IMPORTANCE OF TREE SPECIES AND PRECIPITATION FOR MODELING

HURRICANE-INDUCED POWER OUTAGES

A Thesis

by

CHRISTOPHER MICHAEL MADERIA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Steven M. Quiring
Committee Members,	Seth D. Guikema
	Daniel W. Goldberg
Head of Department,	David M. Cairns

August 2015

Major Subject: Geography

ABSTRACT

Hurricanes can be a major threat to electric power systems, often resulting in costly repairs and lengthy restoration times. In addition, power companies often lack the personnel required to restore power in a timely and efficient manner and must rely on outside assistance from other utility companies. Statistical power outage models, such as the Hurricane Outage Prediction Model (HOPM), provide estimates of outages at least 24 hours before a hurricane makes landfall. These models can greatly benefit utility companies by allowing for better allocation of resources and potentially shortening restoration times. This research will investigate the addition of two new variables, tree species and storm-derived precipitation, to the HOPM.

Tree species information was extracted for the service area of a major Gulf Coast utility company. Storm-derived precipitation was also extracted for the service area 24 hours before and after hurricane landfall. The model was then run for the service area with the new variables added, and results were generated that showed the impact of the new predictors on model performance.

Of the two predictors, tree species resulted in the greatest model improvement. Certain tree species, such as sweetgum, may be better predictors of outages than others. Storm-derived precipitation was also an important predictor of outages, particularly in urban areas. Precipitation amounts less than about 7 inches had the greatest impact on outages. Inclusion of tree species and storm-derived precipitation in future versions of

the HOPM may enhance model performance and, in turn, aid utility companies in their

goal of more efficient power restoration.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Among natural disasters hurricanes rank as one of the most deadly and destructive (Emanuel 2005). These violent storms often produce damage to homes, businesses, and infrastructure that can be disruptive. They can also render significant damage to electric power distribution systems, causing major disruptions to service. During Hurricane Katrina in 2005, 82% of customers in one of the largest utility companies along the Gulf Coast lost power, and some customers had to wait up to 12 days before their power was restored (Guikema et al. 2010). Hurricanes can severely damage an electrical power system; thus, predicting their impacts on the power grid prior to landfall will have a significant positive effect on pre-storm planning efforts for utility companies.

Utility companies generally lack sufficient personnel to rapidly restore power after major storm events such as hurricanes. Instead, through mutual aid agreements, they must call on other utility providers for assistance. Before doing so, the utility company making the request must estimate the demand and resources required to restore power quickly and effectively. Requesting more assistance than necessary will result in unnecessary expenditures, while requesting too little assistance will result in significantly longer power restoration times. As part of their pre-storm planning efforts,

utility companies must weigh the anticipated impacts from the storm and make an estimate on how much assistance they will need (Guikema et al. 2010).

Statistical, regression-based, modeling can be used to provide an estimate of a hurricane's impacts on the power grid prior to landfall; this enables crews to be placed in the areas of greatest impact ahead of time. It also allows for a better estimate of the extra resources the utility company may require to restore power quickly and efficiently. These models use data about power system performance during past hurricanes to predict the performance during an approaching hurricane (Guikema et al. 2010).

Past hurricane power outage models have incorporated environmental predictors such as land cover and antecedent precipitation; however, one of the primary limitations in these models, as identified by Liu et al. (2008), was their lack of tree-related variables such as number, type, age, and tree trimming frequency. Nateghi et al. (2014) incorporated tree trimming and determined it to be an important predictor. Liu et al. (2005) determined that precipitation in the week prior to hurricane landfall is a statistically significant predictor, and Han et al. (2009b) successfully used the Standardized Precipitation Index, a measure of deviation of precipitation from normal conditions, over varying time scales.

This thesis will investigate whether incorporating new tree species and precipitation variables increases the accuracy of the hurricane outage prediction model (HOPM). The hypothesis is that these new predictor variables will further improve model performance and predictive power.

1.1 Research objectives

The main purpose of this research is to evaluate whether the addition of two new predictor variables, tree species and 48-hour precipitation, increases the accuracy of the HOPM. To accomplish this purpose, the thesis has three main objectives:

1) Identify the most suitable version of the HOPM for evaluating the influence of tree species and 48-hour precipitation.
2) Determine if including tree species data in the HOPM will improve model accuracy.
3) Determine if including precipitation data in the HOPM will improve model accuracy.

The first objective will evaluate model error and outage prediction accuracy for all versions of the HOPM with and without the new variables. The results of this evaluation will then be used to select the most appropriate model for further analysis of tree species and 48-hour precipitation.

The second objective will investigate the impact of prevalent tree species on HOPM predictive accuracy. Various measures of accuracy and importance will be calculated to evaluate the impact of tree species on outage prediction. This process will not only determine if the tree species variable can improve model performance, but it may also help identify which tree species may be more likely to cause outages.

The third objective will involve collecting radar-derived, 48-hour precipitation totals, 24 hours prior to and following hurricane landfall. The radar-derived precipitation

data will be used as a proxy for the precipitation estimates predicted by numerical weather models. These will then be integrated into the HOPM to explore the impact of storm-derived precipitation on power outage prediction. Importance rankings and accuracy measures will be calculated using the same approach as Objective 2.

Introduction of these two new predictor variables into the HOPM will potentially lead to improvements in model performance and will aid in selection of the most influential variables to include in future versions of the model. This information will also benefit utility providers and emergency managers in their pre-storm preparations by providing a more accurate pre-landfall estimate of number and location of outages as well as identifying areas more prone to outages.

