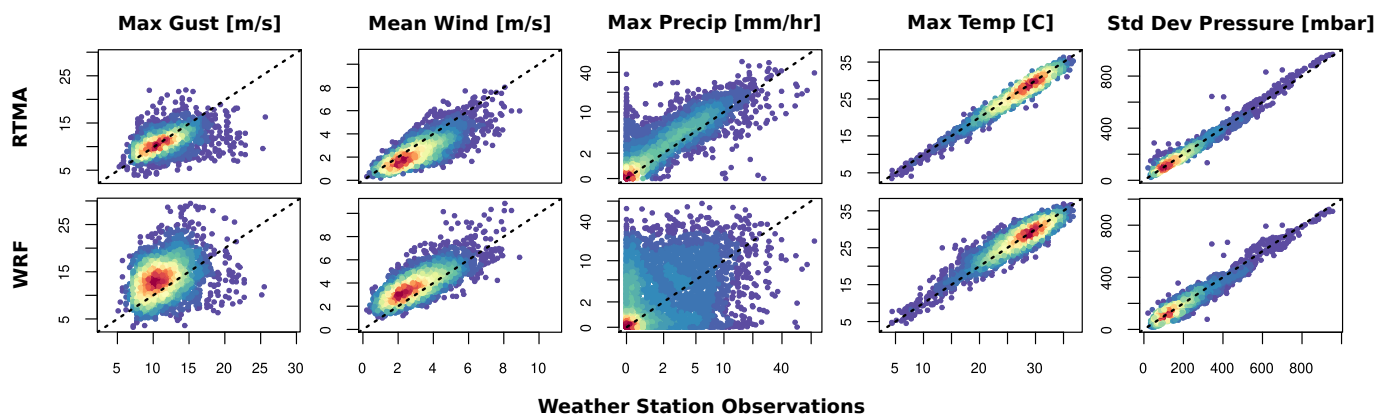
Figure 2. Point-to-point comparison of the NOAA analysis parameters (RTMA, (**Top**)) and the WRF simulation parameters (WRF, (**Bottom**)) versus weather station observations for select variables, describing 24 h thunderstorm events.

3.2. The Outage Models

The RTMA-based outage model performs slightly better than the WRF-based model based on all metrics used in our analysis as seen in Table 3 and Figure 3.

Table 3. Error metrics of the event-level performance of the cross-validation of the outage prediction systems.

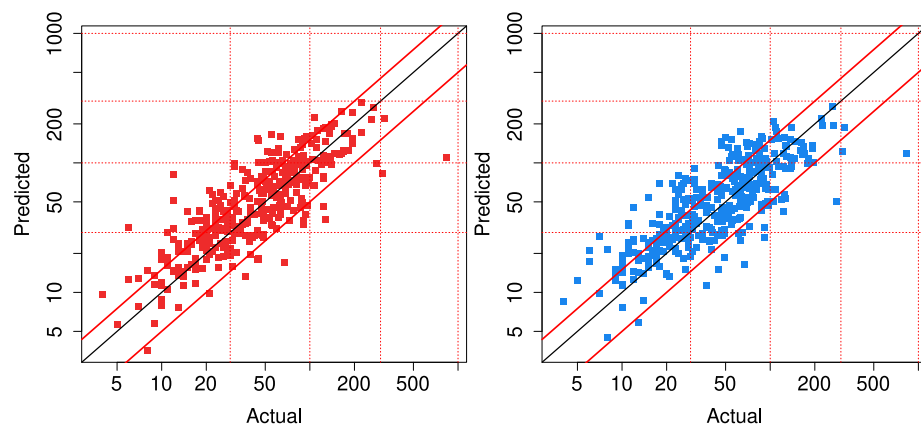|	MdAPE	MAPE	CRMSE	R^2	NSE
RTMA	31%	46%	50	0.39	0.37
WRF	35%	50%	51	0.36	0.35

Figure 3. Scatterplots of cross-validation predictions versus actual outages for all thunderstorm events for RTMA- (red, (**left**)) and WRF (blue, (**right**))-based outage prediction systems.

While a direct comparison is not particularly fair because of the differences in the events used in the analysis and the domains of the models, both outage models presented here perform reasonably well in comparison to other outage prediction models of a similar architecture. Wanik et al. [10] describe a warm weather outage model that has a slightly higher MdAPE (35.1 to 38.7%). In Cerrai et al. [9], the best overall outage model has an overall MdAPE of 43%, a MAPE of 59% and an NSE of 0.53. In Yang et al. [17], their conditional outage prediction system designed for severe events has a MdAPE of 38%, MAPE of 46%, and NSE of 0.79. In Watson et al. [18], their best performing rain/wind storm model has a MdAPE of 38%, MAPE of 57%, and an NSE of 43%. The thunderstorm outage models described here have competitive APE metrics, but have a comparatively low NSE, in part because of one under-predicted extreme event.

Overall, the cross-validation results indicate that the models presented here are sensitive to the overall severity of the different thunderstorms. The model has a good dynamic range, especially if one considers that the median daily outages for CT, WMA, EMA, NH, and UI are 35, 6, 20, 22, and 25, respectively. The models shown here demonstrate a dynamic range of around 10 times the typical daily outage level for each service territory, depending on storm severity.

3.2.1. Spatial Skill

As seen in Figure 4, the RTMA-based outage model has slightly better spatial performance than the WRF-based model, but the differences between the outage models are small in comparison to the differences between the events and territories. While many thresholds were evaluated, we show the results for a threshold of 0.111 damage locations, which correspond to having one damage location smoothed out in a 3 × 3 pixel area (approximately 7.5 km^2).
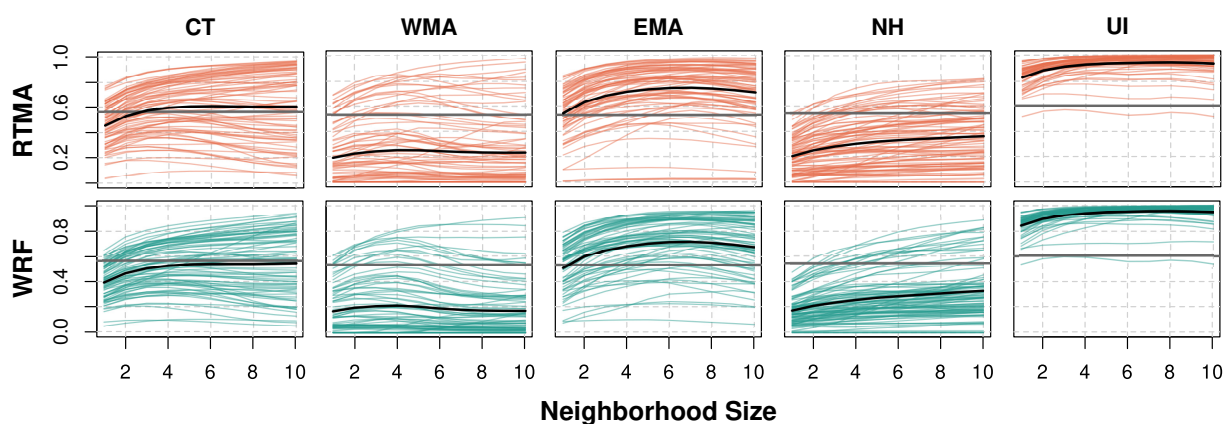

Figure 4. FSS for all events by territory for the RTMA- and WRF-based outage models with a moderate outage risk threshold (0.111 damage locations), plotted for neighborhood sizes 3 × 3 to 21 × 21 grid cells. The colored lines are FSS values for each event; the black line indicates the average FSS over all events; and the horizontal dark grey line indicates the average $\text{FSS}_{uniform}$.

