A Longitudinal Analysis of the Drivers of Power Outages During Hurricanes: A Case Study with Hurricane Isaac

Gina Tonn[a*], Seth Guikema[a], Celso Ferreira[b], and Steven Quiring[c]
[a] Johns Hopkins University, Baltimore, MD, US
[b] George Mason University, Fairfax, VA, US
[c]Texas A&M University, College Station, TX

Abstract: In August 2012, Hurricane Isaac, a Category 1 hurricane at landfall, caused extensive power outages in Louisiana. The storm brought high winds, storm surge and flooding to Louisiana, and power outages were widespread and prolonged. Hourly power outage data for the state of Louisiana was collected during the storm and analyzed. This analysis included correlation of hourly power outage figures by zip code with wind, rainfall, and storm surge using a non-parametric ensemble data mining approach. Results were analyzed to understand how drivers for power outages differed geographically within the state. This analysis provided insight on how rainfall and storm surge, along with wind, contribute to power outages in hurricanes. By conducting a longitudinal study of outages at the zip code level, we were able to gain insight into the causal drivers of power outages during hurricanes. The results of this analysis can be used to better understand hurricane power outage risk and better prepare for future storms. It will also be used to improve the accuracy and robustness of a power outage forecasting model developed at Johns Hopkins University.

Keywords: Power Outages, Hurricanes, Random Forest

1. INTRODUCTION

Hurricane Isaac hit Louisiana in August 2012 and caused substantial power outages. It was a Category 1 hurricane at landfall and 47% of the state's electric customers lost power. The storm was large, slow-moving, and had significant storm surge associated with it. In comparison with other hurricanes, Isaac ranks fourth in customer power outages, behind Hurricanes Katrina, Gustav, and Rita, for the Entergy service area in Louisiana, Mississippi, Texas and Arkansas [1].

Power outages result in direct repair and restoration costs for utility companies, and can also result in loss of services from other types of critical infrastructure that are reliant on power service such as water, transportation, and communications systems. This can delay recovery times for a community that is impacted by a hurricane [2]. Accurate predictions of power outages prior to a storm can benefit both utility companies and government agencies by making planning and recovery more efficient [3].

Power outage prediction is often accomplished through the development of models based on wind field estimates, along with other covariates such as power system data, soil moisture levels, land use and topographical indicators [3]. A number of such statistical models have been developed [2,3,4]. While these models can be very accurate for some storms, they are less accurate for others.

In addition to accuracy of models varying from storm to storm, the causes of the outages can vary geographically across a region, and the existing models do not include some potential causes of power outages, particularly high rainfall. The main goal of this paper is to obtain a better understanding of the causes and geographic variance of power outages, both to improve basic understanding and to provide a stronger basis for improving outage forecasting models. Are the causes the same for a coastal area as an inland area? How important are rainfall and surge relative to wind? Damage to power systems is recorded by utilities, but good data on causes of outages is not generally available, making a longitudinal approach necessary.

[*] Corresponding author: Gina Tonn (e-mail: gtonn2@jhu.edu)

Can statistical analysis of power outage data and covariate data provide a better understanding of the causes of outages? The purpose of this study was to look at power outages longitudinally across the state of Louisiana for Hurricane Isaac to identify how the importance of covariates changes geographically. The results of this analysis may inform power outage prediction models and help to build more resilient infrastructure through improved understanding of the causes of outages.

In Section 2, the data used for the analysis as well as the statistical analysis methods are presented. Results and Discussion are included in Section 3, and Conclusions in Section 4.

2. METHODS AND DATA

2.1 Overview of Methods and Data

We focused on variables related to three key physical hazards associated with hurricanes: wind, storm surge, and rainfall in order to gain a better understanding of relative contribution of these three drivers. We analyzed all variables on an hourly basis, and so included variables that change over time as the storm progresses. We obtained data for the covariates of interest from publically available sources or modeled them based on publically available data. A summary of the covariates, abbreviations used for covariate names, data sources, and a description of each covariate is provided in Table 1. While data was available in varying time increments for each covariate, we performed interpolation to obtain hourly estimates. We chose the hourly change in outages as the response variable, and hours that did not have a positive increase in outages were removed from the analysis to focus the analysis on only the outage occurrence portion of the storm, not the outage restoration part of the storm. A more detailed description of each category of data and the data interpolation are provided in Sections 2.2 through 2.6.