CHAPTER II

BACKGROUND AND LITERATURE REVIEW

2.1 Power outage prediction

Statistical models to forecast weather-related power outages typically integrate several predictors and use them to forecast outages or damage to electrical infrastructure (Table 2.1). The predictors that are common to many of these models include: historical and present storm conditions, power system characteristics, power system reliability, and environmental characteristics (i.e., land cover or land use). These models have been used to predict power outages and damage due to a variety of weather conditions, including ice, lightning, wind, and hurricanes.

Broström and Söder (2007) estimated the risk of power system failure (probability of outages) due to ice storms using a series of weather and component vulnerability models. Longer power lines are found to be more vulnerable than shorter lines with fewer segments (Broström and Söder 2007). Using a similar approach, DeGaetano et al. (2008) estimated ice accretion on distribution lines prior to the storm using a modified version of Jones' (1996) ice accretion model. In place of hourly surface observations, DeGaetano et al. (2008) used hourly forecasted values for temperature, precipitation amount, and wind speed from the Weather Research and Forecasting (WRF) model. This revised ice accretion model improves the underestimation problem found in earlier models.

Reference	Hazard type	Model type	Dependent variable	Predictors
Broström and Söder (2007)	Ice	Weather + component vulnerability	Outages	Ice accretion, precipitation, wind load
DeGaetano et al. (2008)	Ice	Weather + Ice accretion	Ice accretion	Temperature, precipitation, wind speed
Liu et al. (2007)	Ice	AFT (statistical)	Outages, restoration times	Outages (duration, total, start time), ice thickness, number of customers, population density
Liu et al. (2008)	Ice	Spatial GLMM (statistical)	Outages	Protective devices (number), ice thickness, land cover type, soil (drainage and depth)
Balijepalli et al. (2005)	Lightning	Monte Carlo simulation (statistical)	System reliability	Storm intensity/duration, lightning flash counts, outage rates
Zhu et al. (2007)	Lightning	Assessment-based; statistical	Outages	Past outages, weather station observations, lightning (intensity and corridor width)
Brown et al. (1997)	Wind	Monte Carlo simulation (statistical)	Outages	System reliability, wind speed and duration
Reed (2008)	Wind	Fragility analysis, GIS analysis	Outages	System reliability, max wind speed, peak wind gust, temp, precip
Cerruti and Decker (2011)	Various	GLM (statistical)	Power system damage	Forecasted weather conditions (various)
Li et al. (2010)	Various	Poisson regression (statistical)	Outages, damage	Forecasted severe weather conditions (various)
Zhou et al. (2006)	Various	Poisson regression + Bayesian network (statistical)	Outages	Forecasted weather conditions (wind, icing, lightning)

Table 2.1. Summary of studies modeling the impact of weather on power outages and damage to electrical infrastructure.

Liu et al. (2007) took a different approach and estimated power restoration times using statistical, accelerated failure time (AFT) models. The models were fitted using past outage, power system, and environmental data (Table 2.1). Restoration times were then estimated for each county or service area sub-region using accumulated outage duration times and number of customers. Estimated restoration times for the January 2004 ice storm are an average of 5% different than actual restoration times. Building off of the regression model used by Liu et al. (2005) to predict hurricane-induced power outages, Liu et al. (2008) took a slightly different approach by integrating a series of predictors into a spatial generalized linear mixed model (GLMM) to estimate outages occurring due to ice storms. The final model is found to over-predict outages due to ice storms. The number of protective devices and ice thickness had the greatest impact on outages. The same January 2004 ice storm was used to test both of these models (Liu et al. 2008).

Power outage models have also been run to estimate outages due to lightning. For example, Balijepalli et al. (2005) used a Monte Carlo simulation to generate distribution system reliability indices, with storm characteristics (i.e., storm duration, storm intensity, and lightning flash count) and outage rates as the inputs. Zhu et al. (2007) investigated patterns of outage occurrence using past outage and weather station observations for 49 storms; these data were then used to create a statistical model to predict numbers of outages. Over half of distribution system outages in the summer are found to be attributed to lightning, and a separate model is created using only the

lightning data to predict outages. Certain lightning intensities and corridor widths are found to have a higher correlation with power outages (Zhu et al. 2007).

Brown et al. (1997) selected storm events for analysis based on their wind speeds and created a Monte Carlo simulation to determine outage frequency and duration. The simulation was tested using wind data for Snohomish County, Washington. Momentary outages (less than several minutes) and long-term outages were classified separately, and momentary outages are found to compose a large portion of the total outages. Reed (2008) estimated outages due to wind for the Seattle area power distribution system. A fragility analysis was run for four winter storms using outage frequency and duration, minimum temperature, 24-hour precipitation, maximum wind speed, and peak wind gust data, and results were compared with data from other hurricanes and winter storms. Outage durations resulting from wind events are very similar for both winter storms and hurricanes; also, peak wind gust seems to be the best predictor of outage durations (Reed 2008).

Cerruti and Decker (2011) created a generalized linear model (GLM), a type of multiple linear regression model, to predict power system damage using forecasted surface weather. Surface weather for a given day was used to assign that day to a certain "weather mode" (i.e., thunderstorm, warm, mix, cold, heat, wind, none, or questionable), and a different mode of the model was run depending on the condition. When compared with other statistical models, the GLM seems to be most effective at predicting power system damage from weather forecasts (Cerruti and Decker 2011). A Poisson regression model was developed by Li et al. (2010) to predict the number of outages and the

amount of power system damage in response to a variety of severe weather conditions. A set of 19 major storm events were used to train the model, and the model is currently in use by an electric utility company in the northeastern U.S. for its distribution system. Zhou et al. (2006) combined both a Poisson regression and a Bayesian network model to estimate yearly outage rates for a power distribution system in Manhattan, Kansas. The Poisson model was run to predict the number of outages, while the Bayesian model was used to predict the probability of an outage. A variety of weather variables were used as predictors in the model; however, wind, icing, and lightning are determined to be the most influential (Zhou et al. 2006).