3.2.2. Outage Model Variable Importance

The grouped variable importance analysis of the outage models in Figure 5 shows that, while infrastructure-related variables are the most important by far, there are differences between the two models as to which weather parameters contribute the most to the models. While the RTMA-based system finds precipitation information to be very useful, the WRF-based system has stronger preference for winds, temperature, and humidity than the RTMA model. The WRF model also appears to fit more on such environmental variables as land cover, vegetation, and elevation, which do not vary storm-by-storm in a given service territory. The results of an individual variable importance analysis is displayed in Appendix A. Although the importance of any one variable to the model is relatively small, given the large number of variables used, and the logarithmic error metric used to measure the dropout loss only makes the differences appear smaller, there are some interesting differences between the two models. Most notably, the maximum precipitation rate is one of the least important variables in the WRF model but is the second most important variable in the RTMA model.
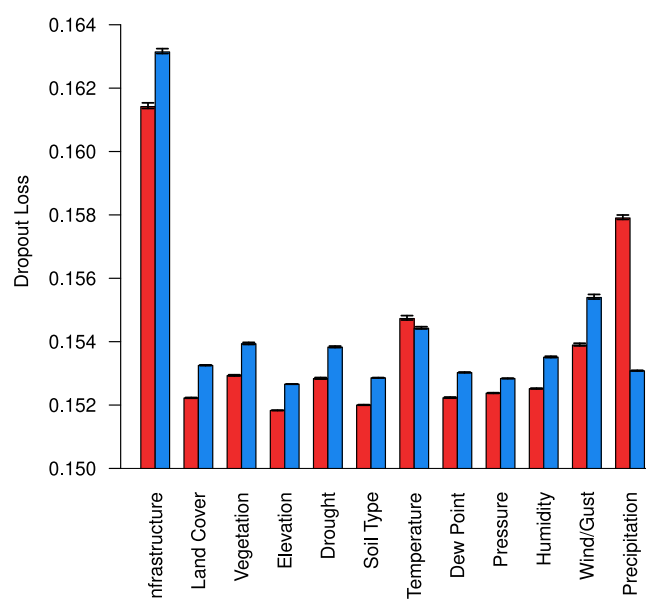
Figure 5. Grouped variable importance as measured by dropout loss (RMSLE) over 10 iterations of permuted groups of variables. The 95% confidence intervals are also shown for both the RTMA-based outage model (red, (**left**)), and the WRF-based outage model (blue, (**right**)).

4. Discussion

Based on these results, several conclusions can be made about the predictability of thunderstorm-related power outages. Firstly, while the NOAA analysis data represent local weather conditions more accurately than the WRF simulation, many weather features used in the outage prediction models have significant errors in both systems. Rather than these errors being simulation or forecasting errors, because of the amount of observational data assimilated into the NOAA analysis system, they are likely due to the representativeness error caused by depicting complex and locally variable phenomena as deterministic and uniform in the 2.5 km × 2.5 km area. This type of error has been documented in the literature for precipitation and winds [74–77], and the errors in the RTMA data for winds and the Stage IV are comparable to the magnitude of representativeness error found in these works.

Secondly, because the NOAA analysis data have higher quality weather data than the WRF simulations, it is unsurprising that the RTMA outage model is more accurate than the WRF-based one. However, what is surprising is how modest the performance differences between these outage models are. Even with the large amount of observational data incorporated into the RTMA and Stage IV analysis products, which have much fewer simulation errors present than the WRF simulations, the outage model is unable to predict thunderstorm-related outages with greater accuracy.

This suggests that the randomness of storm damages is quite significant, and more precise outage predictions may require significantly more precise information. One possibility is that additional factors that are not considered in this study, such as the age of the infrastructure, limit the outage model. However, there are also differences between the two models that suggest other possibilities. As described above, the spatial resolution of the representation of the weather data is a readily apparent source of imprecision in our data. Although all data used in these models, including the environmental and infrastructure information, may suffer from similar representativeness errors, we can see that some weather variables are better represented at 2.5 km × 2.5 km than others. How the precision of the weather data affects the outage models can be understood with a more detailed analysis of the variable importance.

By comparing the R^2 values of the weather feature evaluation and the importance of the weather variables in the outage models, we find that there is a weak but real correlation between the two (0.23 ± 0.07 for RTMA, 0.29 ± 0.07 for WRF). This indicates that the

outage models have a preference for precise and accurate weather information. This may be obvious, but this preference also appears regardless of whether or not the weather phenomena directly causes power outages. Both RTMA and WRF outage systems find temperature and humidity to be somewhat important to its predictions, although these variables are not direct causes of outages in thunderstorms. They are more indicators of convective potential and are, thus, indirectly related to power outages, but because of their accurate representation, the machine learning algorithms of the outage models find them useful for understanding the risk of weather-related damage.

At the same time, there is also a distinct preference for variables that have a more direct causative relationship with weather-related outages. This can best be seen in how the RTMA system has a strong preference for precipitation variables. `maxPREC` is the 2nd most important variable of all for that model, despite it having only a moderate correlation with local conditions (R^2 of 0.5298). It can also be seen in how both models find useful information in wind and gust variables, despite the most precisely predicted variable in that group, `avgWIND`, only having a moderate correlation with local conditions (R^2 of 0.6346 and 0.5879 for RTMA and WRF, respectively). This is because both wind and precipitation are good indicators of the location and intensity of a convective storm, and more direct indicators of the risk of weather-related damages. Indeed, in the case of the RTMA system, the strong preference for precipitation information comes with a comparatively weaker preference for most other variable groups.

This suggests that if the precision of the precipitation and wind information could be increased further, we can expect corresponding increases in the accuracy of outage prediction models for thunderstorms. Additionally, if we consider how the apparent lack of precision in these data is likely from representativeness error, as described above, future directions for research become apparent.

Lastly, the spatial skill of the outage prediction system appears to vary significantly from storm-to-storm as well as territory-by-territory. It is beyond the scope of this paper to speculate about the storm-to-storm variability in the FSS scores, which may also be a function of the accuracy and precision of the weather simulations, but the distinct differences in spatial predictability of outages in different service territories is suggestive of distinct differences between them. It has been documented for precipitation that the FSS calculations change significantly depending on the size of the domain. However, in the case of outages, this effect is likely only moderate because the average value of $\text{FSS}_{uniform}$ does not vary much between territories. The most apparent and potentially impactful difference for outage models between the territories is the densities of the infrastructure. As seen in Figure 5 and Appendix A, infrastructure is a very influential variable for outage modeling, and while all the service territories included in this study contain some urban areas, some are much more consistently urbanized than others. As such, the mean density of overhead lines for each territory varies widely with a minimum of 8.5km per grid cell in WMA, and a maximum 27.5 km per grid cell in UI. If the mean density of the overhead lines and mean FSS as shown in Figure 4 for each territory are compared, we see that the Pearson correlations between the two are 0.927 for RTMA, and 0.946 for WRF: a very strong correlation between the overall spatial predictability of outages and the density of the infrastructure in the region. This is a clear indication of the influence that the infrastructure density has over the spatial predictability of power outages. However, this also may be an indication of over-fitting on the infrastructure features. Infrastructure is by far the most important variable group in this analysis, but in the case of the RTMA outage model, better spatial skill comes with a corresponding lower importance of infrastructure.

