 DOI: 10.1111/j.1539-6924.2011.01618.x

Comparison and Validation of Statistical Methods for Predicting Power Outage Durations in the Event of Hurricanes

Roshanak Nateghi,[1,*] Seth D. Guikema,[1] and Steven M. Quiring[2]

This article compares statistical methods for modeling power outage durations during hurricanes and examines the predictive accuracy of these methods. Being able to make accurate predictions of power outage durations is valuable because the information can be used by utility companies to plan their restoration efforts more efficiently. This information can also help inform customers and public agencies of the expected outage times, enabling better collective response planning, and coordination of restoration efforts for other critical infrastructures that depend on electricity. In the long run, outage duration estimates for future storm scenarios may help utilities and public agencies better allocate risk management resources to balance the disruption from hurricanes with the cost of hardening power systems. We compare the out-of-sample predictive accuracy of five distinct statistical models for estimating power outage duration times caused by Hurricane Ivan in 2004. The methods compared include both regression models (accelerated failure time (AFT) and Cox proportional hazard models (Cox PH)) and data mining techniques (regression trees, Bayesian additive regression trees (BART), and multivariate additive regression splines). We then validate our models against two other hurricanes. Our results indicate that BART yields the best prediction accuracy and that it is possible to predict outage durations with reasonable accuracy.

KEY WORDS: Data mining; power system restoration; survival analysis

1. INTRODUCTION

Hurricanes cause power outages in the United States that can potentially affect millions of customers and lead to social and economic impacts. The longer the outage duration, the greater the incurred losses. Having accurate estimates of outage durations may allow utilities to better plan their recovery efforts by, for instance, assigning more repair crews to areas with longer expected outage durations. Accurate duration estimates would also allow the public and public agencies (e.g., state and federal emergency response agencies) to better plan their responses and coordinate restoration efforts for other infrastructure systems that depend on electric power. It is therefore important to be able to model power outage duration accurately.

A range of models have been proposed in the literature to model power system restoration time after natural disasters. Most of the proposed methods, however, rely on damage assessments made after the occurrence of the extreme events.[1–3] Liu *et al*.[4] were the first to implement a statistical regression model for power outages that could be applied as a storm was approaching rather after damage assessments had been completed.

The objective of this article is to examine different statistical methods for predicting power

[1]Johns Hopkins University, Baltimore, MD, USA.
[2]Texas A&M University, College Station, TX, USA.
*Address correspondence to Roshanak Nateghi, Johns Hopkins University; rnategh1@jhu.edu.

 0272-4332/11/0100-1897$22.00/1

outage durations in the event of hurricanes and to compare their accuracy. This is done by comparing the predictive accuracy of five different statistical models for estimating power outage durations using a range of covariates such as the hurricane wind speed, power system component inventories, the number of customers affected by each outage, the type of protective device that activated, and geographic and climatic variables such as soil moisture levels. More specifically, we compare the accuracy of out-of-sample predictions from five different models including accelerated failure time (AFT) regression, Cox proportional hazards (Cox PH) regression, Bayesian additive regression trees (BART), regression trees, and multivariate adaptive regression splines (MARS). Our models are developed using data from the power outages caused by the landfall of Hurricane Ivan in 2004. The data set consists of 14,320 outage records and 93 covariates. The developed models are then validated against Hurricane Katrina and Hurricane Dennis. This article shows that the Bayesian additive regression tree approach can estimate power outage durations during hurricanes with stronger predictive accuracy than the previously proposed models in the literature and that the accuracy of these predictions is relatively high.

2. BACKGROUND

A range of approaches have been used in the past to model power system restoration after disasters. The previous approaches include engineering fragility curve fitting approaches, deterministic resource constraint methods, probabilistic methods such as Monte Carlo simulations, optimization methods, discrete-state, discrete-transition Markov processes, network-based approaches, and statistical regression approaches. We first provide an overview of the nonmultivariate-regression approaches before providing a more in-depth review of the prior multivariate-regression-based approaches for estimating power outage duration times.