2.2 Hurricane-induced power outage prediction

For over a decade now, predictive models have been used to forecast power outages and damage to power infrastructure resulting specifically from hurricanes. Both non-statistical and statistical models have been used; however, most of these models are statistics-based. Non-statistical models tend to be either assessment-based or use some sort of fragility model. In the sections below, non-statistical models will be discussed first, followed by statistical models, including the HOPM. A summary of all non-HOPM models can be found in Table 2.2 below.

Reference	Model type	Dependent variable	Predictors
Davidson et al. (2003)	Assessment-based	Outages, customers affected	Maximum wind gust, antecedent precip, land cover type, power system
Reed et al. (2010)	Fragility analysis	Outages	Wind speed ratio, power system
Winkler et al. (2010)	Damage and fragility models	Outages, damage	Wind gust speed, tree wind-throw, terrain, network topology
Ouyang and Dueñas-Osorio (2014)	Four-part model	Outages, restoration times	Wind gust speed, tree wind-throw, land cover
Krishnamurthy and Kwasinski (2013)	Non-linear regression (statistical)	Outages, restoration times	Storm surge, storm size, max wind, wind exposure time
Liu et al. (2005)	Negative binomial regression (statistical)	Outages	Wind (max gust, duration, wind field), antecedent precip, land cover, tree type, soil drainage, hurricane indicator (wind speed and rainfall), power system
Liu et al. (2007)	AFT (statistical)	Outages, restoration times	Outages (start time), wind (max gust, duration), antecedent precip, land cover, number of customers, population density
Liu et al. (2008)	Spatial GLMM (statistical)	Outages	Wind (max gust, duration), antecedent precip, land cover type, soil (drainage and depth), protective devices (number)

Table 2.2. Summary of studies modeling the impact of hurricanes on power outages and damage to electrical infrastructure (Quiring et al. 2011).

2.2.1 Review of models

2.2.1.1 Assessment-based, hazard, and fragility models

Prior to the development of a hurricane power outage model, assessments of past storms can be performed to determine the influence of certain variables and their usefulness as predictors in a future model. Davidson et al. (2003) examined outage and other distribution system data from five hurricanes affecting the Carolinas. Various statistical analyses (i.e., correlation, t-tests, and scatterplots) were implemented to assess the importance of maximum wind gust speed, rainfall, and land cover type on the number, distribution, and duration of outages. Wind gust speed is found to have the most influence on the number of outages, while land cover type is found to have a significant influence on the spatial distribution of outages. A combination of the land cover data with reports from electric utility companies indicate that most power system damage results from trees falling on power lines (Davidson et al. 2003). In a different approach, Reed et al. (2010) used various restoration functions and the IEEE[1] performance index for outages to determine the impact of the wind speed ratio on outages and power system damage during Hurricane Rita. The wind speed ratio is the ratio of the maximum sustained wind speed for a parish to the ASCE[2] 2-minute wind speed for Louisiana (the region of study). Results of the analysis show that wind speed is the primary contributor to line damage (Reed et al. 2010).

---

[1] Institute of Electrical and Electronics Engineers
[2] American Society of Civil Engineers

A combination of hazard and fragility models has proven to be effective in modeling hurricane-induced power outages and power system damage with low amounts of error. In a continuation of the above study, Reed et al. (2010) conducted a fragility analysis comparing the wind speed ratio with fragility, the ratio of outages to total number of customers. Results from this analysis show that when the wind speed ratio was at 45%, then 50% of customers in that parish experienced outages (Reed et al. 2010). In Winkler et al. (2010) a hurricane damage prediction model was used in combination with multiple component fragility models to predict power system reliability during hurricanes. The hurricane damage model integrates the most significant predictors from (Han et al. 2009b), terrain and three-second wind gust speed, and associates them with a fragility model to produce probabilities of damage to components. Meanwhile, a series of component fragility models predict the likelihood of failure (outages) for the transmission and distribution networks using wind gust speed and tree type estimates for seven species. Model error for outage prediction is estimated to be ~15%. Power system reliability also correlates well with network topology and structure of the transmission system, and certain topologies, such as the ring mesh topology, have proven more resistant to damage from hurricanes (Winkler et al. 2010).

Ouyang and Dueñas-Osorio (2014) created a single comprehensive hurricane power outage model by combining four separate models: hurricane hazard, component fragility, power system performance, and system restoration. The hurricane hazard model uses the HAZUS software to generate storm scenarios for the study region, which are then applied to the other three models. The component fragility portion integrates the

12

fragility models from Winkler et al. (2010) for five power system components (substations, transmission lines, distribution lines, distribution nodes, and distribution circuits) to estimate outage probabilities for each storm. The power system performance portion then models power system response to component failures, and the system restoration model uses system resource quantities and restoration sequences (i.e., first transmission lines, then substations, then distribution lines) to estimate restoration times. The combination of these four models can also be used to generate a restoration curve relating the time after landfall to the number of customers with power (Ouyang and Dueñas-Osorio 2014).

2.2.1.2 Statistical models

Regression models, which are the main focus of this thesis, have been widely used to model hurricane-induced power outages. They are relatively easy to implement, have proven effective, and allow for the evaluation of a large quantity and variety of predictors, making it easy to improve model performance.