5. Conclusions

While the two thunderstorm-related outage models shown here are acceptably skilled at predicting the total number of damages for each storm event, they have difficulty predicting the location of storm impacts. Both the models based on the NOAA analysis dataset and the WRF simulation dataset appear to fit strongly on the amount of infrastructure

present in an area and a combination of weather variables that are either directly related to storm damages but imprecisely represented (precipitation, winds), or are more general indicators of convective potential but more precisely represented (temperature, humidity).

Because predictions of the weather conditions and power outages appear to have similar limitations for thunderstorms, there are established analytical methods that could be readily applied to improve the modeling of power outages and other impacts associated with thunderstorms. Just as weather ensembles allow meteorologists to predict the potential intensity of thunderstorms beyond the capabilities of deterministic forecasts, an outage model coupled to a weather ensemble may allow us to predict the potential impacts in a similar way. Because of the high uncertainties, rapidly-refreshing outage models, such as that described in Alpay et al. [33], may be more useful in an operational decision-making context for thunderstorm preparedness.

If one considers how strong convective storms are an increasing threat, globally, there is an implicit call to accelerate investment in global weather prediction and the observation infrastructure. The impact models presented here, even with their limitations, are only possible because of the availability of high-resolution nowcasting products in the United States. While recent developments in global convective-allowing NWP systems are encouraging [78], for this type of impact modeling to be applied in other countries, more work in this space is needed.

Based on our findings, we can expect that as better representations of local weather conditions during thunderstorms are developed both in the United States and globally, outage model accuracy, overall as well as spatially, will improve; the outage models will learn more and more of the phenomena directly linked to weather-related power outages, such as strong winds and extreme precipitation, instead of the synoptic patterns that are correlated to them. To progress along that path, a more granular understanding of the weather conditions that cause damage in convective storms and how they can be represented is needed. Further research involving an analysis or modeling of storm impacts based on microscale numerical weather prediction, large eddy simulations, or even observations from radar or lidar instruments could be very informative about how weather information can be generated in a way that improves our ability to understand and anticipate the impacts of convective storms.

Abbreviations

The following abbreviations are used in this manuscript:

NWP	Numerical Weather Prediction
WRF	Weather Research and Forecasting model
RTMA	Real-Time Mesoscale Analysis

RMSLE	Root Mean Squared Logarithmic Error
NSE	Nash–Sutcliffe Efficiency
MAPE	Mean Absolute Percent Error
CRMSE	Centered Root Mean Squared Error
FSS	Fraction Skill Score
CT	Eversource Connecticut
WMA	Eversource Western Massachusetts
EMA	Eversource Eastern Massachusetts
NH	Eversource New Hampshire
UI	AVANGRID United Illuminating
SPI	Standardized Precipitation Index
LAI	Leaf Area Index
DEM	Digital Elevation Model
NLCD	National Land Cover Database
3DEP	3D Elevation Program
GEDI	Global Ecosystem Dynamics Investigation
SSURGO	Soil Survey Geographic Database
ITSP	Individual Tree Species Parameter
WWDT	West Wide Drought Tracker
MODIS	Moderate Resolution Imaging Spectroradiometer
METAR	Meteorological Aerodrome Reports
SPECI	Aviation Selected Special Weather Report
NOAA	National Oceanic and Atmospheric Administration
NCEP	National Centers for Environmental Prediction
USDA	United States Department of Agriculture
USGS	United States Geological Survey
MRLC	Multi-Resolution Land Characteristics

Appendix A. Data Features

Table A1. Description of variables used in outage prediction models. The dropout loss of the top ten variables for each model are in bold. Higher dropout loss indicates greater importance.

Name	Description	Source	Variable Group	RTMA Drp. Loss	WRF Drp. Loss
ohLength	Length of Overhead Line	Utility Company	Infrastructure	**0.153695**	**0.155182**
poleCount	Number of Utility Poles	Utility Company	Infrastructure	**0.152473**	**0.153222**
fuseCount	Number of Fuses	Utility Company	Infrastructure	0.152181	**0.153253**
reclrCount	Number of Reclosers	Utility Company	Infrastructure	**0.152233**	**0.153057**
prec11	Percent NLCD 11—Open Water	NLCD 2016 [51]	Land Cover	0.151933	0.152713
prec21	Percent NLCD 21—Developed, Open	NLCD 2016 [51]	Land Cover	0.152056	0.152876
prec22	Percent NLCD 22—Developed, Low	NLCD 2016 [51]	Land Cover	0.151910	0.152749
prec23	Percent NLCD 23—Developed, Medium	NLCD 2016 [51]	Land Cover	0.152079	**0.152963**
prec24	Percent NLCD 24—Developed, High	NLCD 2016 [51]	Land Cover	0.151989	0.152700
prec31	Percent NLCD 31—Barren	NLCD 2016 [51]	Land Cover	0.151927	0.152714
prec41	Percent NLCD 41—Deciduous Forest	NLCD 2016 [51]	Land Cover	0.151974	0.152783
prec42	Percent NLCD 42—Evergreen Forest	NLCD 2016 [51]	Land Cover	0.151936	0.152731
prec43	Percent NLCD 43—Mixed Forest	NLCD 2016 [51]	Land Cover	0.151861	0.152732
prec52	Percent NLCD 52—Shrub	NLCD 2016 [51]	Land Cover	0.151933	0.152704
prec71	Percent NLCD 71—Grassland	NLCD 2016 [51]	Land Cover	0.151928	0.152699
prec82	Percent NLCD 82—Cultivated Crops	NLCD 2016 [51]	Land Cover	0.151933	0.152715
prec95	Percent NLCD 95—Herbaceous Wetlands	NLCD 2016 [51]	Land Cover	0.151934	0.152713
avgCanopy	Mean Percent Tree Canopy Cover	NLCD Tree Canopy 2016 [52]	Vegetation	**0.152329**	**0.152956**
stdCanopy	Standard Deviation of Canopy Cover	NLCD Tree Canopy 2016 [52]	Vegetation	0.151968	0.152736
avgVegHgt	Mean Vegetation Height	GEDI 2019 [53]	Vegetation	0.152037	0.152906
stdVegHgt	Standard Deviation of Vegetation Height	GEDI 2019 [53]	Vegetation	0.151945	0.152771