In the empirical curve fitting approach, the data are typically derived either from past records or by eliciting expert opinion to plot restoration curves that represent utility performance during extreme events.[1,3] These restoration curves typically show the percentage of customers for whom service has been restored or the percentage of demand met as a function of time since the event. The restoration curves usually do not account for any variables that may help explain spatial variations in restoration times, and the accuracy of restoration curves developed for future events is largely untested. Monte Carlo simulation has also been used to assess power system reliability. For example, Balijepalli *et al.*[5] use a bootstrap method to model lightning storm parameters. They then implement Monte Carlo simulations based on storm parameters and fault rates to estimate system reliability indices (i.e., both momentary and sustained outages). In the deterministic resource constraint method[6] the restoration procedure is modeled in a simplified manner, ignoring uncertainties associated with the restoration time. An optimization model is then implemented to minimize the average time that a customer is without power, using decision variables such as damage assessments, inspections times, and repair times and constraints such as the system's characteristics and the total number of crews available. Sensitivity analysis is then implemented to estimate the impact of different factors (e.g., the number of crews available) on restoration times. Another approach is to use a discrete-transition Markov processes approach to model the performance of individual infrastructure systems after an earthquake occurs. This method is challenging as it requires accurate estimates of model parameters and probability values and hence requires extensive amount of data or expert assessments. The discrete event simulation approach[7] is structurally similar to the discrete-transition Markov approach in that it is a simulation-based representation of the restoration process. However, in the discrete event simulation approach, the actions of each repair crew and each significant item of material (e.g., transformers) are explicitly modeled. The times at which each block of customers has its power restored is based on this explicit, detailed model of the restoration process. However, this is a data-intensive approach requiring detailed knowledge and modeling of the restoration process of a specific utility. In the network approach,[8] the power system is simplified to be comprising a series of supply and demand nodes that are connected to one another via links that could either be functional or nonfunctional. Graph theory and optimization techniques are then implemented to minimize the mean time for restoration, with recovery defined as customers being connected to supply nodes. This modeling approach allows for the consideration of both spatial and temporal variations of the restoration procedures. For example, Xu *et al.*[9] developed an approach for optimizing the order in which substations are repaired after a major disaster based on the detailed simulation model of Cagnan *et al.*[7] This approach seeks to decrease restoration times by making more efficient use of restoration resources. The idea behind using a statistical approach such as regression modeling to estimate power outage durations is to utilize a set of covariates and outage duration data from similar past events to predict the outage durations during future events. A regression approach thus differs from the approaches discussed above in one key aspect: it accounts for a range of variables that can influence outage durations. If the past data set is large enough and is representative of the response of the power system to the approaching hurricane, this approach has the potential to provide accurate estimates at a more spatially detailed level than the other approaches. However, the accuracy of the estimates of the statistical approaches are critically dependent on (1) the appropriateness of the model used and (2) the sufficiency of the underlying data.[10] If an inappropriate or inappropriately developed model is used or if the data are insufficient to support the model development effort, the predictive accuracy of the statistical approach will be poor.

Statistical modeling approaches such as multivariate regression have not been used as widely for estimating power outage durations as the other non-regression approaches. The only formal use of a regression or regression-like approach that we are aware of is that of Liu *et al.*[4] They implemented survival analysis to model power outage restoration times during hurricanes and ice storms. Survival analysis, which will be reviewed in depth in Section 3 below, models the duration of events such as power outages as a function of a number of explanatory variables. Liu *et al.*[4] implemented the two most common types of survival analysis models, AFT and Cox proportional hazard (Cox PH). They recommended AFT over Cox PH largely because the model output is easier to interpret.

In this article we examine survival analysis together with data mining techniques including classification and regression trees (CART), BART, and MARS to investigate which model can estimate power outage duration most accurately. All of these models are briefly discussed in Section 3.2. The results are provided and discussed in Section 4. The article closes with a summary of our main findings.

3. METHODS

3.1. Data Used

Many factors influence the susceptibility of electric power systems in a given geographic area to outages during hurricanes.[11,12] Examples of such factors include the number of power system components, such as the length of distribution lines and the number of poles, switches, and transformers, in the area. Other factors include the number of customers served in different locations, geographic characteristics of the area such as land use and land cover data, and climatic variables such as hurricane duration and intensity, long-term precipitation patterns in the area, and soil moisture levels prior to a hurricane making landfall. Soil moisture plays an important role in explaining the stability of the foundation of power system poles and trees that could potentially fall onto power lines and poles during hurricanes. We used the data from a utility serving the central Gulf Coast region. The service area was covered with 6,681 grid cells with dimensions of 3.66 km (12,000 foot) by 2.44 km (8,000 foot), and each variable used in the model was given at this scale. Below is a brief summary of the types of data that were used in this article.