Krishnamurthy and Kwasinski (2013) used a non-linear regression model to generate county-level indices and curves for maximum number of outages, average outage duration, and restoration times. The curve for maximum number of outages shows the best fit, while the curve for restoration times has a significantly weaker fit. This could be due to regional variance in restoration practices. Storm surge, maximum sustained wind speed, storm size, and the exposure time to tropical storm force winds all

show promise as predictors, with storm surge being the strongest predictor in coastal areas (Krishnamurthy and Kwasinski 2013).

Negative binomial regression models have also proven effective for modeling hurricane-induced power outages. Liu et al. (2005) employed such a model to predict the number of outages at the zip code level. They find that negative binomial regression yields more accurate predictions than Poisson regression, most likely because negative binomial regression models take into account unequal variance among the predictors. The most influential variables appear to be the number of transformers (power system), the company affected (power system), maximum wind gust, and the hurricane indicator (see other predictors in Table 2.2). Rainfall and soil drainage may also be important, but only when trees lie close to power lines (Liu et al. 2005).

As mentioned above, Liu et al. (2007) used AFT models to estimate power restoration times for hurricanes and ice storms. The restoration curves for hurricanes show significant overestimation, especially when compared with those of the January 2004 ice storm, the other storm tested in this study. Despite that, however, 96% of customers had their power restored within the benchmark time that the model predicted 90% of customers would have their power back; for ice storms, this was 91% (Liu et al. 2007). The GLMM used by Liu et al. (2008) tends to over-predict outages due to hurricanes. Number of protective devices and maximum wind gust speed had the greatest impact on hurricane-induced power outages. The same hurricane (Hurricane Charley) and ice storm (January 2004) were used to test both models (Liu et al. 2008).

14

2.2.2 Evolution of the HOPM

      The HOPM is a statistical regression model that utilizes hurricane,

environmental, and power system predictors to estimate the quantity and spatial

distribution of power outages prior to landfall (Quiring et al. 2014). Over time these

models have become increasingly accurate, but they require a large set of input variables

(Nateghi et al. 2014). Examples of each type of variable can be seen in Table 2.3, and a

summary of the variables included in each version of the HOPM can be found in Table

2.4. Power system variables include the length of distribution line and the number of

poles, switches, transformers, and customers in a given geographical area. When

combined, the power system variables serve as a proxy measure for the exposure of the

power system. Environmental variables may vary depending on model version, but can

include soil moisture, long-term antecedent precipitation, or land cover type (Quiring et

al. 2011; Quiring et al. 2014). The model output is a forecast of the number and spatial

distribution of power outages within the utility provider service area.

Hurricane	Environmental	Power system
Saffir-Simpson category	Soil moisture	Number of poles
Minimum central pressure	Land cover type	Number of transformers
Duration of strong winds	Antecedent precipitation	Number of switches
Maximum wind gust speed	Tree trimming frequency	Number of customers

Table 2.3. Common predictors used in the HOPM (Nateghi et al. 2014; Quiring et al. 2011).

Reference	Model type	Dependent variable	Hurricane	Power system	Environmental
Han et al. (2009b)	Negative binomial GLM	Outages	Wind (max gust, duration of max winds, max wind radius), time since last landfall, central pressure	Transformers, poles, lines, switches, customers	Land cover, soil moisture, antecedent precipitation
Han et al. (2009a)	GAM	Outages	Same as Han et al. (2009b)	Same as Han et al. (2009b)	Same as Han et al. (2009b)
Guikema et al. (2010)	BART/CART + GLM/GAM	Damage to poles and transformers	Wind (max gust, duration of max winds)	Transformers, poles, primary lines, switches	Han et al. (2009b) + elevation, slope, topography
Quiring et al. (2011)	CART	Outages	Han et al. (2009b)	Han et al. (2009b)	Han et al. (2009b) + soil properties, topography
Guikema and Quiring (2012)	CART + Poisson GAM	Outages	Han et al. (2009b)	Han et al. (2009b)	Han et al. (2009b)
Nateghi et al. (2014)	Random forest	Outages	Wind (max gust, duration of max winds)	Number of customers	Fractional soil moisture (at two depths), tree trimming
This thesis	**Random forest**	**Outages**	**Nateghi et al. (2014) + storm-derived precip**	**Nateghi et al. (2014)**	**Nateghi et al. (2014) + tree species**

Table 2.4. Versions of the HOPM showing all three types of predictors (Quiring et al. 2011).

The first version of the HOPM predicted power outages using negative binomial GLMs, a type of regression-based model that permits analysis of count data (Guikema et al. 2010). This version resembles the model of Liu et al. (2005), but replaced the indicator variables with a more extensive variable set that could be measured prior to landfall. Its predictions were based on maximum wind gust, duration of strong winds, time since the last hurricane, radius of maximum winds, and central pressure deficit (Han et al. 2009b; Quiring et al. 2014). Soil moisture levels from three layers of soil and the Standardized Precipitation Index (SPI), representing antecedent precipitation, were also incorporated into this version of the model. One disadvantage of this model is that it overestimates the number of outages in urban areas and underestimates outages in rural regions (Nateghi et al. 2014).

Han et al. (2009a) improved on the accuracy of the first version of the HOPM by using generalized additive models (GAMs). Unlike GLMs, GAMs allow for nonlinearity and, thus, can better fit the power outage data, which is important for model development and testing (Guikema et al. 2010). GAMs were found to more accurately predict the number and spatial distribution of outages and overcame many of the over-prediction issues found in the GLMs (Han et al. 2009a; Quiring et al. 2014). They may also give a better idea of overall system response (Han et al. 2009a).

