Storm Outage Modeling for an Electric Distribution Network in Northeastern US

D. W. Wanik[1], E. N. Anagnostou[1], B. M. Hartman[2], M. E. B. Frediani[1], M. Astitha[1]

[1] Department of Civil and Environmental Engineering, University of Connecticut, Storrs, CT

[2] Department of Mathematics, University of Connecticut Storrs, CT

Submitted to *Natural Hazards*

Initially submitted to reviewers: December 2014

Returned to authors for major revisions: March 2015

Resubmitted to reviewers with revisions: May 2015

Corresponding Author: Prof. Emmanouil Anagnostou, Civil and Environmental Engineering, University of Connecticut, Storrs, CT 06269 | Email: manos@engr.uconn.edu| Tel.: 860-486-6806

Abstract

The interaction of severe weather, overhead lines and surrounding trees is the leading cause of outages to an electric distribution network in forested areas. In this paper, we show how utility-specific infrastructure and land cover data, aggregated around overhead lines, can improve outage predictions for Eversource Energy (formerly Connecticut Light & Power), the largest electric utility in Connecticut. Eighty nine storms from different seasons (cold weather, warm weather, transition months) in the period 2005 – 2014, representing varying types (thunderstorms, blizzards, nor'easters, hurricanes) and outage severity, were simulated using the Weather Research and Forecasting (WRF) atmospheric model. WRF simulations were joined with utility damage data to calibrate four types of models: a decision tree (DT), random forest (RF), boosted gradient tree (BT) and an ensemble (ENS) decision tree regression that combined predictions from DT, RF and BT. The study shows that ENS model forced with weather, infrastructure and land cover data was superior to the other models we evaluated, especially in terms of predicting the spatial distribution of outages. This framework could be used for predicting outages to other types of critical infrastructure networks with benefits for emergency preparedness functions in terms of equipment staging and resource allocation.

Keywords: Electric distribution network, critical infrastructure damage modeling, data mining, numerical weather prediction, land cover, hurricanes.

1. Introduction

The severe storms of 2011 and 2012 will resonate in the minds of Connecticut's populous for years to come. For the first time since Hurricane Gloria (1985) impacted Connecticut, prolonged power outages (longer than ten days) occurred three times within the span of fifteen months during Storm Irene (2011), the October nor'easter (2011), and Hurricane Sandy (2012). The storms affected hundreds of thousands of customers and each caused hundreds of millions of dollars of damage to the State. Several investigative reports by regulators (McGee et al. 2012) and consultants (Davies Consulting 2012, O'Neill et al. 2013, Witt Associates 2011) followed the major events, resulting in several improvement recommendations for Connecticut's utilities. One of the recommendations from the reports was that electric utility companies should use outage prediction models to support utility emergency preparedness efforts before a storm event. Such a model could aid the pre-storm deployment of crews and resources (i.e. poles, transformers, and conductors), thereby decreasing restoration times and increasing reliability to customers. In this paper, we present new research on predicting outage locations ("outages") from severe weather in Connecticut. We define outages as locations that require a manual intervention to restore power, which is separate from modeling the number of customers affected ("customer outages").

Much research has been done on storm-related impacts to the electric distribution network; including predicting damages to overhead lines (i.e. broken poles) (Guikema et al. 2010); predicting the number of outages that need to be repaired (Guikema et al. 2014a, Mensah and Duenas-Osorio 2014); predicting the associated customers affected by power outages (Guikema et al. 2008, Guikema et al. 2014b, Han et al. 2009b), and modeling the length of outage durations during major storm events (Liu et al. 2007, Nateghi et al. 2011, Nateghi et al. 2014b). While each of these are distinct research topics, the underlying fundamentals of each problem are similar; critical infrastructure and environmental data are related to actual utility data (i.e. outages, customers or damages), which tend to be zero-inflated data with nonlinear response thresholds (Guikema and Coffelt 2009). In addition, modeling utility-related problems is complex due to different interactions involved (e.g. tree conditions, soil saturation, infrastructure age). To address this complexity, an assortment of methods have been used for utility-related problems; including generalized linear models (GLMs) (Cerruti and Decker 2012, Hongfei Li et al. 2010), spatial and non-spatial generalized linear mixed models (GLMMs) (Guikema and Davidson 2006, Liu et al. 2008), generalized additive models (GAMs) (Han et al. 2009a), classification and regression trees (CART) (Quiring et al. 2011), random forest (Nateghi et al. 2014a) and Bayesian additive regression trees (BART) (Nateghi et al. 2011). In addition to count data models, probabilistic models have also been coupled with physical models of the electric system with the aim to predict failures on both transmission and distribution lines (Mensah and Duenas-Osorio 2014). The evolution of the implementation of these models is also interesting; many of these models have been implemented as i) individual models, ii) average of multiple individual models, or iii) as part of a hybrid two-stage model (Guikema and Quiring 2012).

Recent literature (Nateghi et al. 2014a) has shown that the random forest model is superior to other models that have been built on the same set of hurricane data (Guikema and Quiring 2012, Han et al. 2009a, Han et al. 2009b). In addition to modeling improvements, the quality and granularity of utility-specific data (i.e. tree-trimming and overhead line infrastructure) and environmental data (i.e. soil conditions, aspect ratio, and elevation) used as forcing parameters has led to models better representing the physical processes that cause outages. As a complement to these data-intensive/utility-specific models, there has been additional research dedicated to investigating whether publicly available data can be used in lieu of proprietary, utility-specific data (i.e. using population counts for Census data rather than using actual customer data) such that the calibrated models can be generalized to other areas. We authors believe this area of research is exciting and important as such early warning tools can lead to better emergency preparedness efforts. Recent work by Nateghi et al. (2014a) has shown that these generalized models have a marginal yet acceptable decrease in accuracy than the utility-specific models, which allows for the calibrated models to be applied to other service territories for which outage models don't currently exist. In addition to short-term outage models, other research has extended these generalized models into a long-term evaluation of tropical cyclone risk from climate change (Staid et al. 2014).

Related research by Guikema et al. (2014b) have taken their utility-specific customer outage model for a Gulf Coast utility, called Hurricane Outage Prediction Model (HOPM), to create the Spatially Generalized Hurricane Outage Prediction Model (SGHOPM) to predict customer outages for Hurricane Sandy along the Eastern seaboard. Although SGHOPM did well for many regions (including Massachusetts and Rhode Island), it underestimated customer outages that impacted Connecticut. The authors suggest that a large amount of customer outages in Connecticut might have been caused by storm surge which wouldn't be captured by SGHOPM, though conceding this required further investigation. Although the authors are correct that storm surge was abundant and catastrophic in Connecticut during Sandy, with some coastal stations reporting >12 feet of surge (Fanelli et al. 2013), the storm surge only contributed to a minor fraction of the customer outages in the Eversource and neighboring United Illuminating service territories. According to sources at each utility, the majority of outages were actually caused by trees interacting with the overhead lines. This might highlight that not all distribution utilities respond similarly to severe weather; a 50 mph wind gust may have a different impact in

Connecticut than it would in a Gulf Coast state, which we believe is a function of the overhead infrastructure and the surrounding vegetation. Connecticut is among the most forested and densely populated regions in the country as measured by the amount of wildland-urban interface (Radeloff et al. 2005), which makes the region especially susceptible to tree-related utility damage (Wagner et al. 2012).

In this paper, we build on the current research regarding modeling outages on the overhead distribution lines. While most outage models have focused on hurricanes, we will use high-resolution numerical weather simulations for 89 storms of varying type (e.g. hurricanes, blizzards, thunderstorms) and severity (from 20 outages to >15,000 outages). We will attempt to answer the following questions: (1) if utility outage data exists, how accurately can a predictive model relate high-resolution numerical weather simulation data to outages for a range of storm types, severities and seasons (e.g. warm weather, cold weather and transition months)?; and (2) how much added performance does the utility-specific data (e.g. land cover data aggregated around overhead lines and distribution infrastructure data) contribute to magnitude (count of outages) and spatial (distribution of predicted outages) error metrics?

The paper is organized into the following additional sections. Section 2 explains the study area and datasets used in the model. Section 3 covers the models used to predict utility damages and the model validation strategy. Section 4 presents the results of all models evaluated, as well as a selection of the best model overall. Section 5 focuses on the hurricanes outage modeling results for the most simple and complex models we evaluated. Section 6 provides discussion of all the results and comparison to other outage models in the literature, followed by the conclusion in Section 7.

2. Study Area and Datasets

Eversource Energy (“Eversource”), formerly the Connecticut Light and Power Company (CL&P), is the largest investor-owned utility in Connecticut and distributes electricity to 1.2 million customers in 149 towns across Connecticut via >18,000 miles of overhead distribution lines (Connecticut Light & Power 2014). Each of the 149 towns belongs to one of 18 Area Work Centers (AWCs) which are used to organize restoration crews (note that AWCs can have up to 15 associated towns, Figure 1). Although Eversource also serves customers in Massachusetts and New Hampshire, we focus solely on Connecticut in this paper. Connecticut has a wide variety of land cover conditions; from a southerly coastal landscape, to urban centers in central and southwestern Connecticut, to the forested uplands of eastern and western Connecticut. Population density is most concentrated in the central and coastal areas.

2.1 Weather Simulations

We simulated the weather for 89 storms that impacted the Eversource service territory between 2005 and 2014 using the Advanced Research (ARW) dynamics core of the Weather Research and Forecasting Model (WRF) model version 3.4.1 (Skamarock et al. 2008). The events were dynamically downscaled from analyzed fields provided by the Global Forecast System (GFS, at 6-hourly intervals with 1.0 degree grid resolution) produced by the National Center for Environmental Prediction (NCEP). In order to minimize initial condition (IC) and boundary condition (BC) errors, the events were modeled as hindcasts (e.g. the analyses are used to derive both the model's IC and BC updates).

For the WRF setup, three nested domains (Figure 2) were created to gradually downscale from the 1.0 deg GFS analysis to a 2 km resolution: an outer domain with resolution of 18 km, an inner-intermediate domain with 6 km, and the focus area with 2 km with a topography dataset at 30 arc-second (~1000 m) resolution. A subset of the inner most domain provides the modeled atmospheric conditions, which are derived from the grid cells within the area of this study.

WRF was configured to use a 30 second timestep, 2-way feedback between nested grids, and 28 vertical levels. The schemes used to parameterize the physical processes included the Thompson for cloud microphysics (Thompson et al. 2008); Grell 3D for convection (Grell and Devenyi 2002), with the 2 km inner nest solved explicitly; RRTM for Long Wave radiation (Mlawer et al. 1997), initialized each 18, 6, and 2 minutes for each domain, respectively; Goddard for Short Wave radiation (Chou and Suarez 1994); MM5 similarity for Surface Layer (Zhang and Anthes 1982); Unified NOAH for Land Surface Model (Tewari et al. 2004); Yonsei for Planetary Boundary Layer (Song-You Hong et al. 2006); and topographic correction for surface wind to represent extra-drag from sub-grid topography and enhanced flow at hill tops (Jimenez and Dudhia 2012); all the others settings were left to the default configuration.

For each event, the model was initialized 6 hours prior to the time of the first damage report running in the Eversource network for a 60 hour simulation time with hourly outputs. From these outputs, various wind and precipitation variables were derived and reduced to sustained mean and maximum representative value per grid cell (Table 1). The maximum value per grid cell refers to the maximum value over the duration of the 60 hour simulation, while the sustained mean value per grid cell refers to the maximum 4 hour mean from a “running window” during the simulation. In addition to these sustained and maximum values, we also calculated the duration of 10m wind speed and wind gust above a range of thresholds (e.g. 9, 13, 18 m/s for 10 m wind speed, and 18, 27, 36, 45 m/s for gust winds). In terms of precipitation we used WRF-derived storm-total accumulated liquid and solid (snow and ice) precipitation and soil moisture. The impact of heavy rains has been shown to be significant in cases of stationary storms (typically associated with complex terrain areas), which can exemplify wind effects due to saturated soils (Guikema et al. 2014b). On the other hand, blizzards and freezing rain can enhance the effect of winds on tree damages.