To incorporate wind field characteristics of the hurricane in our models, we included estimates of the maximum three-second gust wind speed and the length of time that the winds were above 20 m/sec (44.7 miles/h) for each grid cell. The hurricane wind data were estimated by Impact Weather, a commercial forecasting service.[13]

In order to take into account the potential differences in outage restoration times for different land uses, we included information about land cover and land use. The land cover data used in this article are publicly available from the National Land Cover Database (NLCD) 2001.[25] The information on each of the 21 land cover classes is provided with a resolution of one arc-second (approximately 30 m). We categorized the 21 land cover types into eight aggregated classes, namely, water, developed (including residential, commercial, and industrial), barren, forest, scrub, grass, pasture, and wetland.

As in the work by Han *et al.*[11,12] we also incorporated soil moisture information, antecedent precipitation, and mean annual precipitation into our models to be able to better explain the variability of outage duration times. Soil moisture and antecedent precipitation were included because in preliminary discussions with utility personnel, the utility experts felt that saturated soils would increase outage durations due to repair trucks becoming stuck or avoiding those areas with particularly wet soil until the soil had a chance to dry some after the storm. Soil moisture was simulated at 1/2 degree (latitude/longitude) resolution by implementing the variable infiltration capacity (VIC) model.[14,15] Soil characteristics were obtained from the State Soil Geographic (STATSGO) database[16] created by the U.S. Natural Resource Conservation Service (NRCS). The STATSGO database was primarily designed to facilitate broad planning and management uses that cover state, regional, and multistate regions. It was constructed by generalizing the comprehensive soil survey data to a mapping scale of 1:250,000. The number of soil polygons per quadrangle map ranges from 100 to 400, and the minimum area mapped is approximately 625 hectares. Examinations of model performance have shown that VIC can accurately simulate the wetting and drying of the soil.[17,18] Quantifying antecedent precipitation before hurricane landfall was facilitated using the Standardized Precipitation Index (SPI).[19,20] The SPI is a statistical measure of precipitation compared to normal conditions (e.g., a measure of drought or wetness) and can be calculated for any time period. The SPI was calculated for six different time periods: 1, 2, 3, 6, 12, and 24 months, using monthly precipitation data from 1915 to 2005 at 1/2 degree (latitude/longitude) resolution. Mean annual precipitation and potential evapotranspiration are associated with the particular varieties of natural vegetation that tend to grow in an area.[21] Since some types of trees (e.g., pines) are more likely to be prone to being blown onto power lines during a hurricane than others, it is essential to take spatial variations in vegetation into consideration. We used mean annual precipitation in our models to characterize the spatial variability in the distribution of plant communities, as spatially detailed vegetation data were not available on a state-wide basis.

Mean annual precipitation (mm) was calculated at 1/2 degree (latitude/longitude) resolution using daily precipitation obtained from the National Oceanic and Atmospheric Administration (NOAA) Cooperative Observer (COOP) network data (1915–2004). It is important to note that mean annual precipitation is only a proxy that accounts for the broad patterns in vegetation as other factors such as soil fertility and human activities also strongly influence the distribution of vegetation. The soil moisture, SPI, and mean annual precipitation data were all downscaled to the utility company grid size, by employing an inverse-distance weighting (IDW) algorithm with a radius of influence of 100 km.

We also used information about the power system such as the number of transformers, poles, switches, customers, and the length of overhead and underground line in each grid cell. This information provides a measure of the extent of power system exposure to high winds during hurricane landfalls. We obtained the duration of each outage (14,320 in total) in each grid cell during Hurricane Ivan. A power outage is defined by the utility providing our data as the activation of a protective device leading to a nontransitory loss of power that requires repair. This excludes short-duration interruptions that were automatically cleared by protective devices in the system. A single power outage could affect varying numbers of customers (i.e., it could affect only a few or thousands of customers).

3.2. Models Implemented

We implemented both AFT and Cox PH models, two common regression models for the duration of events, and CART, BART, and MARS, three data mining methods. Each of these approaches is summarized below.

3.2.1. Regression Models

The AFT model is a parametric survival analysis model for time-to-event data. Such data, typically referred to as survival data, are generally nonsymmetrically distributed, often with a positive skew. It is therefore unreasonable to assume a normal distribution for them. Examples of the types of event durations appropriate to address with AFT models include individual lifetimes after particular medical treatments or disease diagnoses or the duration of failure events such as power outages from onset to resolution.

AFT relates the survival time to the explanatory variables through a linear relationship, as shown in Equation (1):

$$\ln(T_i) = X_t^T \beta + \varepsilon_i, \tag{1}$$

where T_i is the survival time random variable, X_i is the vector of covariates, β is the vector of parameters, and ε_i is the vector of errors that is assumed to be independently distributed. AFT is most typically fit using the method of maximum likelihood.