Table 1: Summary of Covariates

Covariate	Abbreviation	Source	Description
Cumulative Precipitation	cumprecip	NCDC	Total precipitation amount during storm duration to hour of analysis in inches
Hourly Precipitation Total	precip	NCDC	Precipitation amount in hour of analysis in inches
Wind Speed	windspeed	Texas A&M model	Wind speed in meters/second for zip code in hour of analysis
Wind Gust Duration	windduration	Texas A&M model	Duration of wind gust >20 meters/second for zip code in hour of analysis
Previous Outages	prevout	Entergy website	Number of outages in previous hour of analysis for zip code
Population	population	US Census Bureau	Population estimate for zip code
Average Surge	surgeAVG	George Mason University model	Average storm surge depth for zip code in hour of analysis
Minimum Surge	surgeMIN	George Mason University model	Minimum storm surge depth for zip code in hour of analysis
Maximum Surge	surgeMAX	George Mason University model	Maximum storm surge depth for zip code in hour of analysis
Surge Variance	surgeVAR	George Mason University model	Variance of storm surge depth for zip code in hour of analysis

After completing the data collection and interpolation, we ran a Random Forest model for the entire data set. The most important variables were identified through Random Forest based importance measures for use in additional analysis as described further below. Using this reduced set of covariates, we ran the Random Forest analysis for each zip code separately. We plotted the results in map format for analysis of spatial trends. Then we used a Quantile Regression Forest for selected zip codes to gain insight into model accuracy. The modeling and analysis methods are described in more detail in Sections 2.7 through 2.9.

2.2 Outage Data

Power outage data was collected during the duration of the storm from August 27 to September 5, 2012. The data was harvested from the Entergy Louisiana website by a team of researchers at Johns Hopkins University [5]. The data was collected on a half-hourly basis during periods of peak outages, and was collected less frequently during non-peak outage periods. Data collected included the number of current customer outages by zip code. In order to standardize the data for use in analysis, we performed linear interpolation to estimate the number of outages for each zip code at the top of each hour for the duration of the storm.

We chose the change in outages (termed delta outages in this paper) for each hour of analysis for each zip code as the response variable for this analysis. Total power outages for the previous hour of analysis for each zip code was included as a covariate to account for the fact that the number of customers already without power impacts the number of power outages occurring in a given hour.

2.3 Precipitation Data

Precipitation data were obtained from the National Climatic Data Center (NCDC) website. Data was available for 36 stations in Louisiana. The time intervals at which the precipitation data were recorded varied by station, but were typically hourly or half-hourly. The data obtained was the hourly total rainfall [6]. In order to standardize the data for use in analysis, we interpolated the data set to estimate the hourly precipitation (precipitation that occurred in the previous 60 minute period) at the top of the hour for each station. Because our analysis was performed on a zip code basis, we needed rainfall estimates for each zip code. Based on the geographic coordinates of the zip code centroids and on the locations of the stations, we generated hourly rainfall estimates for each zip code using inverse distance weighted interpolation based on the spatially sparser set of rainfall stations that we had available.

2.4 Storm Surge Model

We used the coupled version of the 2-Dimensional Depth Integrated version of the Advanced Circulation (ADCIRC) model and the wave model SWAN [7] to simulate hurricane storm surge. The ADCIRC model [8] is a finite element, shallow water model that solves for water levels and currents at a range of scales and is widely used for storm surge modeling (e.g., Ferreira *et al.* 2013) [9]. This version of the program solves the Generalized Wave Continuity Equation (GWCE) and the vertically integrated momentum equations. SWAN is a third generation spectral wave model [10] that computes the time and spatial variation of directional wave spectra. We used the pre-validated numerical mesh *SL15* presented in Bunya *et al.* (2010) [11] and validated by Dietrich *et al.* (2010) [12] with resolution up to 30 meters in some areas. The hurricane surge model was forced by wind and pressure fields developed by a parametric asymmetric wind model [13] that computes wind stress, average wind speed and direction inside the Planetary Boundary Layer (PBL) based on the National Hurricane Center (NHC) best track data [14] meteorological conditions (e.g., central pressure, forward speed and radius to maximum wind). The simulations for Hurricane Isaac included tides (Tidal potential components M2, S2, N2, K2, K1, O1 and Q1) and neglected rivers inflows. Simulation results were recorded at 15 minute intervals for every model node in the study region. The water levels for each model node within each zip code were extracted from the entire model domain and inundation levels were converted to the NAVD88 vertical datum. Covariates based on the storm surge model include average storm surge, maximum storm surge, minimum storm surge, and storm surge variance.