2.2 Weather Simulation Evaluation

Given the significance of winds on the outage predictions, numerical weather simulations of selected major storm events were evaluated against wind speed observations using data from airport stations provided by the National Centers for Environmental Prediction (NCEP) ADP Global Upper Air and Surface Weather Observations (NCEP/NWS/NOAA/USDC 2008). Specifically, wind speed at 10 m above ground is compared to modeled wind speed taken at the gridded location of each airport station. The error analysis was performed on the data pairs (WRF and NCEP ADP stations) of 10 m sustained wind speed (SWS) time series and the corresponding maximum 10 m sustained wind speed values from each station location. Sustained wind speed is calculated similarly to the way used in the DPM model, namely, taking a 4 hour running window that spans the entire duration of the simulated event. Error analysis results are presented for three major storms: Storm Irene ("Irene", 2011), Hurricane Sandy ("Sandy", 2012) and the Nemo blizzard ("Nemo", 2013). Details on the statistical metrics, including name conventions and mathematical formulas, are provided in the Appendix.

The model predictions of sustained wind speed at 10 m above ground have shown acceptable agreement with the observations. This step was necessary to gain confidence in the numerical weather prediction of extreme events for northeastern U.S. and use the model data as one of the drivers of the damage prediction model. Although precipitation was not evaluated in this study, it is noted that winds and precipitation processes resolved in the model are based upon the same atmospheric physics, with precipitation imposing added complexity due to microphysical processes. Scatter plots of observed versus modeled data show the linear correlation between calculated and measured horizontal wind fields (Figure 3). The model simulations for Irene and Sandy exhibit similar patterns with predictions close to observed values. The mean bias (MB) is low (0.12 - 0.56 ms^{-1}) and the correlation varies from 0.6 to 0.8 depending on the atmospheric variable (Table 2). Similar performance is seen for the Nemo blizzard, with the only difference that the model slightly under-predicted the observations (negative MB and NSE) with an overall high correlation coefficient (0.74) for both wind parameters (sustained and maximum wind). Among other metrics, the statistical metric denoted as percentage within a fraction of 2 (FAC2) has been widely used in the atmospheric and air quality modeling community for the evaluation of predicted values (Astitha et al. 2010, Builtjes 2005, Chang and Hanna 2004, Hendrick et al. 2013). FAC2 uses the multiplicative bias (model/observation) for each model-observation pair instead of the difference between the values. The percentage within a factor of two shows how many model-observation pairs are within an acceptable range (predicted values must be between half and twice the observations, with 1 being the ideal situation).The fraction of SWS within a factor of 2 for a series of model-observation pairs was 93% for Nemo, 92% for Irene and 70% for Sandy. The fraction of SWS within a factor of 1.5 was 88% for Nemo, 84% for Irene and 66% for Sandy. This statistical metric is considered more robust than the traditional correlation coefficient since it is not sensitive to outlier data pairs (high or low) (Chang and Hanna 2004). In all cases, the model correctly captured the diurnal variation of the wind field in the majority of the stations (not shown here). In addition, the uncertainty ratio (characterized as the ratio of standard deviation from modeled to observed fields) in the cases shown herein varies between 1.03 and 1.2 indicating strong similarity in the predicted and observed variability of wind simulations. Although, precipitation parameters are also used in forcing damage prediction model, errors in forecasts of precipitation are not evaluated herein. Precipitation information has limited contribution to the damage prediction, mainly as an index for enhancing the impact of severe winds in damage prediction. Future investigations based on more accurate spatial precipitation data (such as those derived from weather radar) could be used to enhance the use of precipitation information in damage prediction.

2.3 Seasonal Categorization

Storms were categorized based on their month of occurrence: "warm weather storms" included storms from June through September, "cold weather storms" included storms from December through March, "transition storms" occurred in April, May, October and November (Table 3). The labeling of storms allows this categorical variable to be included in the model. Each season category has an average leaf index associated with it, and storm characteristics tend to be more similar per season. For example, trees would hold leaves during warm weather storms, not during the cold weather storms, and hold some for transition storms. Warm weather storms tend to be predominately driven by convective and mesoscale processes, while cold weather storms tend to be predominately driven by synoptic scale processes, and transition storms can be characterized by either mesoscale or synoptic processes, as well as nor'easters (Jiménez et al. 2008, Wallace and Hobbs 2006).

2.4 Utility Outages

The response variable in our models was the count of outages per grid cell. Outages are defined by Eversource as "extended interruptions (>5 minutes) of service to one or more customers, that usually requires human intervention to restore electric service (Connecticut Light & Power 2014)." For reference, the median number of outages on a normal day with low wind is typically around 40. During Hurricane Sandy and Storm Irene each event had >15,000 outages, which is equivalent to more than an entire year's worth of outages caused by one storm (calculated as total outages divided by median number of outages per day).

Eversource provided detailed records of outages outputted from their Outage Management System (OMS) for each of the storms we simulated. The OMS records included geographic coordinates, nearest substation, customers affected, town, regional operating center, date, time, outage length and circuit affected. In general, analysts should use caution when working with OMS data, as much as the data inputted by lineman can be erroneous; in an effort to save time, the lineman may enter the first entry of a dropdown list into a data collection system, even if incorrect. However, per a personal communication with System Engineering, we were authorized to delete duplicate records and records with "cause codes" not related to storm damages (i.e. damage caused by animals or vandalism.) The events that were deleted represented approximately 5% of all observations.

Eversource does not track outages at individual metered locations; instead they rely on its customers to notify them of outages. After that, predictive algorithms automatically approximate the location of the damage to the nearest isolating device (i.e. transformers, fuses, reclosers, switches). Once the possible outage is recorded into the OMS, a crew is dispatched to find and repair the damage, and closes out the outage record once restoration is complete.

2.5 Distribution Infrastructure

Eversource provided detailed geographic data about their electric distribution system in the form of polylines of the overhead distribution lines and point shapefiles of isolating devices and poles. Although overhead distribution lines and pole locations were provided, these ultimately were excluded from the model because outages are recorded at the nearest isolating device (and not the nearest pole). Holding everything else constant, a grid cell with one mile of overhead lines and one isolating device will theoretically only have one outage attributed to it, while a grid cell with one mile of overhead lines and 100 isolating devices will likely have many more outages attributed to it. Outages can occur anywhere on the overhead lines, and the isolating device may be of any type. Rather than aggregating the data by isolating device type (i.e. counts of transformer per grid cell), the total number of all isolating devices was aggregated by grid cell as a term called “sumAssets”. As the sum of isolating devices increases per grid cell so does the opportunity that a trouble spot will be recorded simply by virtue of an isolating device to be there. Overhead line length was not used as a variable in the models directly but was used to calculate the percentage of land cover around overhead lines per grid cell, which we discuss next.

2.6 Land Cover

Overhead lines directly interact with the environment that surrounds them. Trees are the predominant cause of damages to the Eversource distribution system (Connecticut Light & Power 2014), and vegetation management (colloquially referred to as “tree trimming”) has been shown to decrease customer outages (Guikema et al. 2006a). Specific trees that have the potential to damage the overhead lines are referred to as “hazard trees”. The interaction between trees and overhead lines is inherently localized and complex, and because “hazard tree” data does not currently exist for Eversource, we investigate whether land cover data surrounding the overhead lines can be used as a surrogate for grid cells that may have high amounts of “hazard trees”. Land cover data aggregated by grid cell has previously shown to help generalize models where utility-specific distribution infrastructure data is not available without significantly affecting model performance (Quiring et al. 2011).

Thirty-meter resolution land cover data was attained from the University of Connecticut Center for Land Use Education and Research (CLEAR). The 2006 Landsat satellite imagery was processed by CLEAR into various land cover categories (University of Connecticut 2006) of which coniferous forest, deciduous forest and developed categories were included in the damage models. To determine the land cover categories around the overhead lines, the overhead lines were first overlaid with the land cover data. Given that the resolution of the land cover data was 30 m, a point was placed uniformly every 30 m on the overhead lines shapefile and spatially joined to the land cover data. The counts of points per land cover category were aggregated for every 2 km grid cell, and the total counts of points per category were then divided by the total number of points in the grid cell to calculate the percentage of land cover category that surrounded the power lines in each grid cell. Initially, there was an overwhelming abundance of developed land cover (> 66%, Table 4) when the count of points was summed per grid cell. We suspected that roadways might be interfering with our land cover analysis: a typical two lane road with two shoulders is approximately 48 ft (16 m) (Stein and Neuman 2007) and thus may constitute >50% of a grid cell. To counteract this phenomenon, a 60 m buffer was drawn around the overhead lines and points were uniformly placed every 30 m. Table 4 provides a comparison of service-territory percentages of land cover categories by using different aggregation methods. Our analysis shows that overhead lines were mostly located along deciduous forest and developed areas, and were least likely to be located near wetland areas. Additionally, Figure 4 shows the classification for “developed” land cover around overhead lines, which is most concentrated in central and coastal Connecticut.

3. Models

3.1 Overview

Three decision tree models (decision tree, random forest, boosted gradient tree), with full and reduced datasets, and an ensemble decision tree that uses as input the predictions from the three decision tree models, were evaluated to determine which combination of method and data would yield the best damage predictions on the Eversource electric distribution network in Connecticut. The reduced subset consisted of only the weather variables, while the full model consisted of the weather variables along with infrastructure and land cover variables (Table 5). Models ending with an "A" subscript refer to models that use the reduced set of variables (i.e. "Model DT_A" is a decision tree model using the reduced dataset), and models with a "B" subscript refer to models that use the full set of variables (i.e. "Model DT_B" is a decision tree model using the full dataset). Although variable importance is interesting and has been investigated by other papers (Davidson et al. 2003, Nateghi et al. 2014a), our focus is the predictive accuracy of the models, so we will not include a section on variable importance.

3.2 Decision Tree Regression (DT)

The decision tree regression (DT) model, as described by Breiman et al. (1984), was the simplest model evaluated in this study and was selected because it is among the easiest of models to interpret and apply. A decision tree is a collection of logical "if-then" statements (called "branches") that relates explanatory variables (i.e. wind gust, wind duration above a threshold, etc.) to a response variable (i.e. outages) by recursively partitioning the explanatory variables into bins (called "leaves") that minimize the sum of square error (SSE). Recursive partitioning can either be an interactive process with the analyst selecting which splits should occur, or an automatic process that uses a stopping criterion (i.e. a node reaching purity (SSE = 0) or a decrease in the validation R^2) to grow the tree. Although not required, pruning can improve the robustness of a decision tree model by removing extraneous leaves.

3.3 Random Forest Regression (RF)

Random forest regression (RF), also described by Breiman (2001), is an extension of the decision tree model that tends to yield more robust predictions by stretching the use of the training data partition. Whereas a decision tree makes a single pass through the data, a random forest regression bootstraps 50% of the data (with replacement) and builds many trees (as specified by the analyst). Rather than using all explanatory variables as candidates for splitting, a random subset of candidate variables are used for splitting, which allows for trees that have completely different data and different variables (hence the term random). The prediction from the trees, collectively referred to as the "forest", are then averaged together to produce the final prediction. One hundred trees were included in our random forest model, with six terms sampled per split, a minimum of ten splits per tree, and a minimum split size of 256.

3.4 Boosted Gradient Tree Regression (BT)

Boosted gradient tree regression (BT), a common model used in ecology (Kint et al. 2012) and in business analytics (Pittman et al. 2009), is a set of large additive decision trees built by building a series of small trees on the residuals of the previous trees (SAS Institute Inc. 2013). The small trees, also known as "decision stumps" because of their limited depth (e.g. splits per tree), are considered "weak learners". While the first small trees are not very useful, or interesting on their own, the collection of small trees built on residuals of the previous small trees that can become a sophisticated predictive model. As more layers are added to the tree, the contribution from each small tree is regulated via a "learning rate". As the depth of the tree increases, the sum of predictions becomes more accurate while the additive tree becomes increasingly complex. Our boosted gradient tree was initialized with a learning rate of 0.1, fifty layers and three splits per tree.