We implemented AFT in the R software package. Due to a high degree of collinearity among the covariates and the high dimensionality of the data set, AFT was implemented subsequent to principal components (PC) transformation of the explanatory variables. We evaluated a number of different distributions for the error term ε: Weibull, log-normal, exponential, and log-logistic. The AFT models with different underlying distributions were compared using the Akaike information criterion (AIC),[22] which is calculated using the equation below:

$$\mathrm{AIC} = -2\ \log L + 2p, \tag{2}$$

where L is the likelihood of the fitted model and p represents the number of parameters used in the model. The AIC value for each of the assumed distributions is presented in Table I.

It can be seen from the results shown in the table that assuming a Weibull distribution for outage duration times leads to the lowest AIC value. Therefore, we chose Weibull to be the underlying distribution of the power outage duration times.

A Cox PH model is similar to an AFT model in that it is a survival analysis regression model. However, in a Cox PH model, the potential covariates are linearly related to the logarithm of the hazard rate in a semiparametric approach, as shown in Equation (3):

$$\log(h_i) = \alpha + \beta_1 x_1 + \cdots + \beta_k x_k, \tag{3}$$

where the hazard function h_i is defined to be the instantaneous rate of power restoration conditioned on the fact that the outage has lasted to time t, as illustrated in the equation below:

$$h(t) = \lim \frac{pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t}. \tag{4}$$

In contrast to AFT, Cox PH is a semiparametric model and makes no assumptions about the distribution of the baseline hazard function. The fact that this model relates the covariates to the hazard rate and not directly to the response function makes the interpretation of the model output more challenging.

Table I. AIC Values for AFT Models Based on Different Distributions

Distribution	AIC Value
Weibull	253,628.4
Log-normal	257,947.8
Exponential	254,180.6
Log-logistic	256,200.8

3.2.2. Data Mining

We implemented three data mining methods: CART, BART, and MARS. We provide a summary of each of these methods in this section.

CART are built by binary splitting of the data space into terminal nodes.[24] In building regression trees the best splits s are chosen such that the sum of squared errors (or least absolute deviation) within each node t is minimized. The data space is split recursively until each terminal node contains no more than a certain predefined minimum number of records. The average (or median value) is then assigned to the terminal nodes. The tree is subsequently pruned back based on minimizing a cost complexity criterion, defined in Equation (5), to avoid overfitting:

$$C_\alpha T = \sum_{m=1}^{|T|} N_m Q_m T + \alpha |T|. \tag{5}$$

In Equation (5), $|T|$ represents the total number of terminal nodes, N is the total number of cases, and Q_m represents within-node residual sum of squares. α is the tuning parameter that determines the tradeoff between the tree size and its fit. The larger the value of α, the smaller the tree.

A BART model is fully Bayesian and nonparametric.[23] In a BART model the final estimate consists of the summation of the estimate from m small trees, each of which is a weak leaner (See Equation (6)):

$$Y = \sum_{j=1}^{m} g(x, T_j, M_j)\tau\varepsilon, \tag{6}$$

where $\varepsilon \sim (0, \sigma^2)$.

In Equation (6), g_i denotes a regression tree. For each of the tree structures denoted by T_j, and its associated terminal node parameter denoted by M_j, the function $g(x, T_j, M_j)$ assigns the mean value (μ) to the vector of covariates x.

BART is both a fully Bayesian probability model with a prior and a likelihood and a data mining method. A BART model comprises a set of small trees with each tree constrained by a prior to restrict each tree's contribution to the final model, making each individual tree a weak leaner. Fit and inference in BART are achieved through a Markov chain Monte Carlo algorithm. BART also provides an approach for estimating the relative importance of the different explanatory variables. BART does this based on the frequency with which each variable is used in the small trees comprising the overall model. Those variables used most frequently are regarded as the most important explanatory variables. This facilitates identifying the covariates that are most important in explaining the variations in the response variable.