2.5 Wind Model

The parametric wind field model of Willoughby et al. (2006) [15] is used to generate wind estimates for the duration of the hurricane at the zip code level for Hurricane Isaac. Parametric hurricane models are formulated from a physical understanding of hurricane wind fields. That is, winds are calm in the eye of the hurricane and they are typically at a maximum in the eyewall. Outside the eyewall the wind decreases with radius, although not always monotonically, and become near zero at some distance from the center of circulation. This wind field model was previously used in Han et al. [2, 16]. Two of the covariates are based on output from this model. The first is maximum wind speed in meters per second in the previous hour. The second is wind gust duration greater than 20 meters per second, with duration being taken cumulatively over the life of the storm for each zip code. Both of these covariates are simulated for the centroid of each zip code polygon based on running the wind field model every 60 minutes over the duration of the storm.

2.6 Other Data

Population estimates for each zip code were obtained from the US Census Bureau American Community Survey. These estimates were based on the US Census Bureau data for the year 2011. Because the US Census bureau does not track population on a zip code basis, the population data is an estimate based on census tract data (US Census) [17].

2.7 Random Forest and Quantile Regression Forest Methods Overview

A Random Forest is a non-parametric ensemble data mining method [18]. In the method, a large number of regression trees are developed, with each tree based on a bootstrapped sample of the data set. Random Forest models are good for data sets with non-linear data, outliers, and noise. Two types of output from the Random Forest model fit very nicely with the objectives of this analysis. The first is variable importance, which is a measure of the contribution of a given covariate to the model prediction accuracy. The second is the partial dependence plot. These plots show the marginal effect of a covariate on the response variable. The randomforest package in R was used for this analysis [19].

Quantile Regression Forests provide a non-parametric way of estimating conditional quantiles based on an underlying Random Forest model [20]. Quantiles give more information about the distribution of the response variable as a function of the covariates than just using the conditional mean as a standard Random Forest model does. In this method, regression trees are grown as in the Random Forest method. Then the weighted distribution of the observed response variables is used to estimate a conditional distribution. The difference between Random Forest models and Quantile Regression Forest models is that Random Forest models keep only the mean predictions and disregard other information. Quantile regression forests estimate the quantiles of the predictions based on the trained forest (Meinshausen, 2006). The quantregForest package in R was used for this analysis [21].

2.8 Statistical Analysis

A statewide Random Forest model was run using the data for all covariates and zip codes. Variable importance was reviewed to identify the variables that are most significant for predictive accuracy. Based on the variable importance, one covariate from each category of covariates (precipitation, wind, storm surge, and outages) was retained for individual zip code analysis in order to better understand the influences of the different variables. Partial dependence plots were generated for each of these covariates, and were reviewed to understand the marginal effects of these covariates on the response variable.

In order to understand the relative importance of the four covariates, and how that importance varies geographically, plots of importance for each of the covariates were generated. Because the magnitude of variable importance is not the same for each Random Forest run, comparing the variable importance between zip codes would not be useful. Instead, we calculated a percent variable importance for each zip code. The variable importance for the four covariates (wind speed, cumulative precipitation, maximum storm surge, and previous outages) was summed to calculate the total importance value for each zip code. Then the percent of total importance accounted for by each covariate was calculated. For each of the four covariates, we plotted the percent variable importance by zip code. We visually reviewed these plots to identify how the percent importance for each covariate differed geographically. The plots were also evaluated in light of the plots of the covariate values (Figure 4), so that the magnitude of the covariates was accounted for in evaluating the percent importance trends.

In order to better understand the predictive accuracy of the random forest model, a Quantile Regression Forest model was run on three selected zip codes. The zip codes were chosen so that different geographic areas in the state were represented. These zip codes include 71220 (north), 70129 (southeast), and 70525 (southwest).

3. RESULTS AND DISCUSSION

3.1 Variable Importance

The variable importance results for the Random Forest model with all covariates included are shown in Figure 1. Cumulative precipitation, wind speed, and previous outages are the most important variables, followed by population and hourly precipitation. All of the surge variables, along with wind gust duration had considerably lower variable importance. This differs from some previous work where wind gust duration was shown to be an important variable (e.g., Han et al. 2009) [16] and may be specific to this hurricane for which wind speeds were lower than for the hurricanes included in the Han et al. work [2].