Table A1. *Cont.*

Name	Description	Source	Variable Group	RTMA Drp. Loss	WRF Drp. Loss
avgHardBA	Mean Hardwood Basal Area	ITSP [56]	Vegetation	0.151903	0.152714
stdHardBA	Standard Deviation of Hardwood BA	ITSP [56]	Vegetation	0.151939	0.152722
avgHardSDI	Mean Hardwood Stand Density Index	ITSP [56]	Vegetation	0.151963	0.152686
stdHardSDI	Standard Deviation of Hardwood SDI	ITSP [56]	Vegetation	0.151919	0.152698
avgSoftBA	Mean Softwood Basal Area	ITSP [56]	Vegetation	0.151914	0.152702
stdSoftBA	Standard Deviation of Softwood BA	ITSP [56]	Vegetation	0.151881	0.152691
avgSoftSDI	Mean Softwood Stand Density Index	ITSP [56]	Vegetation	0.151927	0.152677
stdSoftSDI	Standard Deviation of Softwood SDI	ITSP [56]	Vegetation	0.151891	0.152684
avgBA	Mean Total Basal Area	ITSP [56]	Vegetation	0.151950	0.152737
stdBA	Standard Deviation of Total Basal Area	ITSP [56]	Vegetation	0.151910	0.152776
avgSDI	Mean Total Stand Density Index	ITSP [56]	Vegetation	0.151951	0.152698
stdSDI	Standard Deviation of Total SDI	ITSP [56]	Vegetation	0.151981	0.152767
avgDQ	Mean Total Quadratic Mean Diameter	ITSP [56]	Vegetation	0.151929	0.152718
stdDQ	Standard Deviation of Total DQ	ITSP [56]	Vegetation	0.151936	0.152760
avgTF	Mean of Total Frequency	ITSP [56]	Vegetation	**0.152247**	**0.153137**
stdTF	Standard Deviation of TF	ITSP [56]	Vegetation	0.151984	0.152783
avgTPA	Mean of Trees per Acre	ITSP [56]	Vegetation	0.151881	0.152653
stdTPA	Standard Deviation of TPA	ITSP [56]	Vegetation	0.151932	0.152690
LAI	Leaf Area Index	MODIS [9,79]	Vegetation	0.152083	0.152848
avgDEM	Mean Elevation	3DEP [54]	Elevation	0.151837	0.152660
stdDEM	Standard Deviation of Elevation	3DEP [54]	Elevation	0.151924	0.152720
elvDiff	Difference of avgDEM and weather elevation	3DEP [54], RTMA [37], WRF [80]	Elevation	0.151931	0.152706
spi1	One Month Standardized Precipitation Index	WWDT [57]	Drought	0.151987	**0.153005**
spi3	Three Month Standardized Precipitation Index	WWDT [57]	Drought	0.152001	0.152830
spi12_0	12 Month SPI, current	WWDT [57]	Drought	0.151998	0.152773
spi12_1	12 Month SPI, 1 year prior	WWDT [57]	Drought	0.152137	0.152853
spi12_2	12 Month SPI, 2 years prior	WWDT [57]	Drought	0.152075	0.152831
spi12_3	12 Month SPI, 3 years prior	WWDT [57]	Drought	0.152155	0.152853
spi12_4	12 Month SPI, 4 years prior	WWDT [57]	Drought	0.152027	0.152809
spi12_5	12 Month SPI, 5 years prior	WWDT [57]	Drought	0.151939	0.152801
hydNo	Percent not hydric soils	SSURGO [55]	Soil Type	0.151954	0.152771
siltTotal	Percent Silt Content	SSURGO [55]	Soil Type	0.151943	0.152747
clayTotal	Percent Clay Content	SSURGO [55]	Soil Type	0.151929	0.152717
rockTotal	Percent of Rock Content	SSURGO [55]	Soil Type	0.151966	0.152738
soilDepth	Depth of Soil	SSURGO [55]	Soil Type	0.151847	0.152661
orgMat	Percent of Organic Material	SSURGO [55]	Soil Type	0.151949	0.152743
soilDens	Soil Density	SSURGO [55]	Soil Type	0.151950	0.152730
kSat	Saturated Hydraulic Conductivity	SSURGO [55]	Soil Type	0.151945	0.152732
satP	Soil Porosity	SSURGO [55]	Soil Type	0.151961	0.152716
avgTMP	Mean Air Temperature	RTMA [37], WRF [80]	Temperature	0.152191	0.152650
stdTMP	Standard Deviation of Air Temp	RTMA [37], WRF [80]	Temperature	0.152156	0.152819
maxTMP	Maximum Air Temperature	RTMA [37], WRF [80]	Temperature	**0.152685**	**0.153235**
minTMP	Minimum Air Temperature	RTMA [37], WRF [80]	Temperature	0.151925	0.152873
sumTMP	Sum of Air Temperatures	RTMA [37], WRF [80]	Temperature	0.152029	0.152741
peakTMP	Mean Temp during peak winds	RTMA [37], WRF [80]	Temperature	0.152020	0.152780
avgDPT	Mean Dew Point Temperature	RTMA [37], WRF [80]	Dew Point	0.151976	0.152767
stdDPT	Standard Deviation of Dew Point	RTMA [37], WRF [80]	Dew Point	0.152013	0.152804
maxDPT	Maximum Dew Point Temperature	RTMA [37], WRF [80]	Dew Point	0.151926	0.152687
minDPT	Minimum Dew Point Temperature	RTMA [37], WRF [80]	Dew Point	0.151941	0.152832
sumDPT	Sum of Dew Point Temperatures	RTMA [37], WRF [80]	Dew Point	0.152012	0.152792
peakDPT	Mean Dew Point during peak winds	RTMA [37], WRF [80]	Dew Point	0.152007	0.152723
avgPRES	Mean Surface Pressure	RTMA [37], WRF [80]	Pressure	0.151914	0.152716
stdPRES	Standard Deviation of Pressure	RTMA [37], WRF [80]	Pressure	**0.152297**	0.152797
maxPRES	Maximum Surface Pressure	RTMA [37], WRF [80]	Pressure	0.151950	0.152735
minPRES	Minimum Surface Pressure	RTMA [37], WRF [80]	Pressure	0.151946	0.152737
sumPRES	Sum of Surface Pressures	RTMA [37], WRF [80]	Pressure	0.151943	0.152706
peakPRES	Mean Pressure during peak winds	RTMA [37], WRF [80]	Pressure	0.151960	0.152694
avgSPFH	Mean Specific Humidity	RTMA [37], WRF [80]	Humidity	0.152062	0.152817
stdSPFH	Standard Deviation of Spec. Humidity	RTMA [37], WRF [80]	Humidity	0.152018	0.152836
maxSPFH	Maximum Specific Humidity	RTMA [37], WRF [80]	Humidity	0.151949	0.152751
minSPFH	Minimum Specific Humidity	RTMA [37], WRF [80]	Humidity	0.152082	0.152905
sumSPFH	Sum of Specific Humidities	RTMA [37], WRF [80]	Humidity	0.152163	0.152752