3.5 Ensemble Decision Tree Regression (ENS)

Lastly, an ensemble decision tree regression (ENS) was investigated to determine if the predictions from the decision tree, random forest and boosted gradient tree regression could be used to predict storm damages better than the simple average of all models or any model alone. The ensemble decision tree can be likened to asking three different people what they expect the damage to be from a storm, and to then fit a model based on their predictions (one method may better predict extreme damage; and a separate method may better predict low or no damage); any number of these scenarios can be accounted for in the framework of the ensemble decision tree regression.

3.6 Model Validation

3.6.1 Repeated Random Holdout

Model validation on out-of-sample data is used to test the predictive accuracy of the model, and as such, only model validation results will be presented in this paper. There are many ways to look at model validation (repeated holdout, k-fold, stratified k-fold, leave-one-storm out validation). A 10-times repeated random holdout was conducted using 2/3 of the data as training and 1/3 of the data as validation. One drawback of the repeated holdout exercise is that some data may be used for validation exclusively while other data are used only for model training. We completed an analysis (not shown here) and found that more than 97% of observations were used for model validation at least once, and of those 97% of observations, each was used an average of 3.09 times (std. dev = 1.38). Given the small number of covariates (26) relative to the large data record size (>250,000 records), the large size of the validation partition (33% relative to the 10% or 20% used in other studies), and the overall coverage of available observations (97% observations were used on average 3 times for validation), we believe this represents a fair validation of our models.

We conducted our repeated random holdout as follows: of the >250,000 records in our database (2,851 grid cells for each of the 89 storm events), 2/3 of the data was used for training and 1/3 was used for validation. For fair comparison, the same training and validation partitions were used to evaluate each of the eight model combinations. Below we discuss the different models used in this study. Each of the eight models was built on the training data and used to predict the holdout data which was used for validation. The error metrics were calculated for each model in the validation partition, then the training and validation were recombined and the random holdout process was repeated a total of 10 times.

3.6.2 Definition of Accuracy Metrics

Outage predictions are aimed to inform emergency preparedness about the 1) total number of storm outages upon which a utility can decide on the number of crews needed to repair damages and 2) the spatial distribution of those outages so that they know where to place crews before a storm. To evaluate the model's predictive accuracy relative to these utility-specific needs, we opted to decouple the magnitude (count of outages) and spatial (distribution of outages) evaluations of each model. We next present two subsets of metrics to explore the magnitude and spatial accuracy of the trouble spot predictions to compare the eight different models we evaluated separately.

The absolute error (AE) per storm measures the accuracy of the predictions aggregated by storm. It was calculated by taking the absolute value of the difference between the actual (θ) and predicted (θ_o) predicted outages per storm (Equation 1).

$$AE = |\theta - \theta_o| \qquad \text{(Eqn. 1)}$$

Similarly, the percentage error (PE, Equation 2) per storm per holdout sample is calculated by dividing the absolute error by the corresponding actual outages per storm. AE and APE values that are 0 are perfect, while anything greater than 0 is considered less than perfect.

$$APE = \frac{|\theta - \theta_o|}{\theta} \qquad \text{(Eqn. 2)}$$

The four metrics calculated based on the above error definitions include: i) mean absolute error (MAE), ii) median absolute error (MdAE), iii) mean absolute percentage error (MAPE), iv) median absolute percentage error (MdAPE). The mean and median AE or PE of each model can be calculated by taking the mean (median) of the distribution of AE or PE across all holdout samples, respectively (89 storms times 10 holdout samples equals 890 values for mean and median to be calculated).

It's worth noting that most outage models in the literature use MAE per grid cell as the metric to evaluate model performance – given that our storms represent a variety of sizes and severities (from 20 to 15,000 outages), we consider it appropriate to present error metrics by storm rather than by grid cell. To compare our models to other hurricane outage models in the literature, we will present MAE per grid cell to evaluate Storm Irene and Hurricane Sandy in Section 5.

To evaluate the spatial accuracy of the predicted outages we calculated the proportion of actual outages by town. For each storm and model, the actual outages per town were divided by the total actual outages across the service territory per storm. Additionally, we created corresponding proportions for the predicted outages. We then calculated the Spearman's rank correlation coefficient, r_s, between these proportions for each holdout sample (resulting in 10 unique r_s values per model, which we presented as boxplots in Section 4.1.2). To ensure that correlation was not a scale dependent phenomenon, we also created proportions for AWCs and calculated Spearman's rank correlation coefficient for each holdout sample. Although our model predicts outages per 2-km grid cell, towns and AWCs are natural aggregations for correlation because these are the geographic units by which Eversource allocates crews and resources. As mentioned earlier, Eversource is divided into 149 towns which are grouped into 18 AWCs (note that not all AWCs have the same number of towns or geographic boundary). We expect the r_s to be improved for AWCs over towns because of the aggregation.

Spearman's correlation is a nonparametric test for determining the strength of the relationship between variables and is more resilient to outliers than Pearson correlation (Wilks 2011); Spearman's correlation is the Pearson correlation computed using the ranks of the data. The two assumptions required for Spearman correlation are 1) variables are measured on ordinal, interval or ratio scale, and 2) a monotonic relationship between the variables. We chose to use Spearman instead of Pearson because the distribution of proportion of actual outages per town and AWC was skewed right, whereas the distribution of the predicted proportion of outages was normally distributed. Spearman's rank correlation coefficients that are close to 1 have a strong positive relationship (though not necessarily linear), values close to 0 have no relationship, and values close to -1 have a strong negative relationship (though not necessarily linear).

In addition to the spatial correlation error metric, we used maps to qualitatively compare the spatial accuracy of model predictions to actual outages for two of the largest and most impactful events in our database (Irene and Sandy, Section 5). We will also present maps of the MAE and MdAE per town of our best overall model, Model ENS_B, in Section 4 (computed as the mean and median of the AE per town from all 10 holdout samples).

4. Results

4.1 Model Validation Results

4.1.1 Magnitude Results

In this section, we will present the storm-wise results from each holdout (i.e. 89 storms times 10 holdouts equals 890 validation points were used to create each graph). Figure 5 shows boxplots of the absolute error and percentage error per storm for all holdouts. Note that diamond symbols on Figure 5 represent the mean absolute error (MAE) and mean absolute percentage error (MAPE), and the thick black lines represents the median absolute error (MdAE) and median percentage error (MdAPE), respectively. The MAPE values are skewed for all models due to over-prediction of smaller storm events. For example, a storm with 20 actual outages can be off by 500% if 100 outages are predicted for that storm. In addition, the MAE values are skewed for all models due to the errors from predicting the largest storm events (hurricanes, which can be up to two orders of magnitude larger than other events in our database).

Though the meaning of MdAE and MdAPE is different than MAE or MAPE, we believe the MdAE and MdAPE are better metrics for model evaluation than MAE and MAPE because the median is a good representation of the center of the distribution. Table 6 shows MdAE and MdAPE for each model by season. Cold weather storms (storms occurring between December and March) had both lowest MdAE and MdAPE values, and transition storms tended to have slightly improved MdAPE than warm weather storms, and had similar MdAE values. Though beyond the scope of this paper, we believe that cold weather storms might be easier to predict than warm weather or transition storms because the trees have lost their leaves and the soil is generally frozen during these months, so most damage is associated with wind effects. Model ENS_A had the lowest MdAE, MdAPE, MAE, and MAPE values, and so we can say with respect to magnitude that it was the best performing model. Model ENS_B had a similar, slightly less improved performance than Model ENS_A; it also had a slightly wider interquartile range (IQR) and higher MAE, MAPE, MdAE and MdAPE values. If desired, the first and third quantiles (Q1, Q3) of the AE and PE can be read from Figure 5.

4.1.2 Spatial Accuracy Results

In this section, we will present the r_s values for all towns or AWCs for each holdout (recall that each boxplot in Figure 6 was constructed from 10 r_s values, one for each holdout sample). Figure 6 shows the values for r_s for each model and holdout sample for both towns and AWCs. As mentioned earlier, we prefer to use proportions rather than actual values in order to evaluate the accuracy of the model to predict the spatial distribution of outages (even if the territory-total predicted number of outages is over or underestimated). The range of r_s values for the different models was between 0.2 and 0.5 (p-value <0.001), which indicates a weak positive correlation between observed and predicted proportions of outages in each town and AWC for each model. As expected, r_s increased for each model when aggregating from towns to AWCs. The mean value of r_s across all holdout samples is close to the median (Figure 6). These r_s values can be interpreted as follows: when the proportion of predicted outages increases, so does the actual proportion of outages, which implies that there is spatial accuracy (albeit, weak spatial accuracy overall). The model with the best distribution of r_s values was Model ENS_B, which had the highest values for both town and AWC spatial scales of aggregation. Interestingly, Model RF_B had similar spatial correlation to Model ENS_B, which may imply that Model RF_B may also be used to forecast the spatial distribution of damages; however this model was shown to exhibit stronger biases in the total magnitude of outages.

4.3 Selecting the best overall model

In an operational context, we believe that all models should be considered to represent the range of possible outage scenarios. However, utilities need to select the most likely scenario during decision-making, which will be based on the performance of the models in terms of both magnitude and spatial accuracy metrics. We believe that Model ENS_B was the best model in the overall evaluation because of its combination of spatial accuracy and magnitude metrics. For brevity, we will only explore the magnitude error metrics for Model ENS_B as we have already discussed that all models have weak positive spatial correlation. Figure 7 shows the actual vs. predicted outages per storm by season; Figure 8 shows how storm percent error decreases and absolute error increases as a function of storm severity. In order to show where the model tends to have the most error, we also present the MAE and MdAE per town (Figure 9). Note how the model tended to have the highest MAE and MdAE in areas that have the highest population (central and coastal Connecticut) and highest "developed" land cover around overhead lines (Figure 4); less populated areas tended to have less MAE and MdAE per town, which may be a function of having less customers, less isolating devices or better vegetation management practices. However, further research is needed to understand why some areas are more resilient than other areas.

5. Predicting Outages for Hurricanes

Hurricanes are among the most costly, disruptive and serious of all storm events to impact electric distribution networks. Though significant, hurricanes represented only two of the 89 storms in our database. If the damage before a hurricane could be accurately forecasted, then emergency-preparedness efforts could be vastly improved by deploying restoration crews and supplies ahead of a hurricane's land fall. We next compare the simplest model (DT_A) with our most sophisticated model (ENS_B) to show how model and data forcing complexity might influence hurricane outage predictions.

Figure 10 shows the distribution of average outages per town across all 10 holdout samples of Irene and Sandy for actual data vs. model DT_A and model ENS_B outage predictions. On average, Irene predictions were underestimated 10.2% by model DT_A (463 outages) and 11% by model ENS_B (498 outages). On average, Sandy was underestimated 1.7% (80 outages) by model DT_A and overestimated 4.4% by model ENS_B (212 outages). While both models are shown to accurately predict the aggregate total number of outages for the two hurricane storm events, model ENS_B was superior in predicting the spatial distribution of these outages, especially for the towns that were hardest hit (Figure 10). As a complement to the actual and predicted outage maps, Figure 11 shows a scatterplot of the holdout averages for comparison of actual and predicted values from the average of all holdout samples for Irene and Sandy. The improved spatial accuracy of model ENS_B renders the model useful as an input for other related models such as a storm restoration duration model or customer outage model because it correctly predicts the worst hit areas.