Guikema et al. (2010) then developed a new series of models that combined both the GLMs and the GAMs and added to them two data mining approaches, classification and regression trees (CART) and Bayesian regression trees (BART). CART uses a single tree to determine the relationship between the explanatory (i.e., hurricane and

17

environmental predictors) and response (i.e., number of outages or number of poles damaged) variables; whereas, BART constructs multiple, smaller trees to determine the same relationship. This new approach outperformed both of the previous regression-based approaches and showed that the BART and CART models had significantly more accurate predictions than the GLM and GAM models. A limitation of this type of statistical model is their difficulty in handling datasets containing many groups of zeros, or zero-inflated data (Quiring et al. 2014).

Similar to Guikema et al. (2010), Quiring et al. (2011) used a regression tree approach but used only CART instead of BART, CART, and a Poisson GAM. A large number of soil and topographic predictors were also entered into the model. The focus was on understanding the role of soil moisture and topography in predicting power outages. The results indicate that the addition of soil and topographic variables do not significantly improve predictive accuracy. One major finding was that some land cover variables could be used as proxies for power system data. This may be an effective way to generalize outage models and could be especially useful in areas where power system information is not available.

The next in the series of the HOPM models was created by Guikema and Quiring (2012) and is a two-stage model that combines both CART and a Poisson GAM to predict the number of power outages. The CART first predicts where the outages will occur; then the Poisson GAM estimates the number of outages in each of those locations. This model has relatively strong predictive accuracy (Quiring et al. 2014) and offers a better alternative for analyzing zero-inflated data (Guikema and Quiring 2012).

The last in the series of the HOPM models was used by Nateghi et al. (2014) to predict power outages across the service areas of two Gulf Coast states. This version builds off previous versions by using hurricane predictor variables of maximum wind gust and duration of strongest winds; however, it greatly reduces the number of environmental variables. Soil moisture from the 2001 National Land Cover Database (NLCD), SPI, and mean annual precipitation, and land cover type are all carried over from previous models, but all other environmental variables are removed. All other hurricane variables are also removed except for the two mentioned previously. A tree trimming factor, based on prior tree trimming history, is added as an environmental variable, and number of customers is added as a power system variable (Nateghi et al. 2014). Instead of a regression-based model such as GLM or GAM, Nateghi et al. (2014) used random forest, which is a data-mining technique similar to BART and CART. Random forest develops a large number of regression trees from data resampling; the final prediction is the average of predictions from all of the trees. A holdout analysis was used to test the ability of the model to predict outages. For each model (i.e., model with all variables, model with only the six variables) a portion of the data were held out, while the remaining data was used to train the model. The hold out sample was then used to make predictions and test the model. Results from testing this version of the HOPM show that it can explain much of the variance in outages with R^2 values of 0.70 for State One and 0.85 for State Two. Results also show that six fundamental variables are able to provide nearly as accurate an estimate as the model with all predictor variables. These six fundamental variables include three-second wind gust speed, duration of strongest

winds (above 20 m/sec), total number of customers per grid cell, tree trimming, and fractional soil moisture (Nateghi et al. 2014). This thesis uses this latest version of the HOPM and adds tree species and storm-derived precipitation (Table 2.4).

2.3 Tree species

Previous versions of the HOPM have included land use and land cover as predictors, but have not gone as specific as tree species. However, research has shown that tree species can be an important indicator of tree fall during a hurricane, which in turn can have a significant impact on power outages in that immediate region. According to Davidson et al. (2003), tree species may be an important predictor because certain types are more prone to breakage or have shallower root systems.

2.3.1 Prior usage in power outage models

A limited number of previous studies have incorporated tree species into hurricane power outage prediction models. As part of a hurricane power outage model, Winkler et al. (2010) developed a flying debris model using tree species, wind intensity, and tree diameter at breast height (DBH). The model estimates the probability of an outage due to flying debris in the vicinity of a power line. Each grid cell is randomly assigned one of seven local tree species to simulate natural variations in tree population (Winkler et al. 2010). Ouyang and Dueñas-Osorio (2014) applied a similar method to

estimate the probability of tree wind-throw (trees being uprooted or broken by the wind), except land cover type is used instead of tree species.

Liu et al. (2005) integrated tree type information, along with other data such as maximum wind gust, land cover type, and soil drainage, into a regression model to assist in predicting power outages in the Carolinas. Tree type data for 22 species were obtained from the U.S. Forest Service at a 1 km resolution. At this resolution, tree species is not a significant predictor of outages, and other high resolution data such as land cover and soil drainage have low influence as well; however, a larger grid cell size may change those results (Liu et al. 2005). In a different approach, Han et al. (2009a) used mean annual precipitation as a proxy for variations in tree species in a hurricane power outage model. This thesis will evaluate the importance of tree species as a predictor in hurricane power outage models.

Other research has shown that trees do play a significant role in causing outages during hurricanes. For example, according to outage-cause data from Duke Power, trees accounted for about half of all outages in the service area during Hurricanes Opal, Fran, and Floyd, while each of the other causes accounted for 10% or less (Davidson et al. 2003). Also, the amount of tree trimming is considered to be one of the top six predictors in the latest version of the HOPM (Nateghi et al. 2014).