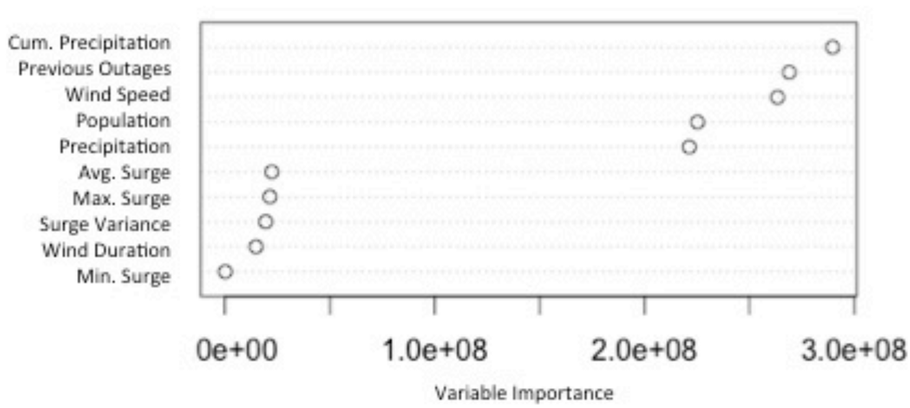

Figure 1: Variable Importance, all covariates included

Based on these results, four variables were selected as part of a reduced covariate set to be used for the remainder of the analysis. These variables were: cumulative precipitation, wind speed, previous outages, and maximum surge. Maximum surge depth was selected over average surge depth in each zip code because it had a clearer physical interpretation than the average surge depth yet had nearly the same importance score. Population was not included because the remainder of the analysis was done on an individual zip code basis wherein population is constant. The Random Forest model for the entire state was rerun with this reduced set of covariates. The resulting variable importance plot is included as Figure 2. In this model, the previous outages covariate has the highest variable importance, followed closely by cumulative precipitation and wind speed. Maximum surge has a lower importance, as should be expected since only a small portion of the state was impacted by storm surge.

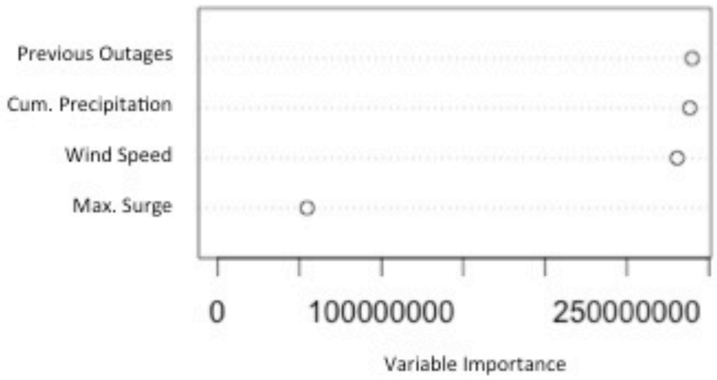

Figure 2: Variable Importance, reduced covariate set

3.2 Partial Dependence

Partial dependence plots were generated for the four covariates in the reduced set, and are provided as Figure 3. Partial dependence provides insight into the marginal impact of the covariate on the response variable, increase in outages.