6. Discussion

In this section, we will address the investigative questions we asked at the beginning of this paper. Regarding weather-only outage models and the value of added utility-specific data, we conclude that a reasonably performing outage model to predict outages for the entire service territory can be developed without additional utility data so long as actual outage locations are available for historic events. Similar to (Han et al. 2009b), our model tended to overestimate in the most populated (urban) areas; and similar to Nateghi et al. (2014a), our simpler weather-only model (ENS_A) exhibited similar (even slightly better) magnitude metrics to the corresponding utility-specific model (ENS_B). Consequently, there is an opportunity to readily expand the models presented in this study to other utilities that are within the inner 2-km weather simulations domain (Figure 2), which includes utilities in Massachusetts, New Hampshire, New York, New Jersey and Rhode Island, so long as historic outage data are available for the simulated storm events. An additional benefit of our models is that they can be used for different storm types from different seasons, which is represented by the error metrics determined for many different storms. It was shown that the models predicted best the cold weather storm events, which may attributed to trees having lost their leaves and frozen soil, so most damage is caused by wind. Although cold weather storms were predicted best, recall that our paper excludes ice storms which are among the most damaging storms for electric distribution networks (Liu et al. 2008). The higher MdAE and MdAPE values were shown for the convective warm weather storms, which are the most difficult to predict with numerical weather models due to their short timescale and localized nature.

These “weather-only “outage models can be valuable tools for utilities in the short term that can be used until data becomes available to build more mature models. The limitation of weather-only outage model is that they cannot account for dynamic conditions of the distribution network; such as system hardening improvements (Han et al. 2014), the topology of the network (Winkler et al. 2010), or vegetation management (Guikema et al. 2006b, Nateghi et al. 2014a, Radmer et al. 2002). The benefit of the added utility data was that it had higher spatial accuracy than the weather-only models. Our storms exhibited weak positive correlation between actual and predicted proportion of outages. Model ENS_B had a mean correlation (r_s = 0.37 for towns, r_s = 0.45 for AWCs) that is similar to other correlation values found in the literature (Angalakudati et al. 2014), though they compute correlation for actual vs. predicted outages (not proportions) at “platforms”, which are similar to Eversource’s AWCs.

We now compare model ENS_B to other existing models in the literature. Given that most of the literature focuses on hurricanes, so follows our discussion. Many hurricane outage papers have reported the MAE per grid cell across all storms evaluated, which we calculated from the grid cell predictions from all 10 holdout samples for Irene and Sandy. For model ENS_B the MAE per grid cell was 3.13 outages (std. dev. = 4.4 outages) for Irene and 3.15 outages (std. dev. = 4.9 outages) for Sandy. These are comparable error magnitudes to those presented in other papers - Nateghi et al. (2014a) used a random forest model and reported MAE per grid cell values between 0.26 and 2.24 depending on which State they modeled; Han (2009a) used generalized additive models and reported MAE per grid cell values between <0.001 – 72 outages depending on the hurricane that was predicted (however, MAE per grid cell values were tabulated as function of the actual number of outages, which made it difficult to do a direct comparison). With respect to storm totals, our outage models predicted Storm Irene within 5% and Hurricane Sandy within 11%, which is similar to other hurricane outage models (Winkler et al. 2010). However, direct comparison of our results must be taken with caution: our northeastern US service territory has different environmental and infrastructure attributes than Gulf Coast utilities, and we are not using the same storms for evaluation (our average size hurricane caused >15,000 outages compared to the Han et al. (2009a) data, in which the average hurricane caused 6,169 outages).

7. Conclusions

We have investigated the performance of four types of models with two different subsets of data to determine what combination of data and method yields the best prediction of outages to Eversource's electric distribution network in Connecticut. Of the eight models evaluated, an ensemble decision tree regression (ENS_B) forced with predictions from decision tree, random forest and boosted gradient tree regressions proved to be the best model overall. The ensemble decision tree regression modeling framework could be implemented operationally to predict future weather-related threats to the distribution system (as well as other types of critical infrastructure such as water or gas distribution systems). Now that outages can be forecasted in anticipation of a storm event, other models could be built from our predictions such as a customer outage model or an outage duration model. Should data become available, this modeling framework lends itself to the inclusion of vegetation management (e.g. tree trimming) and "hazardous tree" data. Further, other utility-specific data, such as conductor material and circuit type (backbone or lateral), may prove important to future models. Although all electric distribution networks are relatively unique (i.e. each utility has different topology, different tree species that interact with overhead lines and different vegetation management strategies), we believe this model can be applied elsewhere as long as the necessary outage data is available.

Acknowledgements: The authors would like to gratefully acknowledge the support of Eversource Energy for the data and funding that sponsored this research. Many thanks to those who contributed their talents to this effort: William J. Burley, Xinxuan Zhang, Jichao He, Jaemo Yang, Dominic Scerbo, Tingyang Xu and Zachary Goulet. We also acknowledge and appreciate valuable input during the course of this study by Dean Desautels, Thomas Layton, Ted Crisuculo, Ray Litwin, Mike Zappone, Peter Clark, Kenneth Bowes and Camilo Serna. In addition, we are grateful to two anonymous reviewers for their thoughtful comments and useful feedback which have helped us contribute a better paper to the outage modeling and emergency preparedness communities.

References

1. Angalakudati M, Calzada J, Farias V, Gonynor J, Monsch M, Papush A, Perakis G, Raad N, Schein J, Warren C, Whipple S, Williams J (2014) Improving emergency storm planning using machine learning. T&D Conference and Exposition, 2014 IEEE PES. doi: 10.1109/TDC.2014.6863406

2. Astitha M, Kallos G, Spyrou C, O'Hirok W, Lelieveld J, Denier Van Der Gon HAC (2010) Modelling the chemically aged and mixed aerosols over the eastern central Atlantic Ocean-potential impacts. Atmos Chem Phys. doi: 10.5194/acp-10-5797-2010

3. Breiman L (2001) Random forests. Mach Learn. doi: 10.1023/A:1010933404324

4. Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) . Classification and Regression Trees

5. Builtjes P (2005) Review of Methods for Assessing Air Quality at Regional/Continental Scale. Air4EU 6th Framework Programme Deliverable D5.1

6. Cerruti BJ, Decker SG (2012) A statistical forecast model of weather-related damage to a major electric utility. Journal of Applied Meteorology and Climatology 51:191-204

7. Chang JC, Hanna SR (2004) Air quality model performance evaluation. Meteorol Atmos Phys. doi: 10.1007/s00703-003-0070-7

8. Chou M, Suarez MJ (1994) An efficient thermal infrared radidation parameterization for use in general circulation models. NASA Technical Memo:84

9. Connecticut Light & Power (2014) Transmission and Distribution Reliability Performance Report

10. Davidson RA, Liu H, Sarpong K, Sparks P, Rosowsky DV (2003) Electric power distribution system performance in Carolina Hurricanes. Nat Hazards Rev. doi: 10.1061/(ASCE)1527-6988(2003)4:1(36)

11. Davies Consulting (2012) Final Report: Connecticut Light and Power's Emergency Preparedness and Response to Storm Irene and the October Nor’easter

12. Fanelli C, Fanelli P, Wolcott D (2013) Hurricane Sandy 2012 Water Level and Meteorological Data Report
. NOAA

13. Grell GA, Devenyi D (2002) A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. Geophysical Research Letters 29:1693

14. Guikema S, Han S-, Quiring S (2008) Estimating power outages during hurricanes using semi-parametric statistical methods. Proc Struct Congr - Struct Congr : Crossing Borders

15. Guikema SD, Nateghi R, Quiring SM (2014a) Storm power outage prediction modeling. Saf , Reliab Risk Anal : Beyond Horiz - Proc Eur Saf Reliab Conf , ESREL

16. Guikema SD, Nateghi R, Quiring SM, Staid A, Reilly AC, Gao M (2014b) Predicting Hurricane Power Outages to Support Storm Response Planning. Access, IEEE. doi: 10.1109/ACCESS.2014.2365716

17. Guikema SD, Quiring SM (2012) Hybrid data mining-regression for infrastructure risk assessment based on zero-inflated data. Reliab Eng Syst Saf. doi: 10.1016/j.ress.2011.10.012

18. Guikema SD, Quiring SM, Han S- (2010) Prestorm Estimation of Hurricane Damage to Electric Power Distribution Systems. Risk Anal. doi: 10.1111/j.1539-6924.2010.01510.x

19. Guikema SD, Coffelt JP (2009) Practical considerations in statistical modeling of count data for infrastructure systems. J Infrastruct Syst. doi: 10.1061/(ASCE)1076-0342(2009)15:3(172)

20. Guikema SD, Davidson RA (2006) Modelling critical infrastructure reliability with generalized linear (mixed) models. Proc Int Conf Probabilistic Saf Assess Manage , PSAM

21. Guikema SD, Davidson RA, Liu H (2006a) Statistical models of the effects of tree trimming on power system outages. IEEE Trans Power Delivery. doi: 10.1109/TPWRD.2005.860238

22. Guikema SD, Davidson RA, Liu H (2006b) Statistical models of the effects of tree trimming on power system outages. IEEE Trans Power Delivery. doi: 10.1109/TPWRD.2005.860238

23. Han S-, Rosowsky D, Guikema S (2014) Integrating Models and Data to Estimate the Structural Reliability of Utility Poles During Hurricanes. Risk Anal. doi: 10.1111/risa.12102

24. Han S-, Guikema SD, Quiring SM (2009a) Improving the predictive accuracy of hurricane power outage forecasts using generalized additive models. Risk Analysis

25. Han S-, Guikema SD, Quiring SM, Lee K-, Rosowsky D, Davidson RA (2009b) Estimating the spatial distribution of power outages during hurricanes in the Gulf coast region. Reliab Eng Syst Saf. doi: 10.1016/j.ress.2008.02.018

26. Hendrick EM, Tino VR, Hanna SR, Egan BA (2013) Evaluation of NO2 predictions by the plume volume molar ratio method (PVMRM) and ozone limiting method (OLM) in AERMOD using new field observations. J Air Waste Manage Assoc. doi: 10.1080/10962247.2013.798599

27. Hongfei Li, Treinish LA, Hosking JRM (2010) A statistical model for risk management of electric outage forecasts. IBM Journal of Research and Development. doi: 10.1147/JRD.2010.2044836

28. Jimenez PA, Dudhia J (2012) Improving the representation of resolved and unresolved topographic effects on surface wind in the WRF model. Journal of Applied Meteorology and Climatology 51:300

29. Jiménez PA, González-Rouco JF, Montávez JP, Navarro J, García-Bustamante E, Valero F (2008) Surface wind regionalization in complex terrain. Journal of Applied Meteorology and Climatology

30. Kint V, Vansteenkiste D, Aertsen W, de Vos B, Bequet R, van Acker J, Muys B (2012) Forest structure and soil fertility determine internal stem morphology of Pedunculate oak: A modelling approach using boosted regression trees. Eur J For Res. doi: 10.1007/s10342-011-0535-z

31. Liu H, Davidson RA, Apanasovich TV (2008) Spatial generalized linear mixed models of electric power outages due to hurricanes and ice storms. Reliab Eng Syst Saf. doi: 10.1016/j.ress.2007.03.038

32. Liu H, Davidson RA, Apanasovich TV (2007) Statistical forecasting of electric power restoration times in hurricanes and ice storms. IEEE Trans Power Syst. doi: 10.1109/TPWRS.2007.907587

33. McGee J, Skiff J, Carozza P, Edelstein T, Hoffman L, Jackson S, McGrath R, McGroarty F, Osten C (2012) Report of the Two Storm Panel. State of Connecticut

34. Mensah AF, Duenas-Osorio L (2014) Outage predictions of electric power systems under Hurricane winds by Bayesian networks. Int Conf Probab Methods Appl Power Syst , PMAPS - Conf Proc. doi: 10.1109/PMAPS.2014.6960677

35. Mlawer EJ, Taubman SJ, Brown PD, Iacono MJ, Clough SA (1997) Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. Journal of Geophysical Research 102:16663

36. Nateghi R, Guikema S, Quiring SM (2014a) Power Outage Estimation for Tropical Cyclones: Improved Accuracy with Simpler Models. Risk Anal. doi: 10.1111/risa.12131