MARS is a nonparametric regression technique that allows for nonlinearities and interaction effects in the model and therefore is particularly suitable for modeling high-dimensional data sets. MARS consists of a summation of a series of linear splines that allows the response variable to vary nonlinearly with the covariates. It generates a set of reflected pairs of linear splines (i.e., one being the reflection of the other around the knot point t) that are of the form: $C = (X_j - t)^+, (t - X_j)^+$, where X_j is the vector of input variables and t is the knot point at values of the input variables. The overall MARS model is built from the linear splines:

$$f(x) = \beta_0 + \sum \beta_m h_m(X), \tag{7}$$

where h_m are the linear splines of the input variables, β_0 is the intercept term, and β_m is the vector of the coefficients. The β_m parameters are estimated by minimizing sum of squared errors. MARS is built in a forward manner starting from a basis function of one (the intercept). It then adds the linear functions iteratively, using a greedy algorithm that chooses the reflected pair of splines that lead to the largest reduction in the training error. The model is then pruned back using generalized cross-validation to avoid overfitting. The generalized cross-validation error is of the form shown in Equation (8):

$$GCV = \frac{RSS}{\left(1 - \frac{r + cK}{N}\right)^2}, \tag{8}$$

where RSS stands for residual sum of squares, c is the penalty term, K is the number of knot points, and r is the number of linearly independent basis functions. We used fivefold out-of-sample cross-validation to choose the optimal interaction and penalty terms, and the results of this selection process are shown in Table II. The best predictive accuracy is achieved when the penalty term is 4 and the interaction term is 3.

Table II. Mean Absolute Deviation and Root Mean Square Error for MARS Models with Different Interaction and Penalty Terms

Penalty	Interaction	MAD	RMSE
1	1	1,321.63	1,805.32
2	1	1,319.17	1,791.73
3	1	1,314.12	1,777.49
4	1	1,301.87	1,768.86
4	2	1,117.42	1,649.77
4	3	1,093.75	3,637.92

4. RESULTS

We have examined five distinct statistical approaches to estimate power outage duration times related to Hurricane Ivan in 2004, a strong but not atypical hurricane for the Gulf of Mexico. Comparing the out-of-sample predictive accuracy of the methods indicates which statistical technique predicts outage durations most accurately for this hurricane. We also examined whether the developed models yield improved results compared to having no model and simply using the mean of outage duration times for the in-sample grid cells as the predictor for the out-of-sample grid cells. The outage duration time caused by Hurricane Ivan has a mean of 2,870.2 minutes (approximately 48 hours) and a standard deviation of 2,395.7 minutes. Our prediction errors are calculated based on 50 random hold-out cross-validation tests. In each of 50 independent iterations, 10% of the data are randomly held out to create a validation set. The model is then built on the remaining subset of the data, the training set, and the predictions are tested against the validation set. The root mean squared error (RMSE) and the mean absolute deviation (MAD) values are shown in Table III. They represent the difference between the actual outage duration times and the values predicted by the developed models, averaged over the 50 repeated validation tests.

The results in Table III show that implementing AFT, BART, and MARS models leads to an improved prediction accuracy over having no model. It also indicates that BART yields the best

Table III. Out-of-Sample Prediction Results

Model	MAD	RMSE
AFT	1,642.1	2,130.0
COX PH	2,793.1	3,706.1
BART	471.7	894.0
CART	2,134.7	2,602.7
MARS	1,106.7	1,860.1
No model	1,896.8	2,395.6

out-of-sample predictive accuracy of the tested models by a substantial margin. The differences between the BART errors and the errors of each of the other models were compared based on a two-sided Student's t-test with unequal variances. The maximum of the p-values for both the four hypothesis tests based on MAD and RMSE was less than 2.2×10^{-16}. This indicates that we should reject the null hypothesis of equal means and that the differences between the error vectors are statistically significant. If attention is restricted to the two regression models, the AFT and Cox PH models, Table III shows that the errors are substantially higher for the Cox PH model. The p-value for the test comparing these two models is also less than 2.2×10^{-16}, indicating that the difference between their error vectors is statistically significant and that AFT outperforms the Cox PH model in terms of its out-of-sample predictive accuracy. Overall, the AFT, BART, and MARS models all yield improved predictive accuracy relative to not using a prediction model. BART, in particular, yields a very promising predictive accuracy for this hurricane.

4.1. Physical Interpretation

Table IV shows the top 14 most important variables based on the fitted BART model.

The index column lists the frequency of inclusion of the covariates in the ensemble model. In other words, it represents the total number of times that the particular covariate is included as a splitting variable over all 200 trees in the BART model. The average of this frequency number in our model is 2.1. A higher index number for a covariate indicates that the variable has been used more number of times in the tree ensemble and is consequently more influential in explaining the variability of the response function. The table of variable importance, together with the AFT model output, help with understanding the influence of different covariates on power restoration times (PRT). The complete output of our AFT model is included in Appendix A.