Other tree variables, such as tree abundance and tree trimming, have also been integrated into power outage prediction models. For example, Maliszewski et al. (2012) discovered that the combined interaction between bird abundance, vegetation abundance,

and overhead lines was the second most important factor in predicting distribution

system outages across Phoenix, Arizona. Reed (2008) found that over 85% of outages

during the December 1999 winter storm in France were due to trees falling on lines. In

addition, an average of 46% of outages during three winter storms in Seattle,

Washington were due to trees (Reed 2008). Also, Guikema et al. (2006) found that

increased tree trimming frequency in the Duke Power service area led to a reduction in

outages under normal (non-storm) conditions. In a study done by Simpson and Van

Bossuyt (1996), tree failure was the primary cause for 40% of outages preventable by

regular tree trimming.

2.3.2 Susceptibility to wind

There are many studies that have evaluated the relationship between various tree

attributes (i.e., stand height/condition, basal area, age, health) and wind resistance. This

literature review will focus solely on the relationship between tree species and wind

resistance, particularly with regard to hurricanes.

In a study done by Kupfer et al. (2008) following Hurricane Katrina, tree species

was found to be the third most important variable for determining forest damage in

Southern Mississippi. Merry et al. (2009) found that tree species and tree species

response to storm surge, among other variables, was critical in determining the amount

of forest damage that occurs during a hurricane. They also found that stem strength is

highly dependent on tree species (Merry et al. 2009). Garrigues et al. (2012) discovered

that tree damage during Hurricane Katrina was almost entirely related to tree stand

characteristics, including tree species, and not to storm-related factors, such as wind

speed or storm track.

2.3.2.1 Factors influencing wind resistance

Due to similarities between study area, study storms, and tree species included,

the findings from Duryea et al. (2007) are the primary source of wind resistance data

used for comparison in this thesis. Duryea et al. (2007) studied the impact of winds from

eight hurricanes on urban tree species in Florida. Each of these storms made landfall

between 1992 and 2005, and two of the storms, Dennis and Ivan, are also included in the

HOPM. Urban tree damage was measured within 3 to 6 days after the storm made

landfall; neighborhoods were randomly selected on the strong side of the storm. In total

100 neighborhoods and 18,200 trees were sampled; results were then compared with a

survey of arborists, scientists, and urban foresters in Florida, which contained their

rankings for wind resistance of southeastern United States coastal plains tree species. For

Hurricane Ivan, Duryea et al. (2007) found that tree species with the highest survival

(percent standing after the hurricane) included sand live oak, American holly, southern

magnolia, live oak, wax myrtle, **sweetgum**[3], crape myrtle, **dogwood**, and sabal palm.

For Hurricanes Erin and Opal (1995), dogwood, live oak, sabal palm, sand live oak, and

---

[3] Bolded tree species indicate those present in the HOPM.

southern magnolia also had the highest survival. **Laurel oak** generally had poorer survival than the other oaks (Duryea et al. 2007).

Branch loss may also be an important measure of resilience to hurricane winds. Species with the least branch loss included crape myrtle, **loblolly pine**, American holly, and tulip poplar. Large trees generally lost the most branches, followed by medium trees and then smaller trees. Also, pines tended to snap, while broadleaf trees were more susceptible to uprooting. Trees with large amounts of branch loss may not be considered healthy trees, so survival was recalculated, with standing trees having 50% or greater branch loss being classified as dead (Duryea et al. 2007). Using the recalculated survival for Hurricane Ivan, tree species used in the HOPM rank as follows, from greatest to least survival: sweetgum, dogwood, slash pine, laurel oak, water oak, loblolly pine, and longleaf pine.

Combining the survival and branch loss results from Duryea et al. (2007) with results from the survey and from scientific literature, Duryea et al. (2007) created a wind resistance classification of southeastern U.S. urban tree species (Table 2.5). Duryea et al. (2007) note that this list must be used with caution because there are many factors (i.e., hurricane wind speed, precipitation, soil, tree age and health) that can affect resilience. According to the classification table, tree species in the HOPM with the highest wind resistance include: dogwood; medium-high: sweetgum and swamp chestnut (same genus as chestnut oak); medium-low: white oak, slash pine, longleaf pine, loblolly pine; and lowest: laurel oak and water oak.

Highest	Medium-high	Medium-low	Lowest
Florida scrub	Florida sugar maple	Boxelder	Pecan
Dogwood (60%)	Japanese maple	Red maple	Tulip poplar
Dahoon holly	River birch	Silver maple	Carolina
Inkberry	Ironwood	Sugarberry	Bradford pear
American holly	Pignut hickory	Hackberry	Southern red
Yaupon holly	Mockernut hickory	Camphor	Laurel oak (4%)
Crape myrtle	Red bud	Loquat	Water oak (8%)
Southern magnolia	Fringe tree	Silverdollar	Chinese tallow
Sand live oak	Common persimmon	Green ash	Chinese elm
Turkey oak	White ash	Red mulberry	Southern red
Myrtle oak	Sweetgum (50%)	Wax myrtle	Leyland cypress
Live oak	Sweetbay magnolia	Redbay	Sand pine
Podocarpus	Saucer magnolia	Sycamore	Spruce pine
Sparkleberry	Water tupelo	Black cherry	Washington fan
Baldcypress	Black tupelo	White oak (55%)	
Pondcypress	American	Willow oak	
Pindo or jelly	Chickasaw plum	Weeping willow	
Canary island date	Swamp chestnut (43%)	American elm	
Date	Shumard oak	Slash pine (25%)	
Cabbage, sabal	Post oak	Longleaf pine (56%)	
	Winged elm	Loblolly pine (20%)	

Table 2.5. Wind resistance of southeastern U.S. tree species (Duryea et al. 2007); species included in the HOPM are in black. Percentages indicate the number of trees in a survey sample with a high level of wind resistance.