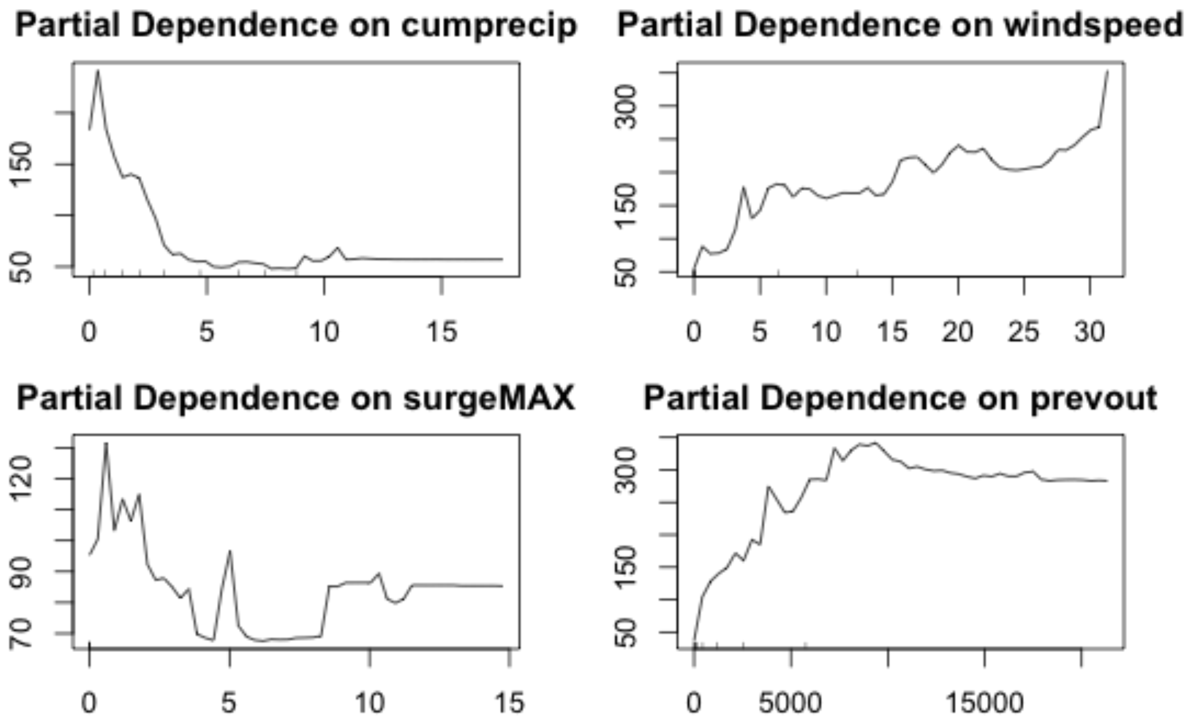

Figure 3: Partial Dependence Plots: a) partial dependence on cumulative precipitation, b) partial dependence on wind speed, c) partial dependence on maximum surge, and d) partial dependence on number of previous outages. The x-axis represents the value of the covariate and the y-axis represents the marginal influence of the covariate on delta outages.

The marginal influence of the cumulative precipitation covariate is highest for about 0 to 4 inches of precipitation. The marginal influence of wind speed generally increases with increasing wind speed. The influence of maximum surge is more variable, which may be due to the fairly low number of zip codes that experience storm surge. The marginal influence of the previous outages covariate increases up to around 10,000 outages, and then slightly decreases, since once a high number of outages occurs in a zip code, additional outages may be small in magnitude.

3.3 Geospatial Analysis

In order to analyze spatial trends across the state, we generated plots to get a sense of the magnitude of precipitation, wind speed, storm surge, and outages, and how the magnitude varied across the state. These plots are presented in Figure 4. Total precipitation (cumulative precipitation) was highest in the southeast part of the state, with more than 12 inches of precipitation recorded in some locations. Maximum wind speed was also highest in the southeast part of the state, where the hurricane made landfall. Maximum storm surge was highest in several zip codes bordering the Gulf of Mexico, as well as in several zip codes bordering the Mississippi River. The maximum numbers of power outages were observed in zip codes in the southeast, around New Orleans, where the population is greatest and the storm impacts were more significant.

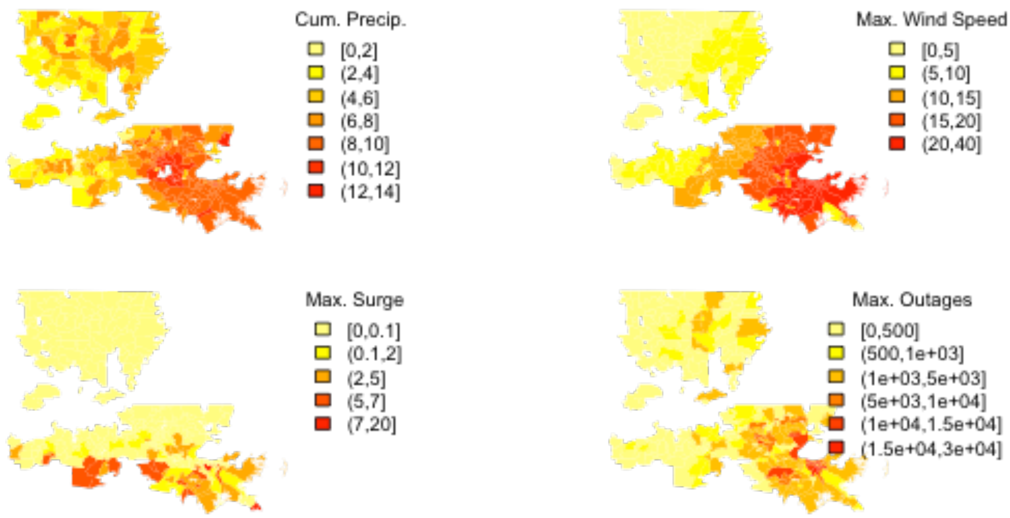

Figure 4: Covariate values for a) cumulative precipitation (inches), b) maximum wind speed (meters/second), c) maximum surge (meters), and d) maximum number of outages. Zip codes not colored are not part of the utility's service area.

Figure 5 illustrates the percent importance for cumulative precipitation, wind speed, maximum storm surge, and previous outages for all zip codes analyzed in Louisiana. In the northern part of the state, both cumulative precipitation and previous outages had the highest percent importance. Wind speed had the highest percent importance in the southeast part of the state. In the southwest portion of the state, percent importance varied from zip code to zip code, with cumulative precipitation, wind speed, and previous outages having varying importance. With the exception of a few zip codes, the percent importance for maximum storm surge was less than 30%, even in coastal areas.