Table IV. Top 14 Variables Identified by BART as Most Important

Parameter	Index
Station-breaker activated (binary)	5.7
Depth to bedrock crossed with a binary variable indicating a primary land type of "water"	5.6
12-month standardized precipitation index	5.3
Low intensity residential land cover	4.9
Outage cause code (tree or/and weather)	4.8
9-month standardized precipitation index	4.6
Soil moisture the day before landfall at the 10–40 cm level	4.6
Number of customers affected by the outage	4.6
Duration of winds above 20 m/sec	4.3
Outage cause code (wind or/and rain)	4.2
Forested land cover type	4.1
Three-second maximum gust wind speed	4.1
Clay content of soil as a percent crossed with a binary water indicator (0 = water)	4.0
Maximum aspect ratio	4.0

Our final models indicate that there is a negative association between PRT and the number of customers affected by the outage. This is reasonable since the areas with more customers are usually prioritized more highly in the restoration efforts. The activation of a station breaker is negatively correlated with PRT, which is also expected since the power company would naturally prioritize a location where the station breaker has been deactivated. Duration of winds above 20 m/sec and three-second maximum gust wind speed are shown to be positively correlated with PRT, indicating that those areas bearing the brunt of the storm tend to have longer outage times. Standardized Precipitation Indices for 9- and 12-month windows are negatively correlated with the PRT. This suggests that drier than normal conditions make trees more prone to snapping, potentially leading to outages that take longer to repair. The "developed" land cover type also has a negative association with PRT, indicating that areas that are predominantly developed are more highly prioritized in restoration efforts. Clay content of the soil is positively associated with PRT suggesting that clay-rich soils prolong the restoration times. We discussed this issue with utility company personnel. They reported that restoration takes longer in areas with higher clay content. Repair trucks are more likely to get stuck in wet clay soils, so repair crews prioritize areas with soils of lower clay content until the high clay soils partially dry out. This delays restoration in areas with clay-rich soils. As depth to bedrock decreases, outage duration times increase, all else remaining constant. The reasons for this relationship are unclear, but if bedrock gets close enough to the surface there would be increased difficulty installing new poles. Soil moisture levels the day before landfall at 10–40 cm level are both negatively associated with PRT, again potentially related to drier soils making trees more prone to snap or break. Finally, it should be noted that the number of crews available is also a very important factor in determining the outage durations. However, that information was not incorporated in our models.

5. FURTHER VALIDATION

In order to further validate the models, we applied the models developed using data from Hurricane Ivan to two other hurricanes, Katrina and Dennis, which both made landfall in 2005. The outage duration times for Hurricane Katrina in the service area of the utility providing the data have a mean of 2,075.7 minutes (approximately 35 hours) and a standard deviation of 2,222.0 minutes. Hurricane Dennis was a smaller hurricane, though it did still cause widespread outages. The mean outage duration for Hurricane Dennis was 1,009.6 minutes (approximately 17 hours) with a standard deviation of 800.0. The results of using the models developed based on Hurricane Ivan data to predict power outage durations for the other two hurricanes are summarized in the tables below. As can be seen from Tables V and VI, BART outperforms all other statistical models in term of its predictive accuracy for hurricanes not used to develop the models. Moreover, all the developed models (except for Cox PH for Hurricane Katrina) yield improved predictive accuracy over having no statistical model and using the mean of the observed outage duration times during Hurricane Ivan as the predictor for the new storms. It should also be pointed out that using Hurricane Ivan's mean duration time as the predictor yields significantly worse results (in terms of predictive accuracy) for Hurricane Katrina and Hurricane Dennis compared to the case of Hurricane Ivan.

Table V. Out-of-Sample Prediction Results for Hurricane Dennis

Model	MAD	RMSE
AFT	582.6	759.6
COX PH	945.4	1186.3
BART	272.5	422.2
CART	746.8	921.4
MARS	336.1	501.6
No model	1,878.9	2,025.3

Table VI. Out-of-Sample Prediction Results for Hurricane Katrina

Model	MAD	RMSE
AFT	1,264.5	1,738.3
COX PH	2,061.4	3,000.8
BART	413.59	782.15
CART	1,646.2	2,235.5
MARS	1,412.1	859.1
No model	2,032.0	2,359.7