In addition to branch loss, other tree species characteristics may also play a

significant role in a species' resistance to hurricane winds. For example, Duryea et al.

(2007) found that a higher amount of defoliation (leaf loss) can lead to a higher rate of

survival following a hurricane; this was true for both survival and recalculated survival

in their study. The tendency for certain species to defoliate during periods of strong wind reduces their likelihood of becoming uprooted or having branches broken and may give them a distinct advantage over other species in terms of survival (Duryea and Kampf 2007). Table 2.6 below compares certain tree species characteristics (breakage, uprooting, and defoliation) with tree species survival rates during six hurricanes (Andrew, Erin, Opal, Charley, Jeanne, and Ivan) from Duryea et al. (2007). Breakage and uprooting are ranked from least to greatest, and defoliation and survival are ranked from greatest to least. Darker shading indicates higher amounts, while lighter shading indicates lower amounts. Rankings for breakage and uprooting were derived from Barry et al. (1993) and Xi and Peet (2008) for southeastern U.S. forests. Rankings for defoliation were derived from Gresham et al. (1991) and Duryea et al. (2007) for Hurricanes Hugo and Ivan, respectively. Generally tree species with lower amounts of breakage and uprooting, and a higher amount of defoliation, had a higher rate of survival during a hurricane. Sweetgum and dogwood both follow this pattern fairly well, with the exception that dogwood may have a higher risk of being uprooted. Many of the pines tend to have much lower survival rates due to more broken branches, a greater chance of uprooting, and greater leaf retention. Survival may also be influenced by the lifespan of a tree species; older trees tend to be more susceptible to disease and more prone to breakage during strong winds. Sweetgum ranks among the longer-living species, with a lifespan of over 100 years, while laurel oaks are fairly short-lived, with lifespans generally under 50 years Also, urban trees may have shorter lifespans than trees in forested areas (Duryea and Kampf 2007).

Breakage	Uprooting	Defoliation	Survival
Sweetgum	Sweetgum	Dogwood	Dogwood (92%)
Dogwood	White oak	Sweetgum	Sweetgum (90%)
Water oak	Chestnut oak	Laurel oak	Laurel oak (87%)
White oak	Longleaf pine	Loblolly pine	Slash pine (85%)
Chestnut oak	Slash pine	Water oak	Longleaf pine (75%)
Virginia pine	Loblolly pine	Longleaf pine	Loblolly pine (74%)
Longleaf pine	Water oak		Water oak (72%)
Slash pine	Dogwood		
Loblolly pine			

Table 2.6. Tree species characteristics (Duryea et al. 2007; Gresham et al. 1991; Stanturf et al. 2007; Xi and Peet 2008) vs. Survival (Duryea et al. 2007), for HOPM species only. Darker shading indicates higher probabilities. Percentage values are the average % survival of a species during six hurricanes (Andrew, Erin, Opal, Charley, Jeanne, and Ivan).

2.3.2.2 Wind resistance of specific tree species

As previously mentioned, the nominal class rankings in Table 2.5 are derived from survival statistics after eight Florida hurricanes, five of which struck the Gulf Coast (Erin, Opal, Charley, Ivan, and Dennis), combined with data from a survey of arborists and other scientists. The percentage values in the table represent the number of trees in a sample with a high level of wind resistance and are derived solely from the survey data.

Some discrepancies can be seen between the two datasets, especially for longleaf pine and white oak. For longleaf pine this may be because, although pines are very sensitive to wind, they may not show immediate signs of damage; instead, death may occur over a period of weeks, months, or even years (Duryea and Kampf 2007). Duryea

and Kampf (2007) found this to be the case during Hurricane Charley when 48% of longleaf pines standing immediately after the storm had died three months later. In addition, a publication produced by the Alabama Cooperative Extension System, lists longleaf pine as having medium-high wind resistance (Tilt et al. 2006). Another possible explanation for this anomaly would be the predominant location in which each of these species is found. In their study of the DeSoto National Forest in southern Mississippi, Kupfer et al. (2008) indicated that pines (especially longleaf, loblolly, shortleaf, and slash pines) generally prevail across upland areas, while oaks and sweetgum dominate the bottomlands. This suggests that pines may have greater exposure to higher wind speeds due to their prevalence in elevated areas such as hilltops, while oak and sweetgum may have reduced exposure to wind. In the same study, however, Kupfer et al. (2008) found that plots with more hardwoods are prone to greater damage; this would then suggest that hardwoods, like oak and sweetgum, would be more prone to damage power lines. Again, this just goes to show that no list of wind-resistant tree species is exact, as other factors such as tree health, topography, or geographic location may play a significant role as well.

Fredericksen et al. (1993) took a different approach and focused on testing wind resistance of loblolly pines using simulated wind stress. Broken limbs are found to be more the result of stem failure than uprooting. Also, longleaf pine may be more resistant to hurricane winds than other species of pine. In a study done by Johnsen et al. (2009) following Hurricane Katrina, stands of longleaf pine suffered less mortality than both slash pine and loblolly pine, with loblolly pine having the highest mortality rate.

Variation in mortality among species were apparent even when taking other factors, such as soil or topography, into account (Johnsen et al. 2009). This supports the finding of Gresham et al. (1991) that identified longleaf pine as suffering less damage than loblolly pine during Hurricane Hugo and the finding that loblolly pines may be more susceptible to wind-throw than other pines (Merry et al. 2009).