These results indicate that the importance of covariates varies geographically. This is due to the storm's track and characteristics, but also potentially due to the interaction of other factors pertaining to the topography and power system. Both wind speed and cumulative precipitation were highest in the southeast, due to the storm's track; however, wind speed generally had greater importance in that area than precipitation. In the northern part of the state, where precipitation was moderately high, but wind speeds were low, precipitation was of greater importance. The previous outages covariate was generally more important in areas that had a lower maximum outages value.

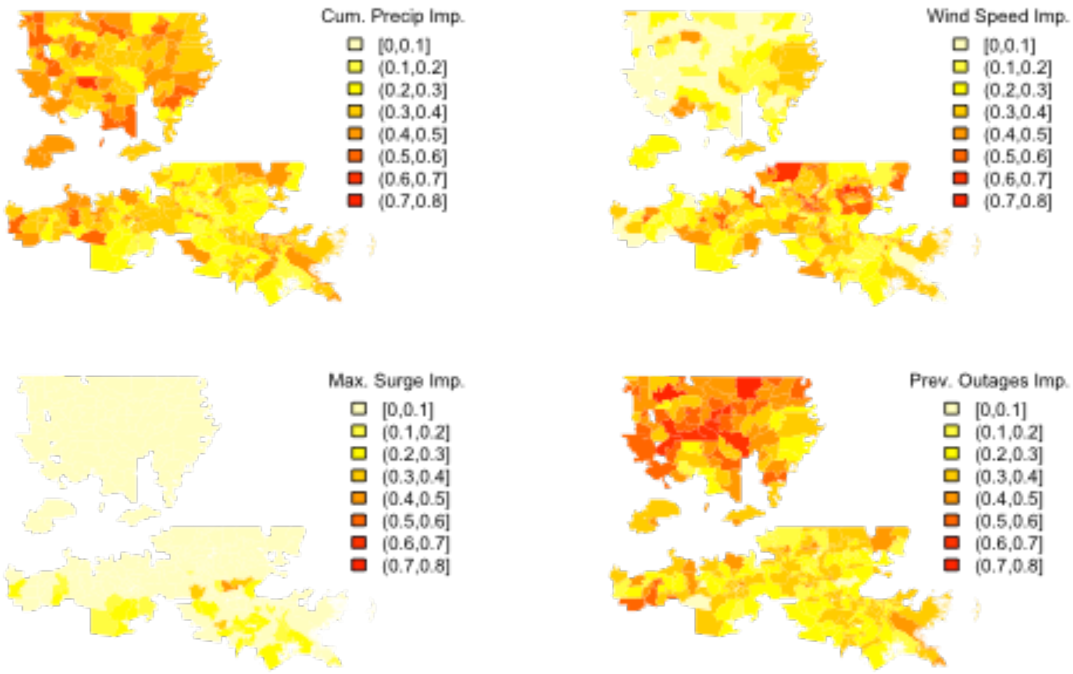

Figure 5: Percent Importance Plots for a) cumulative precipitation, b) wind speed, c) maximum surge, and d) previous outages

3.4 Quantile Regression Forest

We ran a Quantile Regression Forest model on three selected zip codes in order to better understand the predictive accuracy of the model. These zip codes include 71220 (north), 70129 (southeast), and 70525 (southwest). Plots of the Quantile Regression Forest results for these three zip codes are shown in Figure 6. These plots show the 90% prediction confidence intervals and whether predictions using out-of-bag data fall inside or outside of the prediction intervals. As shown on the plots, the majority of the predictions fall within the prediction intervals.