37. Nateghi R, Guikema SD, Quiring SM (2014b) Forecasting hurricane-induced power outage durations. Nat Hazards

38. Nateghi R, Guikema SD, Quiring SM (2011) Comparison and Validation of Statistical Methods for Predicting Power Outage Durations in the Event of Hurricanes. Risk Anal. doi: 10.1111/j.1539-6924.2011.01618.x

39. O'Neill D, Fijnvandraat C, DeSimone D (2013) Best Practices in Utility Emergency Management Docket 12-09-03

40. Pittman SJ, Costa BM, Battista TA (2009) Using lidar bathymetry and boosted regression trees to predict the diversity and abundance of fish and corals. J Coast Res. doi: 10.2112/SI53-004.1

41. Quiring SM, Zhu L, Guikema SD (2011) Importance of soil and elevation characteristics for modeling hurricane-induced power outages. Nat Hazards. doi: 10.1007/s11069-010-9672-9

42. Radeloff VC, Hammer RB, Stewart SI, Fried JS, Holcomb SS, McKeefry JF (2005) The wildland-urban interface in the United States. Ecol Appl

43. Radmer DT, Kuntz PA, Christie RD, Venkata SS, Fletcher RH (2002) Predicting vegetation-related failure rates for overhead distribution feeders. IEEE Trans Power Delivery. doi: 10.1109/TPWRD.2002.804006

44. SAS Institute Inc. (2013) JMP

45. Song-You Hong, Noh Y, Dudhia J (2006) A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes. Mon Weather Rev

46. Staid A, Guikema SD, Nateghi R, Quiring SM, Gao MZ (2014) Simulation of tropical cyclone impacts to the U.S. power system under climate change scenarios. Clim Change. doi: 10.1007/s10584-014-1272-3

47. Stein WJ, Neuman TR (2007) Mitigation Strategies for Design Exceptions FHWA-SA-07-011

48. Tewari M, Chen F, Wang W, Dudhia J, LeMone MA, Mitchell K, Ek M, Gayno G, Wegiel J, Cuenca RH (2004) Implementation and verification of the unified NOAH land surface model in the WRF model:11

49. Thompson G, Field PR, Rasmussen RM, Hall WD (2008) Explicit Forecasts of Winter Precipitation Using an Improved Bulk Microphysics Scheme. Part II: Implementation of a New Snow Parameterization. Mon Weather Rev. doi: 10.1175/2008MWR2387.1

50. University of Connecticut (2006) . Connecticut's Changing Landscape Land Cover

51. Wagner BC, Beach MH, Brooks JP, Clark L, Fernandez BR, Hall KL, Sherman R (2012) Environmental Quality in Connecticut 2011: Part I

52. Wallace JM, Hobbs PV (2006) 10 - Climate Dynamics. In: Wallace JM, Hobbs PV (eds) Atmospheric Science (Second Edition). Academic Press, San Diego, pp 419-465

53. Wilks D (2011) Statistical Methods in the Atmospheric Sciences - 3rd Edition. Academic Press, Oxford, UK

54. Winkler J, Dueñas-Osorio L, Stein R, Subramanian D (2010) Performance assessment of topologically diverse power systems subjected to hurricane events. Reliab Eng Syst Saf. doi: 10.1016/j.ress.2009.11.002

55. Witt Associates (2011) Connecticut October 2011 Snowstorm Power Restoration Report

56. Zhang D-, Anthes RA (1982) A high-resolution model of the planetary boundary layer - sensitivity tests and comparisons with SESAME-79 data. Journal of Applied Meteorology 21:1594

List of Figures

Figure 1: Service territory of Eversource denoting the sum of protective devices (assets: transformers, fuses, reclosers, switches) per 2 km grid cell of the weather forecasting model. Areas in white denote locations not part of the service territory.

Figure 2: Nested grids of Weather Research and Forecasting (WRF) model used to simulate the storm events; the Eversource service territory is within the highest (2 km) model grid resolution.

Figure 3: Scatter plots of 10m (m s^{-1}) sustained wind speed (left panel) and corresponding max values (at each station location) from WRF simulations versus METAR observations. Upper panels: Hurricane Irene; middle panels: Hurricane Sandy; Lower panels: NEMO blizzard. The 45 degree line (1:1 linear relationship) is added in all plots.

Figure 4: Percentage of developed lines per grid cell using 60m buffer classification/aggregation.

Figure 5: Boxplots of absolute error (left) and percentage error per storm by model for all 10 holdouts. Diamonds represent mean values (MAE and MAPE) and bold horizontal lines indicate median values (MdAE or MdAPE) per model.

Figure 6: Boxplots of Spearman's rank-order correlation coefficient for each of the 10 holdout samples, computed by town (left) and AWC (right). Diamonds represent mean values and bold horizontal lines indicate median values.

Figure 7: Total number of actual vs. predicted outages over the validation grid cells for model ENS_B with seasonal grouping for all holdout samples.

Figure 8: Absolute (left) and percentage (right) error per storm of model ENSB for all 10 holdouts as a function of magnitude.

Figure 9: Mean absolute error (left) and median absolute error (right) per town of model ENS_{B} for all 10 holdouts as a function of magnitude.

Figure 10: Maps of actual vs. predicted outages per town for Irene and Sandy for models DT_{A} and ENS_{B}.

Figure 11: Scatterplots of actual vs. predicted outages per town for Storm Irene and Hurricane Sandy.

List of Tables

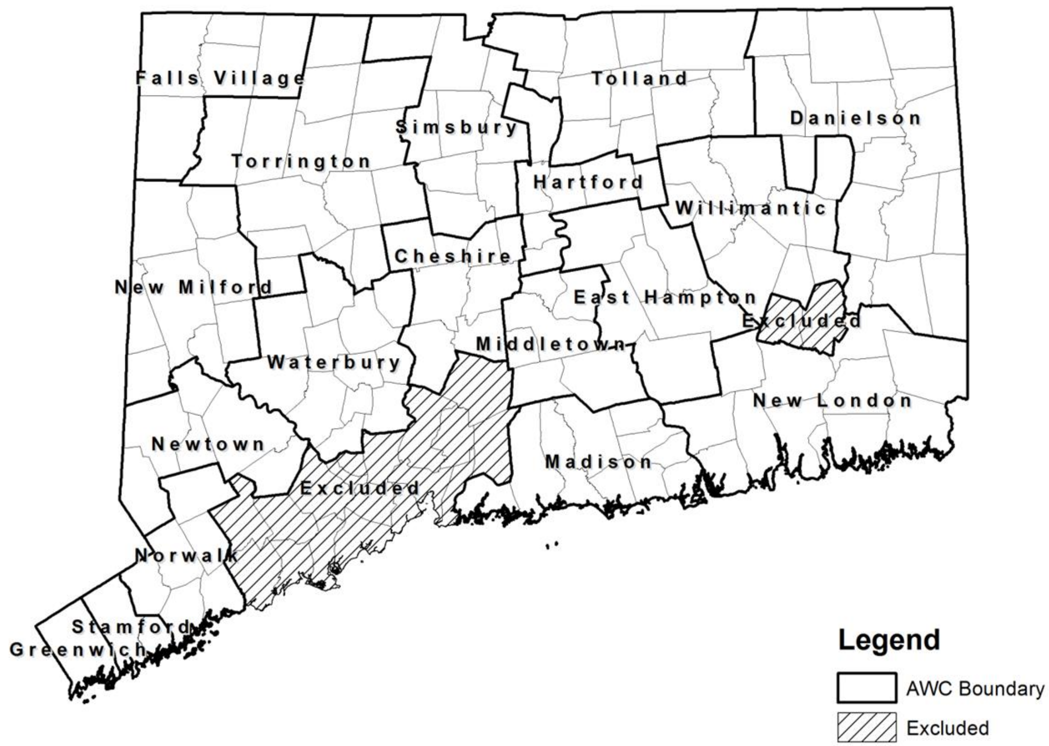

Figure 1: Service territory of Eversource denoting the town and area work center (AWC) boundaries.

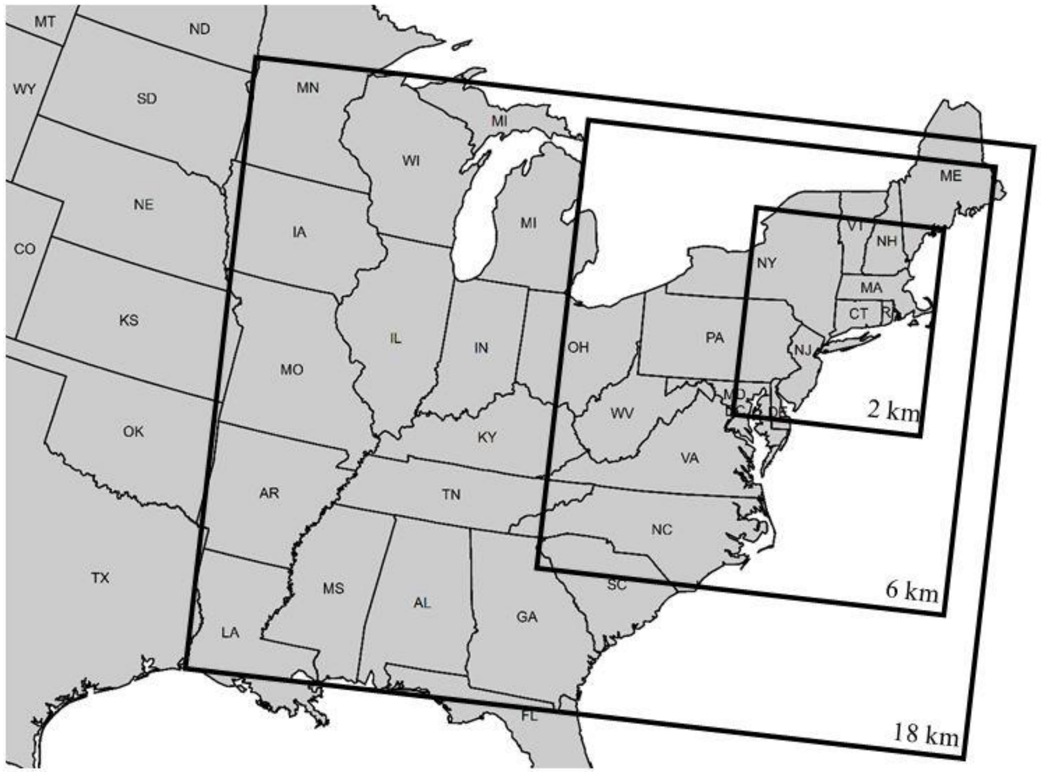

Figure 2: Nested grids of Weather Research and Forecasting model used to simulate the storm events; the Eversource service territory is within the highest (2 km) model grid resolution.