Table A1. *Cont.*

Name	Description	Source	Variable Group	RTMA Drp. Loss	WRF Drp. Loss
peakSPFH	Mean of Spec. Humidity during peak winds	RTMA [37], WRF [80]	Humidity	0.151984	0.152767
avgWIND	Mean 10m Wind Speed	RTMA [37], WRF [80]	Wind/Gust	0.151961	0.152710
stdWIND	Standard Deviation of 10m Wind Speed	RTMA [37], WRF [80]	Wind/Gust	0.151954	0.152750
maxWIND	Maximum 10m Wind Speed	RTMA [37], WRF [80]	Wind/Gust	0.151977	0.152748
minWIND	Minimum 10m Wind Speed	RTMA [37], WRF [80]	Wind/Gust	0.151997	0.152745
sumWIND	Sum of Wind Speeds	RTMA [37], WRF [80]	Wind/Gust	0.151972	0.152716
peakWIND	Mean wind speed during peak winds	RTMA [37], WRF [80]	Wind/Gust	0.151948	0.152742
avgGUST	Mean Wind Gust Speed	RTMA [37], WRF [80]	Wind/Gust	0.152045	0.152836
stdGUST	Standard Deviation of Wind Gust Speed	RTMA [37], WRF [80]	Wind/Gust	0.151985	0.152769
maxGUST	Maximum Wind Gust Speed	RTMA [37], WRF [80]	Wind/Gust	0.152040	0.152752
minGUST	Minimum Wind Gust Speed	RTMA [37], WRF [80]	Wind/Gust	0.152089	0.152746
sumGUST	Sum of Wind Gusts	RTMA [37], WRF [80]	Wind/Gust	0.151988	0.152746
peakGUST	Mean Wind Gust Speed during peak winds	RTMA [37], WRF [80]	Wind/Gust	0.152039	0.152748
avgLFSH	Mean Leaf Stress	MODIS [9,79], RTMA [37], WRF [80]	Wind/Gust	0.151991	0.152744
stdLFSH	Standard Deviation of Leaf Stress	MODIS [9,79], RTMA [37], WRF [80]	Wind/Gust	0.151961	0.152738
maxLFSH	Maximum Leaf Stress	MODIS [9,79], RTMA [37], WRF [80]	Wind/Gust	0.151980	0.152743
minLFSH	Minimum Leaf Stress	MODIS [9,79], RTMA [37], WRF [80]	Wind/Gust	0.151963	0.152755
sumLFSH	Sum of Leaf Stresses	MODIS [9,79], RTMA [37], WRF [80]	Wind/Gust	0.152024	0.152760
peakLFSH	Mean Leaf Stress during peak winds	MODIS [9,79], RTMA [37], WRF [80]	Wind/Gust	0.151961	0.152826
wgt5	Hours of Winds >5 m/s	RTMA [37], WRF [80]	Wind/Gust	0.151974	0.152793
cowgt5	Continuous Hours of Winds >5 m/s	RTMA [37], WRF [80]	Wind/Gust	0.151952	0.152770
ggt13	Hours of Gusts >13 m/s	RTMA [37], WRF [80]	Wind/Gust	0.151967	**0.152997**
ggt17	Hours of Gusts >17 m/s	RTMA [37], WRF [80]	Wind/Gust	0.151932	0.152729
ggt22	Hours of Gusts >22 m/s	RTMA [37], WRF [80]	Wind/Gust	0.151934	0.152717
coggt13	Continuous Hours of Gusts >13 m/s	RTMA [37], WRF [80]	Wind/Gust	0.151935	0.152804
coggt17	Continuous Hours of Gusts >17 m/s	RTMA [37], WRF [80]	Wind/Gust	0.151940	0.152736
coggt22	Continuous Hours of Gusts >22 m/s	RTMA [37], WRF [80]	Wind/Gust	0.151945	0.152719
typWDIR	Typical (mean) wind direction of all storms	RTMA [37], WRF [80]	Wind/Gust	0.152005	0.152712
medWDIR	Median Wind direction of storm	RTMA [37], WRF [80]	Wind/Gust	0.152002	0.152806
difWDIR	Difference between typWDIR and medWDIR	RTMA [37], WRF [80]	Wind/Gust	0.151966	0.152745
avgPREC	Mean Hourly Precipitation Rate	Stage IV [38], WRF [80]	Precipitation	0.152209	0.152784
stdPREC	Standard Deviation of Precip. Rate	Stage IV [38], WRF [80]	Precipitation	**0.152403**	0.152731
maxPREC	Maximum Hourly Precipitation Rate	Stage IV [38], WRF [80]	Precipitation	**0.152844**	0.152773
sumPREC	Total Precipitation	Stage IV [38], WRF [80]	Precipitation	0.152187	0.152746
peakPREC	Mean Precip. Rate during peak winds	Stage IV [38], WRF [80]	Precipitation	**0.152311**	0.152726

Appendix B. Weather Correlations

Table A2. Correlation between RTMA and WRF weather datasets, and METAR and SPECI observations.

Name	Variable Group	RTMA—METAR R^2	WRF—METAR R^2
avgTMP	Temperature	0.9836	0.9129
stdTMP	Temperature	0.9119	0.6448
maxTMP	Temperature	0.9707	0.8686
minTMP	Temperature	0.9443	0.8592
sumTMP	Temperature	0.9119	0.8459
peakTMP	Temperature	0.7814	0.6480
avgDPT	Dew Point	0.9798	0.9461
stdDPT	Dew Point	0.9092	0.7349
maxDPT	Dew Point	0.9608	0.8966
minDPT	Dew Point	0.9511	0.8897
sumDPT	Dew Point	0.9234	0.8921
peakDPT	Dew Point	0.8348	0.7189
avgPRES	Pressure	0.1700	0.1588
stdPRES	Pressure	0.9766	0.9392
maxPRES	Pressure	0.1498	0.1363
minPRES	Pressure	0.2200	0.2038
sumPRES	Pressure	0.0015	0.0013
peakPRES	Pressure	0.1708	0.1469
avgSPFH	Humidity	0.9735	0.9274
stdSPFH	Humidity	0.8878	0.6932

Table A2. *Cont.*

Name	Variable Group	RTMA—METAR R²	WRF—METAR R²
maxSPFH	Humidity	0.9470	0.8648
minSPFH	Humidity	0.9500	0.8799
sumSPFH	Humidity	0.9204	0.8735
peakSPFH	Humidity	0.8219	0.7002
avgWIND	Wind/Gust	0.6346	0.5879
stdWIND	Wind/Gust	0.3217	0.1736
maxWIND	Wind/Gust	0.3327	0.2667
minWIND	Wind/Gust	0.5053	0.3046
sumWIND	Wind/Gust	0.6057	0.5643
peakWIND	Wind/Gust	0.3632	0.3246
avgGUST	Wind/Gust	0.5915	0.5056
stdGUST	Wind/Gust	0.1411	0.0627
maxGUST	Wind/Gust	0.2484	0.1067
minGUST	Wind/Gust	0.0060	0.0091
sumGUST	Wind/Gust	0.5789	0.4957
peakGUST	Wind/Gust	0.1487	0.0625
avgLFSH	Wind/Gust	0.5512	0.5444
stdLFSH	Wind/Gust	0.3583	0.2756
maxLFSH	Wind/Gust	0.2845	0.2249
minLFSH	Wind/Gust	0.4397	0.2735
sumLFSH	Wind/Gust	0.5382	0.5385
peakLFSH	Wind/Gust	0.2939	0.2786
wgt5	Wind/Gust	0.4230	0.4820
cowgt5	Wind/Gust	0.3837	0.3517
ggt13	Wind/Gust	0.4432	0.2149
ggt17	Wind/Gust	0.0352	0.0137
ggt22	Wind/Gust	NA[1]	0.0000
coggt13	Wind/Gust	0.4110	0.1665
coggt17	Wind/Gust	0.0396	0.0105
coggt22	Wind/Gust	NA[1]	0.0000
typWDIR	Wind/Gust	0.0054	0.1378
medWDIR	Wind/Gust	0.3357	0.0304
difWDIR	Wind/Gust	0.2362	0.0219
avgPREC	Precipitation	0.6056	0.0886
stdPREC	Precipitation	0.5589	0.0622
maxPREC	Precipitation	0.5298	0.0538
sumPREC	Precipitation	0.5585	0.0862
peakPREC	Precipitation	0.1989	0.0279

[1] Not enough variance to compute.