6. CONCLUSION

Electric power disruptions create ripple effects in most sectors of society, causing disruptions to other systems that depend on electricity. Electric power is critical to important aspects of our daily lives such as banks, hospitals, transportation, security systems, heating systems, water distribution systems, telecommunication systems, food storage, and businesses. Accurate estimates of power outage durations can help utilities in optimizing their recovery efforts by helping them prioritize areas with longer estimated outage durations and deploying higher numbers of repair crews to them. These estimates can also help the public and public agencies to better plan their responses and coordinate restoration efforts for other infrastructure systems that depend on electric power. In this article we used a data set containing power outage duration times caused by Hurricane Ivan in 2004, a data set provided by a major power company in the central Gulf Coast region. Two survival analysis models together with three data mining techniques were implemented to compare the predictive accuracy of these different models. Our results indicated that BART outperforms all other statistical models with its predictive accuracy being superior to the other models we tested. BART's table of variable importance facilitates identifying the most important factors in predicting power outage restoration times. Moreover, the AFT model provides a further basis for examining the influence of each covariate on the restoration periods. The models developed based on data from Hurricane Ivan were then used to predict outage durations for Hurricanes Katrina and Dennis, providing a further test of the different statistical models. The results of our validation suggest that (1) statistical models can yield accurate, spatially detailed estimates of power outage durations after hurricanes and (2) BART yields the best predictive accuracy of the tested models and, as a result, data mining methods such as BART should be utilized in additional to standard regression-based survival analysis models if the accuracy of outage duration prediction is a major consideration.

APPENDIX

Sign	Coef.	Parameter	Sign	Coef.	Parameter
–	0.05	Number of customers affected	–	0.01	Maximum aspect ratio
–	0.02	1-month standardized precipitation index	+	0.01	Median slope
–	0.01	2-month standardized precipitation index	+	0.01	Mean slope
–	0.02	3-month standardized precipitation index	+	0.01	Maximum slope
–	0.03	6-month standardized precipitation index	+	0.01	Standard slope
–	0.02	9-months standardized precipitation index	+	0.1	Three-second maximum gust wind speed
–	0.02	12-month standardized precipitation index	+	0.08	Duration of wind above 20 m/sec
–	0.02	24-month standardized precipitation index	–	0.04	Activation of jumper
+	0.08	Mean annual precipitation	+	0.01	Activation of transformer
+	0.03	Overhead system	–	0.09	Activation of the station-breaker
–	0.1	Miles of power lines underground	+	0.04	Activation of line-switch
+	0.18	Miles of power lines overhead	+	0.03	Activation of other devices
–	0.01	Cause code: equipment failure	+	0.06	Activation of transmission device
–	0.03	Cause code: other equipment	+	0.04	Activation of source distribution
–	0.03	Cause code: deterioration	+	0.01	Standard elevation
+	0.06	Cause code: wind/rain	+	0.01	Minimum elevation
+	0.02	Cause code: falling trees/ weather	+	0.01	Mean elevation
–	0.02	Cause code: whether or not the outage was scheduled	–	0.01	Hydraulic group A
–	0.03	Cause code: lightning	+	0.01	Hydraulic group B
–	0.01	Cause code: tree-cut	+	0.01	Hydraulic group C
–	0.16	Cause code: animals	–	0.01	Water
–	0.07	Unknown cause code	+	0.01	Clay content of the soil-water
–	0.04	Cause code: vines	–	0.01	Depth to bedrock-water
–	0.04	Cause code: vehicles	–	0.02	Soil moisture the day before landfall at the 10–40 cm level
–	0.02	Cause code: tree-growth	–	0.02	Soil moisture the day before landfall from 40 cm to the bedrock
–	0.02	Cause code: dead-tree	+	0.01	Awch-water
–	0.19	Unclassified cause code	+	0.02	Land cover 1 (water)
–	0.04	Cause code: loose connection	–	0.04	Land cover 3 (barren)
–	0.01	Type: substation	+	0.03	Land cover 4 (forested upland)
–	0.07	Type: transformer	+	0.05	Land cover 5 (shrubland)
+	0.06	Type: primary power lines	–	0.03	Land cover 7 (herbaceous upland)
–	0.02	Type: secondary power lines	+	0.04	Land cover 8 (planted/cultivated)
+	0.01	Level: transformer	–	0.05	Land cover 21 (developed, open space)
–	0.01	Minimum aspect ratio	–	0.04	Land cover 22 (developed, low intensity)
–	0.01	Median aspect ratio	–	0.03	Land cover 23 (developed, medium intensity)
–	0.01	Mean aspect ratio	–	0.01	Land cover 24 (developed, high intensity)

ACKNOWLEDGMENTS

The authors would like to thank U.S. DOE (DE-FG02-08ER64644) and Johns Hopkins Whiting School of Engineering for funding this research and an anonymous power utility for providing the outage duration data.