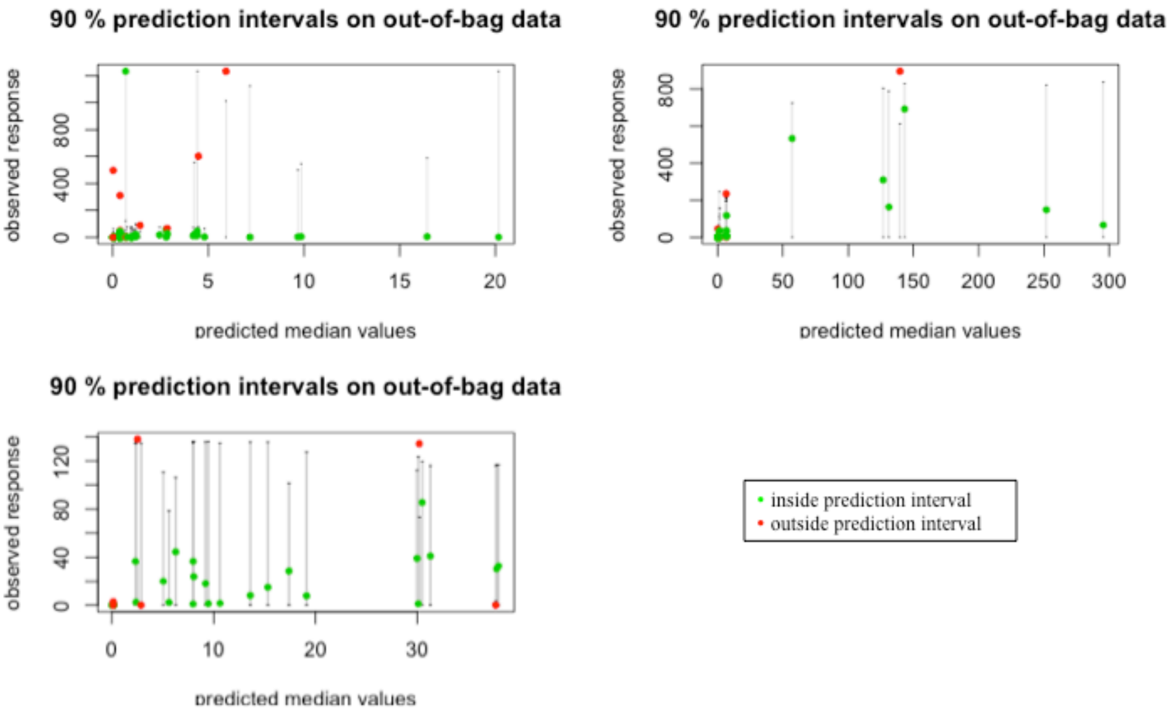

Figure 6: Quantile Regression Forest plots for a) zip code 70129, b) zip code 71220, and c) zip code 70546

Table 2 shows the range of predictions that fall between the 10% and 90% quantiles. For low values of delta outages (0 to 2), the coverage of the 90% interval is very low; the model has little reliability at the lowest level of delta outages. For middle of the range values of delta outages (2 to 75), the model confidence interval coverage is fairly high, ranging from 76% to 100% for a 90% interval for the three zip codes analysed. At the high end of the delta outages range (75+), the coverage accuracy varies significantly. This makes sense given the nature of power outages and the covariates used in the model. Very low increases in power outages are not likely well correlated to wind, precipitation, or previous outages, and are more likely caused by random events occurring at individual houses. Very high increases in power outages can sometimes be correlated with high precipitation or wind, but could also occur due to sudden problems in the power grid.

Table 2: Percent of Predictions within 80% Confidence Interval

Delta Outage Range	Zip Code 70129	Zip Code 71220	Zip Code 70546
0-2	0%	0%	3%
2-75	80%	100%	76%
75+	58%	N/A	100%

4. CONCLUSION

The purpose of this analysis was to provide insight on how rainfall and storm surge, along with wind, contribute to risk of power outages in hurricanes. By conducting a longitudinal study of outages at the zip code level, we were able to gain insight into the causal drivers of power outages during hurricanes. Our analysis showed that the drivers of power outages and the importance of the drivers can vary geographically. In Louisiana, during Hurricane Isaac, rainfall and previous outages were the most important covariates in the north, while wind was more important in the southeast. Rainfall, wind, and previous outages were all relatively important in the southwest. With the exception of a few zip codes, storm surge was generally not an important variable in predicting power outages. The reason the drivers vary geographically is likely due to characteristics of the location and of the storm. In areas where the highest wind speeds are experienced, wind is likely to be the most important covariate. Elsewhere, the importance of covariates differs geographically.

Acknowledgements

The authors would like to acknowledge Michael Gao, a Johns Hopkins University undergraduate researcher for his GIS assistance and members of Seth Guikema's research group at JHU for their assistance with collecting power outage data during Hurricane Isaac. This work was supported by a fellowship to Gina Tonn by the Environment, Energy, Sustainability and Health Institute (E2SHI) and the NSF IGERT Water, Climate, and Health program at Johns Hopkins University as well as by NSF grant number 1149460 from CMMI.