Appendix C. Error Metrics

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \frac{|P - A|}{A} \times 100 \tag{A1}$$

$$\text{CRMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} [(P_i - \bar{P}) - (A_i - \bar{A})]^2} \tag{A2}$$

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{N}(P_i - A_i)^2}{\sum_{i=1}^{N}(P_i - \bar{A})^2} \tag{A3}$$

$$\text{R}^2 = \left(\frac{\sum_{i=1}^{N}(P_i - \bar{P})(A_i - \bar{A})}{\sqrt{\sum_{i=1}^{N}(P_i - \bar{P})^2 \sum_{i=1}^{N}(A_i - \bar{A})^2}} \right)^2 \tag{A4}$$

$$\text{RMSLE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (log(P_i + 1) - log(A_i + 1))^2} \tag{A5}$$

References

1. *Economic Benefits of Increasing Electric Grid Resilience to Weather Outages*; Technical Report; Executive Office of the President: Washington, DC, USA, 2013.
2. Lubkeman, D.; Julian, D. Large scale storm outage management. In Proceedings of the IEEE Power Engineering Society General Meeting, Denver, CO, USA, 6–10 June 2004; Volume 2, pp. 16–22. [CrossRef]
3. Hall, K.L. *Out of Sight, Out of Mind*; Technical Report; Edison Electric Institute: Washington, DC, USA, 2012.
4. Mukherjee, S.; Nateghi, R.; Hastak, M. A multi-hazard approach to assess severe weather-induced major power outage risks in the U.S. *Reliab. Eng. Syst. Saf.* **2018**, *175*, 283–305. [CrossRef]
5. Sander, J.; Eichner, J.F.; Faust, E.; Steuer, M. Rising Variability in Thunderstorm-Related U.S. Losses as a Reflection of Changes in Large-Scale Thunderstorm Forcing. *Weather. Clim. Soc.* **2013**, *5*, 317–331. [CrossRef]
6. Diffenbaugh, N.S.; Scherer, M.; Trapp, R.J. Robust increases in severe thunderstorm environments in response to greenhouse forcing. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 16361–16366. [CrossRef]
7. Scaff, L.; Prein, A.F.; Li, Y.; Liu, C.; Rasmussen, R.; Ikeda, K. Simulating the convective precipitation diurnal cycle in North America's current and future climate. *Clim. Dyn.* **2020**, *55*, 369–382. [CrossRef]
8. Li, Z.; Singhee, A.; Wang, H.; Raman, A.; Siegel, S.; Heng, F.L.; Mueller, R.; Labut, G. Spatio-temporal forecasting of weather-driven damage in a distribution system. In Proceedings of the 2015 IEEE Power & Energy Society General Meeting, Denver, CO, USA, 26–30 July 2015; pp. 1–5. [CrossRef]
9. Cerrai, D.; Wanik, D.W.; Bhuiyan, M.A.E.; Zhang, X.; Yang, J.; Frediani, M.E.B.; Anagnostou, E.N. Predicting Storm Outages Through New Representations of Weather and Vegetation. *IEEE Access* **2019**, *7*, 29639–29654. [CrossRef]
10. Wanik, D.W.; Anagnostou, E.N.; Hartman, B.M.; Frediani, M.E.B.; Astitha, M. Storm outage modeling for an electric distribution network in Northeastern USA. *Nat. Hazards* **2015**, *79*, 1359–1384. [CrossRef]
11. Kankanala, P.; Das, S.; Pahwa, A. AdaBoost$^+$: An Ensemble Learning Approach for Estimating Weather-Related Outages in Distribution Systems. *IEEE Trans. Power Syst.* **2014**, *29*, 359–367. [CrossRef]
12. Han, S.R.; Guikema, S.D.; Quiring, S.M.; Lee, K.H.; Rosowsky, D.; Davidson, R.A. Estimating the spatial distribution of power outages during hurricanes in the Gulf coast region. *Reliab. Eng. Syst. Saf.* **2009**, *94*, 199–210. [CrossRef]
13. Quiring, S.M.; Zhu, L.; Guikema, S.D. Importance of soil and elevation characteristics for modeling hurricane-induced power outages. *Nat. Hazards* **2011**, *58*, 365–390. [CrossRef]
14. Guikema, S.D.; Nateghi, R.; Quiring, S.M.; Staid, A.; Reilly, A.C.; Gao, M. Predicting Hurricane Power Outages to Support Storm Response Planning. *IEEE Access* **2014**, *2*, 1364–1373. [CrossRef]
15. McRoberts, D.B.; Quiring, S.M.; Guikema, S.D. Improving Hurricane Power Outage Prediction Models Through the Inclusion of Local Environmental Factors. *Risk Anal.* **2018**, *38*, 2722–2737. [CrossRef]
16. D'Amico, D.F.; Quiring, S.M.; Maderia, C.M.; McRoberts, D.B. Improving the Hurricane Outage Prediction Model by including tree species. *Clim. Risk Manag.* **2019**, *25*, 100193. [CrossRef]
17. Yang, F.; Watson, P.; Koukoula, M.; Anagnostou, E.N. Enhancing Weather-Related Power Outage Prediction by Event Severity Classification. *IEEE Access* **2020**, *8*, 60029–60042. [CrossRef]
18. Watson, P.L.; Cerrai, D.; Koukoula, M.; Wanik, D.W.; Anagnostou, E. Weather-related power outage model with a growing domain: Structure, performance, and generalisability. *J. Eng.* **2020**, *2020*, 817–826. [CrossRef]
19. Tervo, R.; Láng, I.; Jung, A.; Mäkelä, A. Predicting power outages caused by extratropical storms. *Nat. Hazards Earth Syst. Sci.* **2021**, *21*, 607–627. [CrossRef]
20. Singhee, A.; Wang, H. Probabilistic forecasts of service outage counts from severe weather in a distribution grid. In Proceedings of the 2017 IEEE Power & Energy Society General Meeting, Chicago, IL, USA, 16–20 July 2017; pp. 1–5. [CrossRef]
21. Yue, M.; Toto, T.; Jensen, M.P.; Giangrande, S.E.; Lofaro, R. A Bayesian Approach-Based Outage Prediction in Electric Utility Systems Using Radar Measurement Data. *IEEE Trans. Smart Grid* **2018**, *9*, 6149–6159. [CrossRef]
22. Zhou, Y.; Pahwa, A.; Yang, S.S. Modeling Weather-Related Failures of Overhead Distribution Lines. *IEEE Trans. Power Syst.* **2006**, *21*, 1683–1690. [CrossRef]
23. Kankanala, P.; Pahwa, A.; Das, S. Regression models for outages due to wind and lightning on overhead distribution feeders. In Proceedings of the 2011 IEEE Power and Energy Society General Meeting, San Detroit, MI, USA, 24–28 July 2011; pp. 1–4. [CrossRef]
24. Hohenegger, C.; Schar, C. Atmospheric Predictability at Synoptic Versus Cloud-Resolving Scales. *Bull. Am. Meteorol. Soc.* **2007**, *88*, 1783–1794. [CrossRef]
25. Sun, J.; Xue, M.; Wilson, J.W.; Zawadzki, I.; Ballard, S.P.; Onvlee-Hooimeyer, J.; Joe, P.; Barker, D.M.; Li, P.W.; Golding, B.; et al. Use of NWP for Nowcasting Convective Precipitation: Recent Progress and Challenges. *Bull. Am. Meteorol. Soc.* **2014**, *95*, 409–426. [CrossRef]
26. Yano, J.I.; Ziemiański, M.Z.; Cullen, M.; Termonia, P.; Onvlee, J.; Bengtsson, L.; Carrassi, A.; Davy, R.; Deluca, A.; Gray, S.L.; et al. Scientific Challenges of Convective-Scale Numerical Weather Prediction. *Bull. Am. Meteorol. Soc.* **2018**, *99*, 699–710. [CrossRef]
27. Papadopoulos, A.; Chronis, T.G.; Anagnostou, E.N. Improving Convective Precipitation Forecasting through Assimilation of Regional Lightning Measurements in a Mesoscale Model. *Mon. Weather Rev.* **2005**, *133*, 1961–1977. [CrossRef]
28. Hu, M.; Xue, M. Impact of Configurations of Rapid Intermittent Assimilation of WSR-88D Radar Data for the 8 May 2003 Oklahoma City Tornadic Thunderstorm Case. *Mon. Weather Rev.* **2007**, *135*, 507–525. [CrossRef]