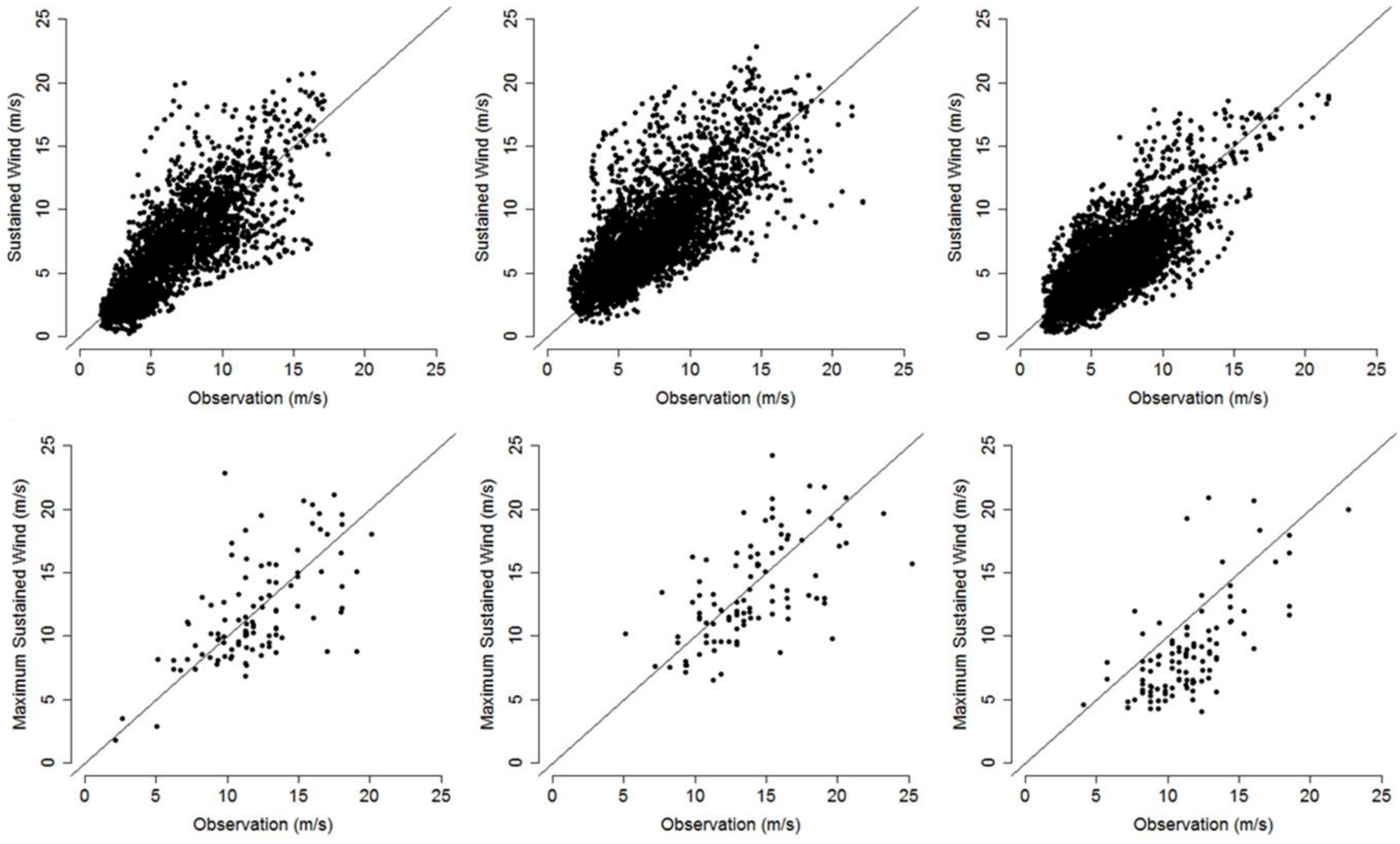

Figure 3: Scatter plots of 10m (m s^{-1}) sustained wind speed (top plot) and corresponding max values at each station location (bottom plot) from WRF simulations versus METAR observations. Left panels: Hurricane Irene; Middle panels: Hurricane Sandy; Right panels: NEMO blizzard. The 45 degree line (1:1 linear relationship) is added in all plots.

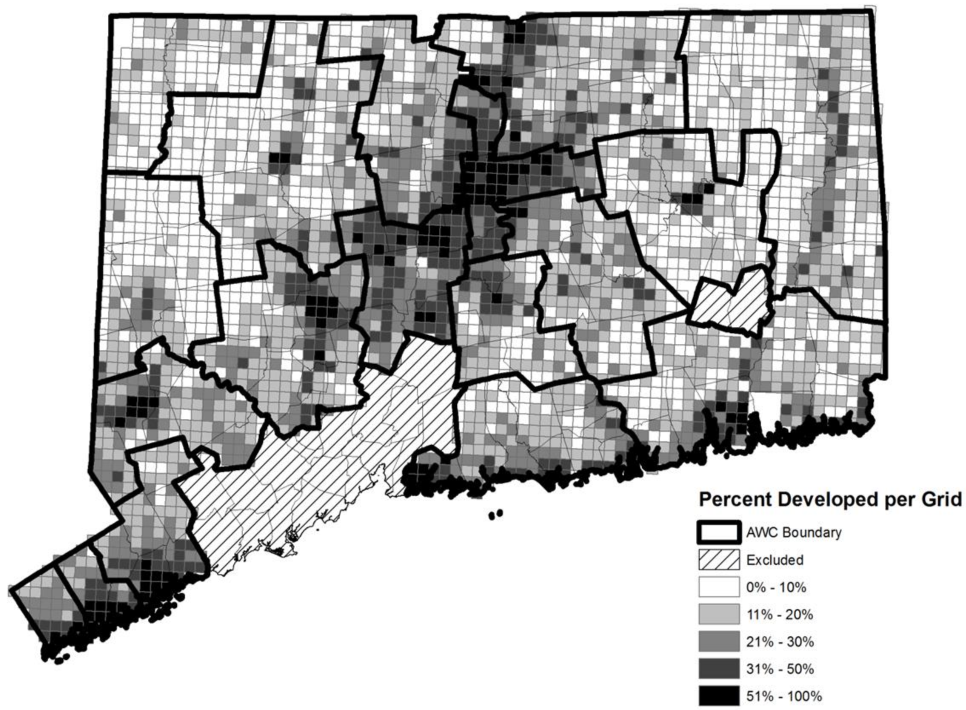

Figure 4: Percentage of developed lines per grid cell using 60m buffer classification/aggregation.

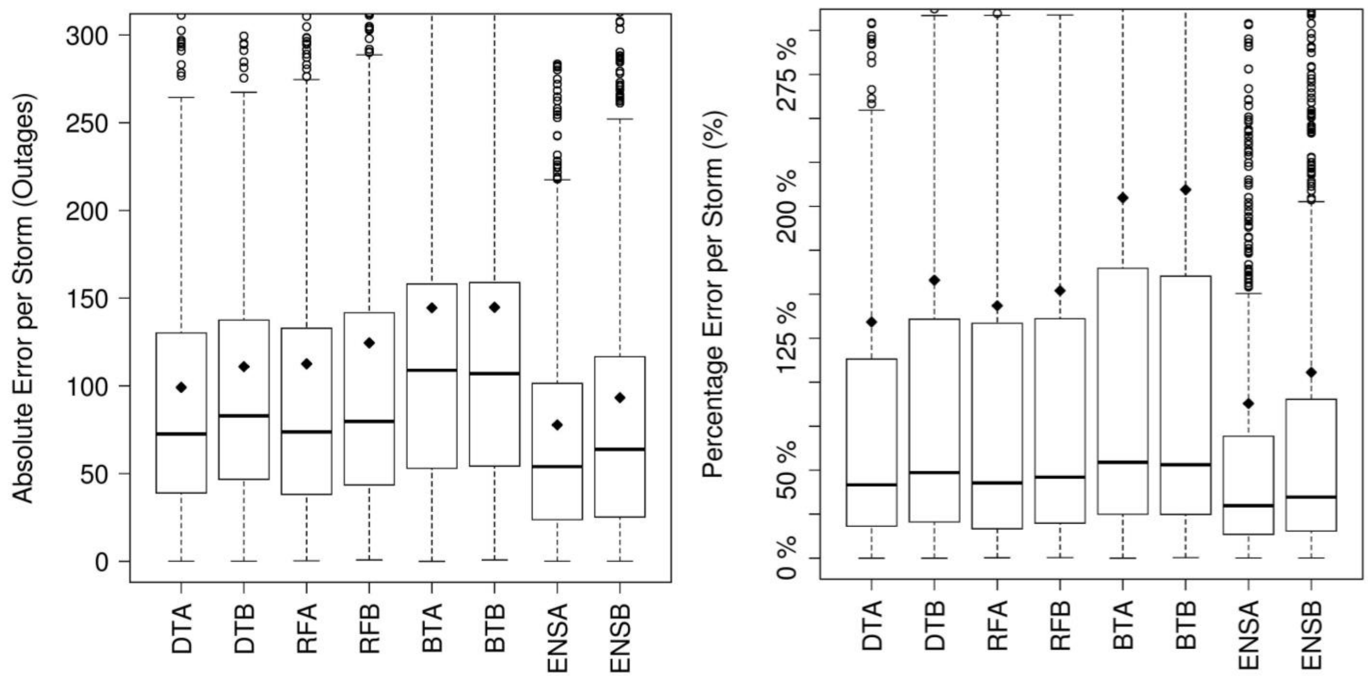

Figure 5: Boxplots of absolute error (left) and percentage error per storm by model for all 10 holdouts. Diamonds represent mean values (MAE and MAPE) and bold horizontal lines indicate median values (MdAE or MdAPE) per model.

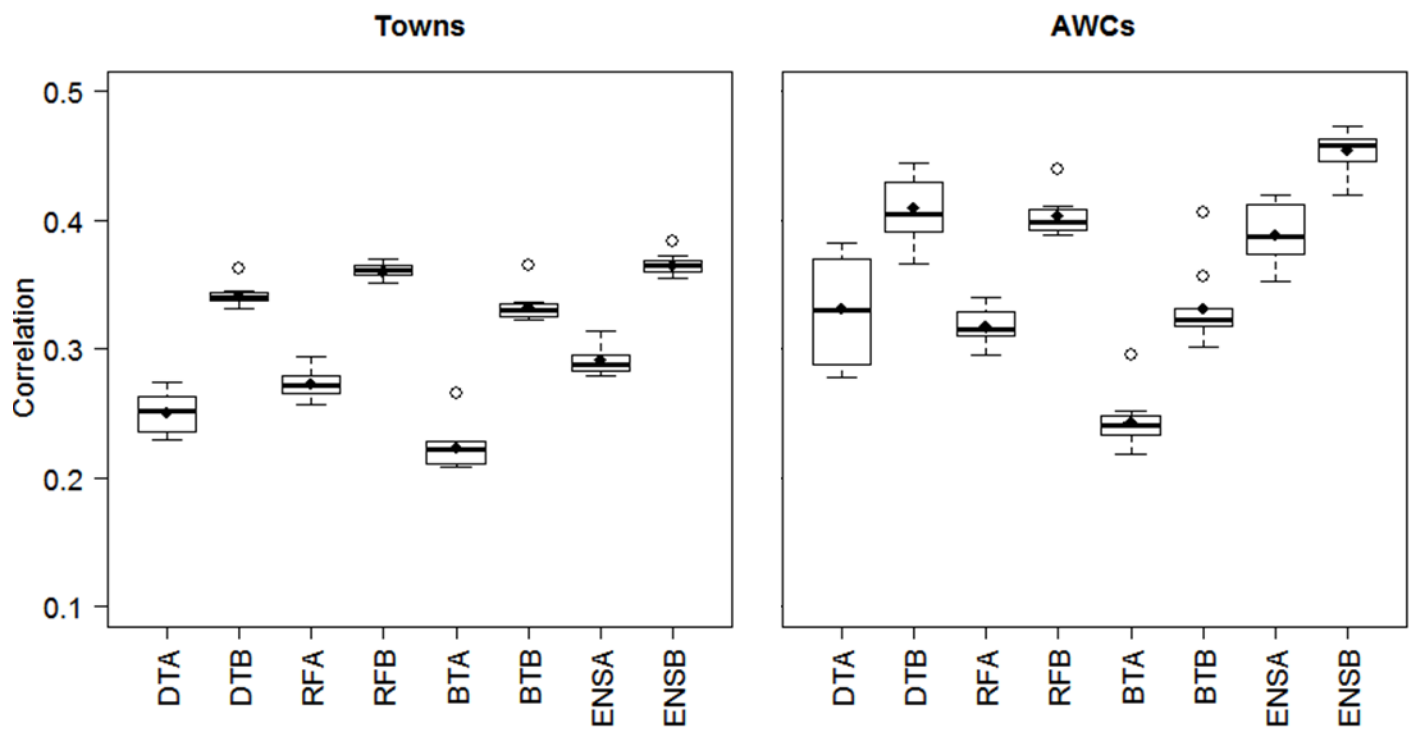

Figure 6: Boxplots of Spearman's rank-order correlation coefficient for each of the 10 holdout samples, computed by town (left) and AWC (right). Diamonds represent mean values and bold horizontal lines indicate median values.