References

[1] "UPDATE 1-Entergy estimates Hurricane Isaac damage at $500 mln." *Chicago Tribune*. 18 Sept. 2012.
[2] S.R Han, et al. "Estimating the spatial distribution of power outages during hurricanes in the Gulf coast region." *Reliability Engineering & System Safety,* vol. 94.2, p. 199-210, 2009.
[3] R. Nateghi, S. Guikema, and S. Quiring. "Power outage estimation for tropical cyclones: Improved accuracy with simpler models." *Risk analysis,* 2013.
[4] S.D. Guikema, R. Nateghi, and S. Quiring. "Predicting Infrastructure Loss of Service from Natural Hazards with Statistical Models: Experiences and Advances with Hurricane Power Outage Prediction," in Proceedings, ESREL 2013, Amsterdam, October 2013.
[5] *Entergy View Outages*. Entergy. http://www.entergy.com/storm_center/outages.aspx, Accessed 27 Aug. – 5 Sept. 2012.
[6] National Climatic Data Center (NCDC). http://www.ncdc.noaa.gov, Accessed 16 Nov. 2012.
[7] J.C. Dietrich, M. Zijlema, J.J. Westerink, L.H. Holthuijsen, C. Dawson, R.A. Luettich, R. Jensen, J.M. Smith, G.S. Stelling, and G.W. Stone. "Modeling Hurricane Waves and Storm Surge using Integrally-Coupled, Scalable Computations," *Coastal Engineering*, vol. 58, p. 45-65, 2011.
[8] R. Luettich, and J. Westerink. "Formulation and numerical implementation of a 2D/3D ADCIRC Finite Element Model Version 4.46." http://adcirc.org/adcirc_theory_2004_12_08.pdf. Accessed 13 Nov. 2010.
[9] C.M. Ferreira, J. Irish, F. Olivera. "ArcStormSurge: Integrating GIS and Hurricane Storm Surge." *Journal of the American Water Resources Association*, 2013.
[10] N. Booij, R.C. Ris, and L. H. Holthuijsen. "A third generation wave model for coastal regions. Model Description and Validation." *Journal of Geophysical Research,* vol. 104, p. 7649-7666, 1999.
[11] S. Bunya, J. Dietrich, J. Westerink, B. Ebersole, J. Smith, J. Atkinson, R. Jensen, D. Resio, R. Luettich, C. Dawson, V. Cardone, A. Cox, M. Powell, H. Westerink, and H. Roberts. "A High-Resolution Coupled Riverine Flow, Tide, Wind, Wind Wave, and Storm Surge Model for Southern Louisiana and Mississippi. Part I: Model Development and Validation." *Monthly Weather Review*, p. 345-377, 2010.
[12] J. Dietrich, S. Bunya, J. Westerink, B. Ebersole, J. Smith, J. Atkinson, R. Jensen, D. Resio, R. Luettich, C. Dawson, V. Cardone, A. Cox, M. Powell, H. Westerink, and H. Roberts. "A High-Resolution Coupled Riverine Flow, Tide, Wind, Wind Wave, and Storm Surge Model for Southern Louisiana and Mississippi. Part II: Synoptic Description and Analysis of Hurricanes Katrina and Rita." *Monthly Weather Review*, p. 378-404, 2010.
[13] C. Mattocks, and C. Forbes. "A real-time, event-triggered storm surge forecasting system for the state of North Carolina", *Ocean Modelling*, vol. 25, p. 95-119, 2008.
[14] National Oceanic and Atmospheric Administration (2013) Atlantic basin hurricane database (HURDAT). http://www.aoml.noaa.gov/hrd/hurdat/, Accessed 08 Jul. 2013.
[15] H.E. Willoughby, R.W.R. Darling, and M.E. Rahn. "Parametric representation of the primary hurricane vortex. Part II: A new family of sectionally continuous profiles." *Monthly Weather Review,* vol. 134.4, 2006.

[16] S.R. Han, S.D. Guikema, and S. Quiring. "Improving the predictive accuracy of hurricane power outage forecasts using generalized additive models." *Risk analysis,* vol. 29.10, p. 1443-1453, 2009.
[17] *American Community Survey*. US Census Bureau. http://www.census.gov/acs/www, Accessed 7 Jul. 2013.
[18] T. Hasti, R. Tibshirani, and J. Friedman. The Elements of Statistical Learning; Data Mining, Inference and Prediction, 1st ed. New York: Springer, 2001.
[19] A. Liaw, and M. Wiener. "Classification and Regression by randomForest." R News, vol. 2(3), p. 18-22. 2002.
[20] N. Meinshausen. "Quantile regression forests." *The Journal of Machine Learning Research* 7, p. 983-999, 2006.
[21] N. Meinshausen. "quantregForest: Quantile Regression Forests." R package version 0.2-3, 2012.