29. Benjamin, S.G.; Weygandt, S.S.; Brown, J.M.; Hu, M.; Alexander, C.R.; Smirnova, T.G.; Olson, J.B.; James, E.P.; Dowell, D.C.; Grell, G.A.; et al. A North American Hourly Assimilation and Model Forecast Cycle: The Rapid Refresh. *Mon. Weather Rev.* **2016**, *144*, 1669–1694. [CrossRef]

30. Clark, A.J.; Gallus, W.A.; Xue, M.; Kong, F. A Comparison of Precipitation Forecast Skill between Small Convection-Allowing and Large Convection-Parameterizing Ensembles. *Weather Forecast.* **2009**, *24*, 1121–1140. [CrossRef]

31. Roberts, B.; Gallo, B.T.; Jirak, I.L.; Clark, A.J. The High Resolution Ensemble Forecast (HREF) system: Applications and Performance for Forecasting Convective Storms. *Meteorology* **2019**. [CrossRef]

32. Bouttier, F.; Marchal, H. Probabilistic thunderstorm forecasting by blending multiple ensembles. *Tellus A Dyn. Meteorol. Oceanogr.* **2020**, *72*, 1–19. [CrossRef]

33. Alpay, B.A.; Wanik, D.; Watson, P.; Cerrai, D.; Liang, G.; Anagnostou, E. Dynamic Modeling of Power Outages Caused by Thunderstorms. *Forecasting* **2020**, *2*, 151–162. [CrossRef]

34. Sheild, S.A.; Quiring, S.M.; McRoberts, D.B. Development of a Thunderstorm Outage Prediction Model. Ph.D. Thesis, The Ohio State University, Columbus, OH, USA, 2018.

35. Kabir, E.; Guikema, S.D.; Quiring, S.M. Predicting Thunderstorm-Induced Power Outages to Support Utility Restoration. *IEEE Trans. Power Syst.* **2019**, *34*, 4370–4381. [CrossRef]

36. Babbage, C. *Passages from the Life of a Philosopher*; Longman, Green, Longman, Roberts, & Green: London, UK, 1864.

37. De Pondeca, M.S.F.V.; Manikin, G.S.; DiMego, G.; Benjamin, S.G.; Parrish, D.F.; Purser, R.J.; Wu, W.S.; Horel, J.D.; Myrick, D.T.; Lin, Y.; et al. The Real-Time Mesoscale Analysis at NOAA's National Centers for Environmental Prediction: Current Status and Development. *Weather Forecast.* **2011**, *26*, 593–612. [CrossRef]

38. Nelson, B.R.; Prat, O.P.; Seo, D.J.; Habib, E. Assessment and Implications of NCEP Stage IV Quantitative Precipitation Estimates for Product Intercomparisons. *Weather Forecast.* **2016**, *31*, 371–394. [CrossRef]

39. NOAA/NWS. *RTMA: Real-Time Mesoscale Analysis Data*; NOAA/NWS: Washington, DC, USA, 2015

40. Hudlow, M.D. Technological Developments in Real-Time Operational Hydrologic Forecasting in the United States. *J. Hydrol.* **1988**, *102*, 69–92. [CrossRef]

41. Environmental Modeling Center; National Centers for Environmental Prediction; National Weather Service; NOAA; U.S. Department of Commerce. *NCEP North American Mesoscale (NAM) 12 km Analysis*; U.S. Department of Commerce: Washington, DC, USA, 2015.

42. Morrison, H.; Thompson, G.; Tatarskii, V. Impact of Cloud Microphysics on the Development of Trailing Stratiform Precipitation in a Simulated Squall Line: Comparison of One- and Two-Moment Schemes. *Mon. Weather Rev.* **2009**, *137*, 991–1007. [CrossRef]

43. Mlawer, E.J.; Taubman, S.J.; Brown, P.D.; Iacono, M.J.; Clough, S.A. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res. Atmos.* **1997**, *102*, 16663–16682. [CrossRef]

44. Chou, M.; Suarez, M. An Efficient Thermal Infrared Radiation Parameterization for Use in General Circulations Models. *NASA Tech. Memo.* **1994**, *3*, 1–85.

45. Jiménez, P.A.; Dudhia, J.; González-Rouco, J.F.; Navarro, J.; Montávez, J.P.; García-Bustamante, E. A Revised Scheme for the WRF Surface Layer Formulation. *Mon. Weather Rev.* **2012**, *140*, 898–918. [CrossRef]

46. Tewari, M.; Chen, F.; Wang, W.; Dudhia, J.; LeMone, M.; Mitchell, K.; Ek, M.; Gayno, G.; Wegiel, J.; Cuenca, R. Implementation and verification of the unified NOAH land surface model in the WRF model. In Proceedings of the 20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction, Seattle, WA, USA, 11–15 January 2004; Volume 1115, pp. 2165–2170.

47. Hong, S.; Noh, Y.; Dudhia, J. A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes. *Mon. Weather Rev.* **2006**, *134*, 2318–2341. [CrossRef]

48. Agostinelli, C.; Lund, U. *R Package Circular: Circular Statistics (Version 0.4-93)*; Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University: Venice, Italy; UL: Department of Statistics, California Polytechnic State University: San Luis Obispo, CA, USA, 2017. Available online: https://cran.r-project.org/web/packages/circular/circular.pdf (accessed on 2 February 2021).

49. Bivand, R.; Keitt, T.; Rowlingson, B. rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R Package Version 1.5-23. 2021. Available online: https://cran.r-project.org/web/packages/rgdal/index.html (accessed on 2 February 2021).