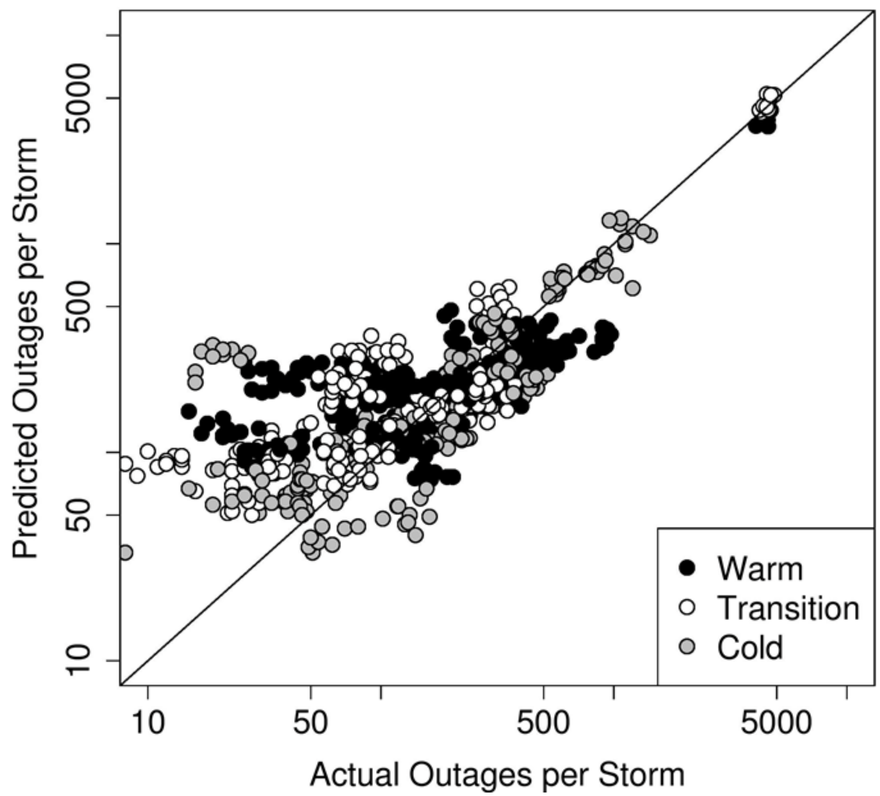

Figure 7: Total number of actual vs. predicted outages over the validation grid cells for model ENS_B with seasonal grouping for all holdout samples, with 45 degree line.

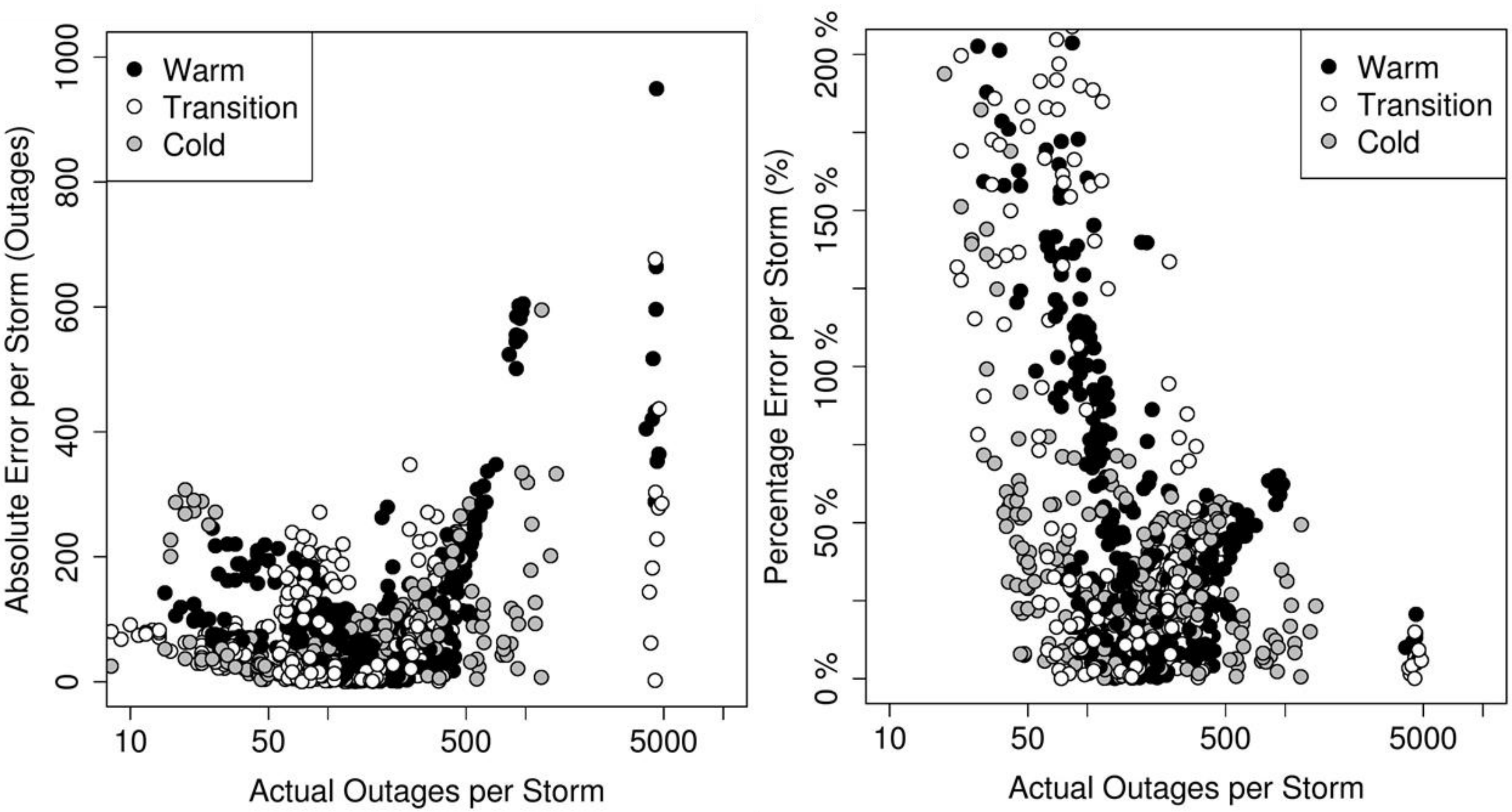

Figure 8: Absolute (left) and percentage (right) error per storm of model ENS_B for all 10 holdouts as a function of magnitude.

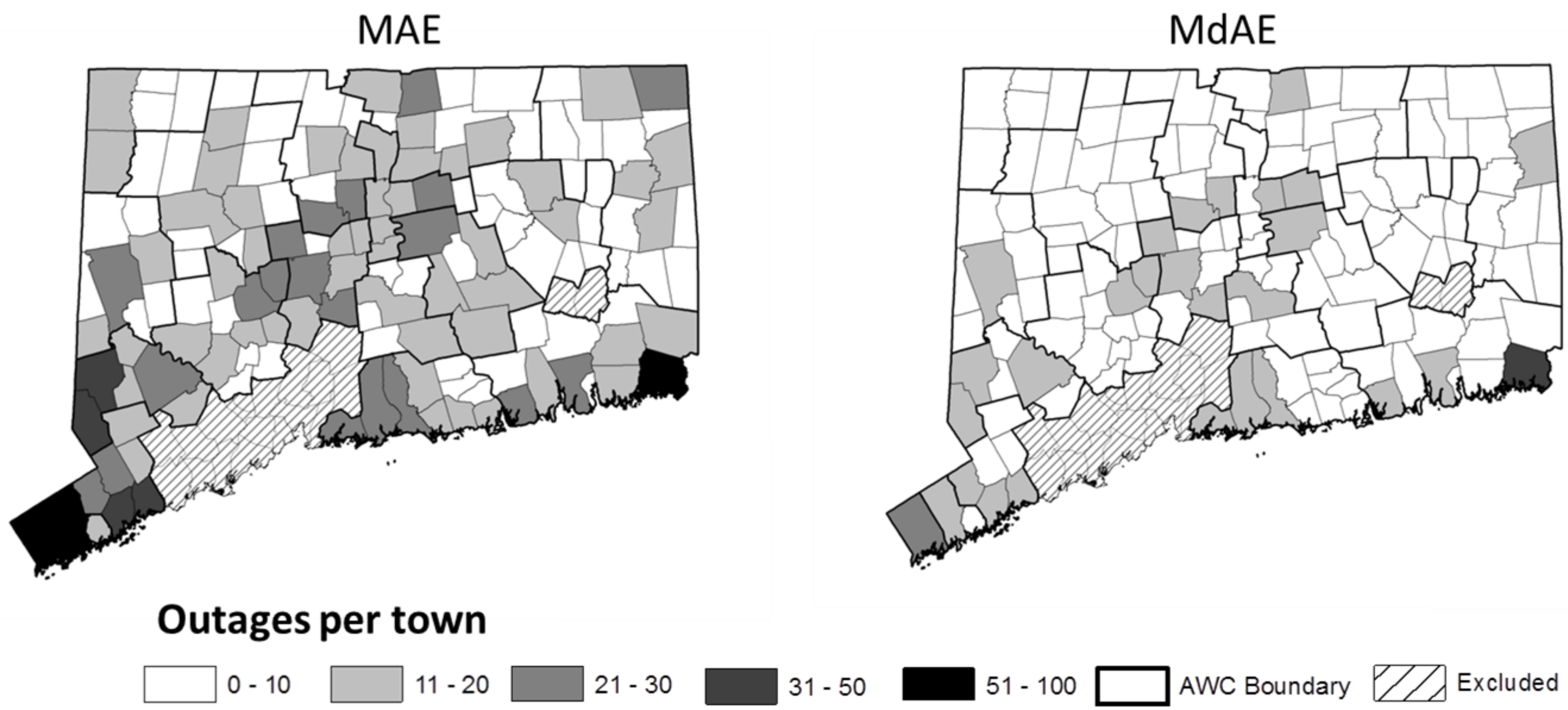

Figure 9: Mean absolute error (left) and median absolute error (right) per town of model ENS_B for all 10 holdouts as a function of magnitude.