REFERENCES

1. Applied Technology Council. Seismic Vulnerability and Impact of Disruption of Lifelines in the Conterminous United States, ATC-25, Redwood City, CA, 1991.
2. Chang SE, Svelka WD, Shinozuka M. Linking infrastructure and urban economy: Simulation of water-disruption impacts in earthquakes. Environment and Planning B: Planning and Design, 2002; 29:281–301.
3. Nojima N, Sugito M. Empirical Estimation of Outage and Duration of Lifeline Disruption Due to Earthquake Disaster. Proceedings of the US–China–Japan Workshop on Lifeline Systems, October, 2002.
4. Lui H, Davidson RA, Apanasovich TV. Statistical forecasting of electric power restoration times in hurricanes and ice storms. IEEE Transactions on Power Systems, 2007; 22(4):2270–2279.
5. Balijepalli N, Venkatta S, Richter C Jr., Christie R, Longo V. Distribution system reliability assessment due to lightning storms. IEEE Transactions on Power Delivery, 2005; 20(3):2153–2159.

6. Isumi M, Shibuya T. Simulation of post earthquake restoration for lifeline systems. International Journal of Mass Emerging Disasters, 1985; 3(1):87–105.
7. Cagnan Z, Davidson R, Guikema SD. Post-earthquake restoration planning for Los Angeles electric power. Earthquake Spectra, 2006; 22(3):1–20.
8. Nojima N, Kameda H. Optimal strategy by use of tree structure for post-earthquake restoration of lifeline network systems. Proceedings of the 10th World Conference on Earthquake Engineering.
9. Xu N, Guikema SD, Davidson RA, Nozick LK, Cagnan Z, Vaziri K. Optimizing scheduling of post-earthquake electric power restoration tasks. Earthquake Engineering and Structural Dynamics, 2007; 36:265–284.
10. Guikema SD. Natural disaster risk analysis for critical infrastructure systems: An approach based on statistical learning theory. Reliability Engineering and System Safety, 2009; 94(4):855–860.
11. Han SR, Guikema SD, Quiring SM, Lee KH, Davidson RA, Rosowsky D. Estimating the spatial distribution of power outages during hurricanes in the Gulf Coast region. Reliability Engineering and System Safety, 2009; 94(2):199–210.
12. Han SR, Guikema SD, Quiring SM. Improving the predictive accuracy of hurricane power outage forecasts using generalized additive models. Risk Analysis, 2009; 29(10):1443–1453.
13. www.impactweather.com, Last accessed on June 28, 2010 at 12:00 pm.
14. Liang X, Lettenmaier DP, Wood EF, Burgess SJ. A simple hydrologically based model of land surface water and energy fluxes for GCMs. Journal of Geophysical Research, 1994; 99(14):14415–14428.
15. Liang X, Wood EF, Lettenmaier DP. Surface soil moisture parametrization of the VIC-2L model: Evaluation and modifications. Global Planetary Change, 1996; 13:195–206.
16. STASTSGO. Available at: http://soils.usda.gov/survey/geography/statsgo/, Last accessed in Jan. 2010.
17. Meng L, Quiring SM. A comparison of soil moisture models using Soil Climate Analysis Network SCAN: Observations. Journal of Hydrometer, 2008; 9:641–659.
18. Robock A, Luo L, Wood EF, Wen F, Mitchell KE, Houser PR, Schaake JC, Lohmann D, Cosgrove B, Sheffield J, Duan W, Higgins RW, Pinker RT, Tarpley JD, Basara JB, Crawford KC. Evaluation of North American Land Data Assimilation System over the southern Great Plains during the warm season. Journal of Geophysical Research, 2003; 108(D22): 8846, doi: 10.1029/2002JD003245.
19. McKee TB, Doesken NJ, Kleist J. The relationship of drought frequency and duration to time scales. 8th International Conference on Applied Climatology, pp. 179–184, 1993.
20. McKee TB, Doesken NJ, Kleist J. Drought Monitoring with Multiple Time Scales. International 9th Conference on Applied Climatology, pp. 233–236, 1995.
21. Mather JR. The Climatic Water Budget in Environmental Analysis. MD: Lexington Books, 1978.
22. Akaike H. Statistical predictor identification. Ann. Statist. Math. 1970; 22:203–217.
23. Chipmman HA, George EI, McColluch RE. BART: Bayesian Additive Regression Trees, arXive: 0806.3286v1. Available at: http:// arxiv.org/ abs/0806.3286, Accessed in June 2008.
24. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Boca Raton, FL: CRC Press, 1998.
25. National Land Cover Database, 2001. Available at: http://www.epa.gov/mrlc/nlcd-2001.html, Last accessed in Jan. 2010.