50. Bivand, R.; Rundel, C. rgeos: Interface to Geometry Engine—Open Source ('GEOS'). R Package Version 0.5-5. 2020. Available online: https://cran.r-project.org/web/packages/rgeos/index.html (accessed on 1 June 2020).

51. Jin, S.; Homer, C.; Yang, L.; Danielson, P.; Dewitz, J.; Li, C.; Zhu, Z.; Xian, G.; Howard, D. Overall Methodology Design for the United States National Land Cover Database 2016 Products. *Remote Sens.* **2019**, *11*, 2971. [CrossRef]

52. Coulston, J.W.; Moisen, G.G.; Wilson, B.T.; Finco, M.V.; Cohen, W.B.; Brewer, C.K. Modeling Percent Tree Canopy Cover: A Pilot Study. *Photogramm. Eng. Remote Sens.* **2012**, *78*, 715–727. [CrossRef]

53. Potapov, P.; Li, X.; Hernandez-Serna, A.; Tyukavina, A.; Hansen, M.C.; Kommareddy, A.; Pickens, A.; Turubanova, S.; Tang, H.; Silva, C.E.; et al. Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sens. Environ.* **2021**, *253*, 112165. [CrossRef]

54. Gesch, D.; Evans, G.; Oimoen, M.; Arundel, S. *The National Elevation Dataset*; American Society for Photogrammetry and Remote Sensing: Bethseda, MD, USA, 2018; pp. 83–110.

55. Soil Survey Staff, Natural Resources Conservation Service. Soil Survey Geographic (SSURGO) Database. 2010. Available online: https://websoilsurvey.nrcs.usda.gov/ (accessed on 21 August 2020).

56. Individual Tree Species Parameter Maps. 2015. Available online: https://www.fs.fed.us/foresthealth/applied-sciences/mapping-reporting/indiv-tree-parameter-maps.shtml (accessed on 19 March 2021).

57. Abatzoglou, J.T.; McEvoy, D.J.; Redmond, K.T. The West Wide Drought Tracker: Drought Monitoring at Fine Spatial Scales. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 1815–1820. [CrossRef]

58. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021.

59. Kursa, M.B.; Jankowski, A.; Rudnicki, W.R. Boruta—A System for Feature Selection. *Fundam. Inform.* **2010**, *101*, 271–285. [CrossRef]

60. Kursa, M.B.; Rudnicki, W.R. Feature Selection with the Boruta Package. *J. Stat. Softw.* **2010**, *36*, 1–13. [CrossRef]

61. Chipman, H.A.; George, E.I.; McCulloch, R.E. BART: Bayesian additive regression trees. *Ann. Appl. Stat.* **2010**, *4*, 266–298. [CrossRef]

62. Sparapani, R.; Spanbauer, C.; McCulloch, R. Nonparametric Machine Learning and Efficient Computation with Bayesian Additive Regression Trees: The BART R Package. *J. Stat. Softw.* **2021**, *97*, 1–66. [CrossRef]

63. Ardia, D.; Boudt, K.; Carl, P.; Mullen, K.M.; Peterson, B.G. Differential Evolution with DEoptim: An Application to Non-Convex Portfolio Optimization. *R J.* **2011**, *3*, 27–34. [CrossRef]

64. Mullen, K.; Ardia, D.; Gil, D.; Windover, D.; Cline, J. DEoptim: An R Package for Global Optimization by Differential Evolution. *J. Stat. Softw.* **2011**, *40*, 1–26. [CrossRef]

65. National Centers for Environmental Information. Integrated Surface Data (ISD) Archive. Available online: https://www.ncei.noaa.gov/data/global-hourly/access/ (accessed on 3 March 2021).

66. Nash, J.; Sutcliffe, J. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290. [CrossRef]

67. Roberts, N.M.; Lean, H.W. Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. *Mon. Weather Rev.* **2008**, *136*, 78–97. [CrossRef]

68. Gilleland, E.; Ahijevych, D.A.; Brown, B.G.; Ebert, E.E. Verifying Forecasts Spatially. *Bull. Am. Meteorol. Soc.* **2010**, *91*, 1365–1376. [CrossRef]

69. Mittermaier, M.; Roberts, N. Intercomparison of Spatial Forecast Verification Methods: Identifying Skillful Spatial Scales Using the Fractions Skill Score. *Weather Forecast.* **2010**, *25*, 343–354. [CrossRef]

70. Laboratory, N.R.A. Verification: Weather Forecast Verification Utilities. R Package Version 1.42. 2015. Available online: https://CRAN.R-project.org/package=verification (accessed on 3 March 2021).

71. Gilleland, E. SpatialVx: Spatial Forecast Verification. R Package Version 0.8. 2021. Available online: https://CRAN.R-project.org/package=SpatialVx (accessed on 3 March 2021).

72. Fisher, A.; Rudin, C.; Dominici, F. All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. *J. Mach. Learn. Res.* **2019**, *20*, 1–81.

73. Biecek, P. DALEX: Explainers for Complex Predictive Models in R. *J. Mach. Learn. Res.* **2018**, *19*, 1–5.

74. Lee, T.R.; Buban, M.; Turner, D.D.; Meyers, T.P.; Baker, C.B. Evaluation of the High-Resolution Rapid Refresh (HRRR) Model Using Near-Surface Meteorological and Flux Observations from Northern Alabama. *Weather Forecast.* **2019**, *34*, 635–663. [CrossRef]

75. Pichugina, Y.L.; Banta, R.M.; Bonin, T.; Brewer, W.A.; Choukulkar, A.; McCarty, B.J.; Baidar, S.; Draxl, C.; Fernando, H.J.S.; Kenyon, J.; et al. Spatial Variability of Winds and HRRR–NCEP Model Error Statistics at Three Doppler-Lidar Sites in the Wind-Energy Generation Region of the Columbia River Basin. *J. Appl. Meteorol. Climatol.* **2019**, *58*, 1633–1656. [CrossRef]

76. Shucksmith, P.E.; Sutherland-Stacey, L.; Austin, G.L. The spatial and temporal sampling errors inherent in low resolution radar estimates of rainfall: Spatial and temporal sampling errors in low resolution radar estimates of rainfall. *Meteorol. Appl.* **2011**, *18*, 354–360. [CrossRef]

77. Moreau, E.; Testud, J.; Le Bouar, E. Rainfall spatial variability observed by X-band weather radar and its implication for the accuracy of rainfall estimates. *Adv. Water Resour.* **2009**, *32*, 1011–1019. [CrossRef]

78. Zhou, L.; Lin, S.J.; Chen, J.H.; Harris, L.M.; Chen, X.; Rees, S.L. Toward Convective-Scale Prediction within the Next Generation Global Prediction System. *Bull. Am. Meteorol. Soc.* **2019**, *100*, 1225–1243. [CrossRef]

79. Leaf Area Index (1 Month—Terra/MODIS). 2017. Available online: https://modis.gsfc.nasa.gov/data/dataprod/mod15.php (accessed on 17 December 2020).

80. Community, WRF. *Weather Research and Forecasting (WRF) Model*; UCAR/NCAR: Boulder, CO, USA, 2000. [CrossRef]