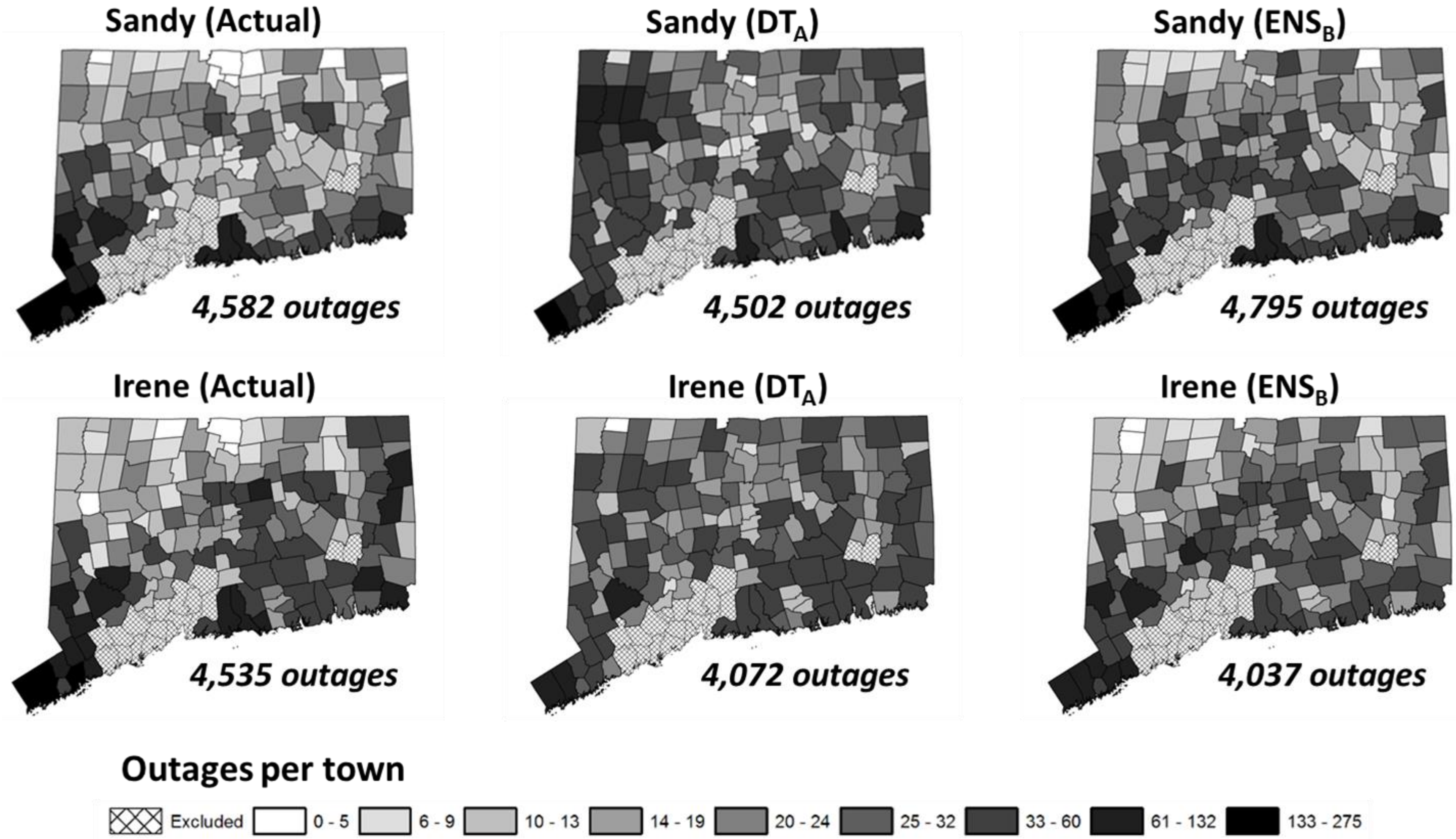

Figure 10: Maps of actual vs. predicted outages per town for Irene and Sandy for models DT$_A$ and ENS$_B$.

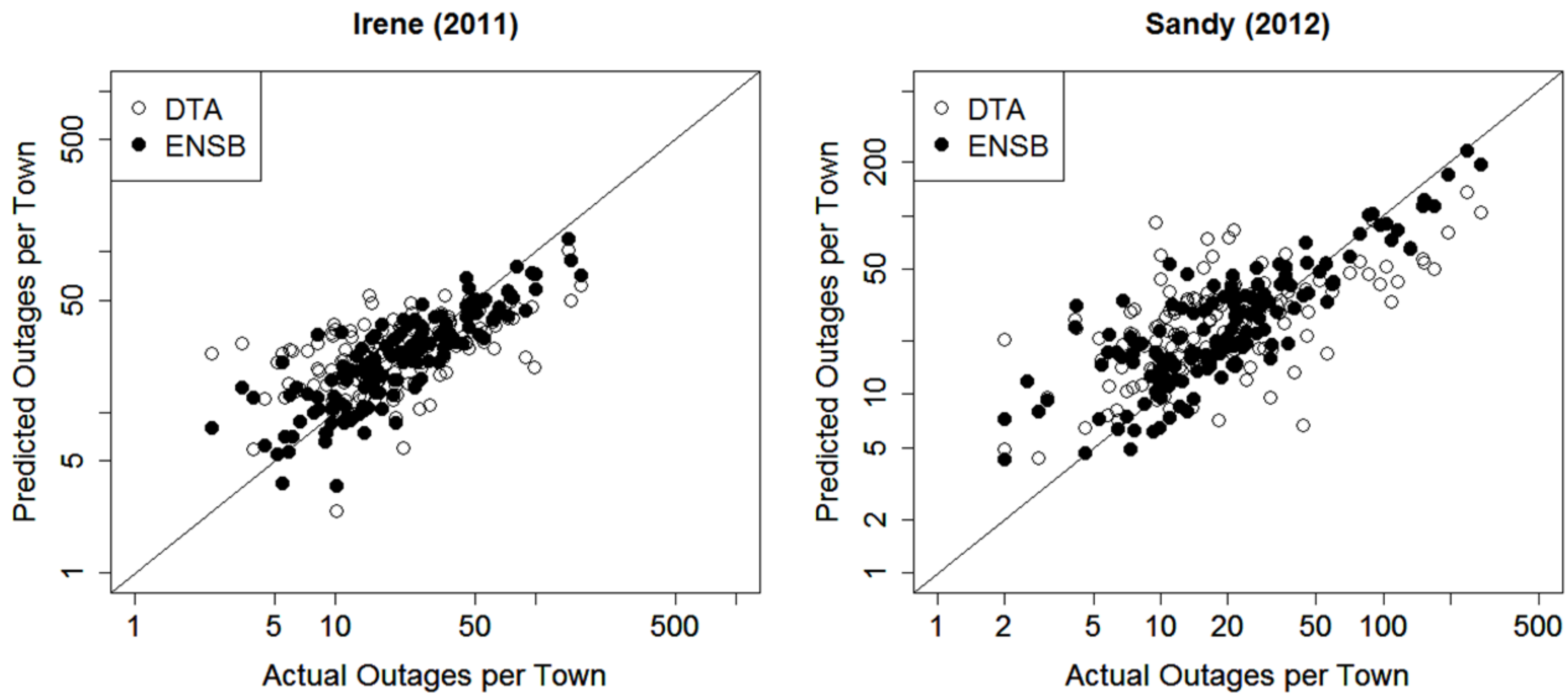

Figure 11: Scatterplots of actual vs. predicted outages per town for Storm Irene and Hurricane Sandy, with 45 degree line.

Table 1: Explanatory data included in the models.

Variable	Abbreviation	Description	Type	Units
Season	seasoncat	Weather	Categorical	-
Duration of wind at 10 meters above 9 m/s	wgt9	Weather	Continuous	hr
Duration of wind at 10 meters above 13 m/s	wgt13	Weather	Continuous	hr
Duration of wind at 10 meters above 18 m/s	wgt18	Weather	Continuous	hr
Duration of wind gusts above 18 m/s	ggt18	Weather	Continuous	hr
Duration of wind gusts above 27 m/s	ggt27	Weather	Continuous	hr
Duration of wind gusts above 36 m/s	ggt36	Weather	Continuous	hr
Duration of wind gusts above 45 m/s	ggt45	Weather	Continuous	hr
Total accumulated precipitation*	TotPrec	Weather	Continuous	mm
Wind stress*	Wstress	Weather	Continuous	unitless
Wind gust*	Gust	Weather	Continuous	m/s
Wind at 10 m height*	Wind10m	Weather	Continuous	m/s
Snow water equivalent*	SnoWtEq	Weather	Continuous	kg/kg
Soil moisture*	SoilMst	Weather	Continuous	mm/mm
Precipitation Rate*	PreRate	Weather	Continuous	mm/hr
Sum of Assets	sumAssets	Infrastructure	Continuous	count
Percent Developed	PercDeveloped	Land Cover	Continuous	%
Percent Coniferous	PercConif	Land Cover	Continuous	%
Percent Deciduous	PercDecid	Land Cover	Continuous	%

***variables have both mean and maximum values, there are 26 unique variables used in the models**

Table 2: Statistical metrics from the WRF evaluation against the METAR station data.

		RMSE	MB	MAE	R	NSE
Irene	SWS[a]	2.28	0.12	1.68	0.81	0.52
(103/3001)[b]	Max SWS	2.89	0.39	2.16	0.60	0.01
Sandy	SWS	2.57	0.53	1.85	0.73	0.36
(102/4159)	Max SWS	3.15	0.56	2.34	0.63	0.06
Blizzard	SWS	2.12	-0.69	1.68	0.73	0.40
(103/4495)	Max SWS	3.00	-1.78	2.54	0.74	-0.11

[a] SWS=sustained wind speed, m/s
[b] Values in parenthesis: (No. of stations/No. of observation-model pairs for SWS)

Table 3: Number of storms per season

Type	Number of Storms	Percentage of Total
Warm weather	38	43%
Cold weather	25	28%
Transition	24	27%
Hurricane	2	2%
Total	89	100%

Table 4: Land use changes for variables classification/aggregation methods.

Land Cover Class	Grid Average	Directly on Circuit	60m Buffer
Developed	5.8%	66.3%	21.7%
Turf & Grass	20.5%	7.8%	16.3%
Other Grasses	4.7%	0.8%	1.9%
Agriculture	12.0%	2.7%	8.4%
Deciduous Forest	29.9%	17.9%	40.1%
Coniferous Forest	10.9%	2.2%	5.6%
Water	3.1%	0.4%	2.0%
Non-forested Wetland	1.1%	0.1%	0.2%
Forested Wetland	8.5%	0.7%	2.1%
Tidal Wetland	0.4%	0.1%	0.3%
Barren Land	2.3%	0.5%	1.0%
Utility ROWs	0.8%	0.5%	0.3%

Table 5: Description of the evaluated models.

Model	Description
Model DT_A	Decision tree model with only weather data
Model RF_A	Random forest model with only weather data
Model BT_A	Boosted gradient tree model with only weather data
Model ENS_A	Ensemble decision tree inputted by models DT_A, RF_A, and BT_A
Model DT_B	Decision tree model with weather, infrastructure and land cover data
Model RF_B	Random forest model with weather, infrastructure and land cover data
Model BT_B	Boosted gradient tree model with weather, infrastructure and land cover data
Model ENS_B	Ensemble decision tree inputted by models DT_B, RF_B, and BT_B

Table 6: Median absolute error (MdAE, outages) and median absolute percentage error (MdAPE, %) by season and model.

Metric	**Season**	$\mathbf{DT_A}$	$\mathbf{RF_A}$	$\mathbf{BT_A}$	$\mathbf{DT_B}$	$\mathbf{RF_B}$	$\mathbf{BT_B}$	$\mathbf{ENS_A}$	$\mathbf{ENS_B}$
MdAE	Transition	88	82	125	97	84	116	54	64
MdAE	Cold	52	59	103	67	66	105	32	39
MdAE	Warm	85	92	100	91	101	102	72	83
MdAPE	Transition	46.7%	43.5%	54.0%	49.8%	44.9%	52.1%	29.8%	32.0%
MdAPE	Cold	33.4%	33.7%	47.6%	39.4%	41.7%	49.1%	23.5%	30.9%
MdAPE	Warm	45.3%	48.0%	56.4%	52.5%	50.6%	57.2%	35.1%	38.7%

APPENDIX

The statistical metrics used in the model evaluation analysis are presented below. The modelled wind speed is represented by M, the observed wind speed by O and N is the total number of data points used in the calculations.

- Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{N}(M - O)^2}$$

- Mean Bias (MB):

$$MB = \frac{1}{N}\sum_{N}(M - O)$$

- Mean Absolute Error (MAE):

$$MAE = \frac{1}{N}\sum_{N}|M - O|$$

- Correlation Coefficient (R):

$$R = \frac{\sum(M - \bar{M})(O - \bar{O})}{\sqrt{\sum(M - \bar{M})^2 \sum(O - \bar{O})^2}}$$

- Nash-Sutcliffe efficiency (NSE):

$$NSE = 1 - \frac{\sum(O - M)^2}{\sum(O - \bar{O})^2}$$

- Percentage within a factor of 2 or 1.5 (FAC2, FAC1.5):

$$FAC2 = 0.5 \leq \frac{M}{O} \leq 2,$$

$$FAC1.5 = 0.5 \leq \frac{M}{O} \leq 1.5$$