In a study done by Conner et al. (2002) sweetgum was found to be more susceptible to wind damage than damage from flooding, along with American elm, ash, red maple, and several species of oak. This supports the findings of Duryea (1997) who determined that sweetgum holds up moderately well to hurricane winds and experienced a branch loss of over 50% during Hurricanes Erin and Opal. Also, during a Texas tornado, sweetgum topped the list of survivors, but also had the most branch damage (Glitzenstein and Harcombe 1988). Despite significant branch loss, sweetgum is ranked by Duryea et al. (2007) as having a medium-high resistance to wind and has been ranked similarly by other studies (Duryea et al. 2007). This relatively high ranking is likely due to its strong root system combined with short, firm branches and long, skinny petioles (stalks that connect the leaves to the branches) that detach easily from the branches during strong winds (Duryea et al. 2007).

Dogwood also has a fairly high rate of survival during exposure to hurricane winds, probably due to its low amount of branch breakage and tendency to defoliate. During Hurricanes Camille and Hugo, dogwoods were found to be one of the most easily uprooted species (Duryea 1997; Gresham et al. 1991); however, Duryea (1997) found

that to be quite the contrary during Hurricanes Erin and Opal, when the majority of dogwoods remained standing.

Among the oaks, live oak appears to have the best rate of survival during hurricanes, probably due to its longer life span, over 100 years, and tendency to defoliate during periods of strong winds. During Hurricane Ivan, defoliation of live oaks was positively correlated with higher wind resistance (Duryea and Kampf 2007). Also during Hurricane Ivan, live oak and sand live oak had a significantly higher rate of survival than laurel oak and water oak. A similar trend was observed for Hurricanes Erin, Opal, and Dennis which also affected the Florida panhandle (Duryea et al. 2007). Laurel oaks and water oaks reside among the less wind-resistant species due to their shorter life span, tall height, and shallow root system, all of which result in a higher risk of stem breakage and uprooting (Duryea and Kampf 2007; Gresham et al. 1991; Hook et al. 1991). Water oak also tends to grow in soils with poor drainage, which may increase its risk of being uprooted (Hook et al. 1991). Along with pines, laurel and water oaks were found to cause the most property damage during Hurricane Ivan (Duryea et al. 2007).

2.4 Precipitation

The version of the HOPM developed in this thesis includes storm-related precipitation. According to Quiring et al. (2011), heavy precipitation occurring during the hurricane may result in inland flooding, thus increasing the number of outages and time to restore power. During Hurricane Katrina, flooding in New Orleans was a major

cause of outages and damage to electrical infrastructure, especially substations and other ground-based equipment (Kwasinski et al. 2009).

Other hurricane power outage models, including previous versions of the HOPM, have included antecedent precipitation as a predictor. Although results vary, antecedent precipitation, especially when combined with other factors, has generally proven to be an important predictor of hurricane-induced power outages.

Han et al. (2009b) and Han et al. (2009a) both used the SPI as a predictor in their hurricane outage models, while Nateghi et al. (2011) identified the SPI and soil moisture the day before landfall as two of the top 14 predictors in their fitted BART model. In the most recent, reduced version of the HOPM used by Nateghi et al. (2014) and in this thesis, fractional soil moisture is identified as one of the six key variables in predicting hurricane-induced power outages.

On the contrary, some studies have shown that antecedent precipitation has very little, if any, effect on hurricane-induced power outages. For example, Davidson et al. (2003) showed that rainfall measured in the seven days prior to landfall had a very weak correlation with outages and restoration times. Also, in the HOPM of Quiring et al. (2011), SPI does not rank in the top 10 variables and therefore is not included in the final version of their model.

It is often the combination of precipitation with other factors, such as tree species or wind, which makes it an important predictor of outages. In a study done over Alberta, Canada, Shen and Koval (1999) found that more outages are due to wind and lightning,

but that it is the combination of wind, trees, rainfall, and soil saturation that leads to increased outages. According to Liu et al. (2005), rainfall or flooding may influence hurricane-induced power outages only when trees are in close proximity to power lines. Guikema and Quiring (2012) integrated long-term precipitation and soil moisture into their hurricane outage prediction models and suggested that soil moisture may be important to determining tree and pole stability during a hurricane. Similarly, Han et al. (2009a) incorporated mean annual precipitation and SPI into their model and proposed that wet conditions may increase the number of downed poles and trees during hurricanes, while unusually dry conditions may result in the snapping of branches (Han et al. 2009a; Nateghi et al. 2011). Nateghi et al. (2011) reported that precipitation may play a role in outage duration due to delays in crews getting to the outage site.

These findings suggest that electric utilities can be impacted in two ways by precipitation. First, rainfall can saturate the ground which can lead to trees and poles falling, especially when combined with wind and already saturated soils. Second, rainfall can lead to flooding, which can cause damage to ground-level infrastructure and delay repair crews from getting to the outage site. Precipitation often works in combination with other factors to cause outages.

CHAPTER III

DATA AND METHODS

3.1 Study region

The region used to test this version of the HOPM is the State A service area of a major Gulf Coast utility provider. The utility provider has divided their service area into 15,213 grid cells across four states, each having a size of 3.66 km (12,000 feet) by 2.44 km (8,000 feet) (Quiring et al. 2014). The State A portion consists of 6,681 grid cells, which represent approximately 44% of the entire service area, or 59,615 square kilometers (Figure 3.1). Most of the grid cells are concentrated near urban areas or where larger numbers of power system equipment (i.e., distribution lines, transformers) are located. Approximately 56% of the state does not fall within a grid cell. These gaps in grid cell coverage indicate areas covered by another utility provider or areas lacking power system equipment; usually these are areas with little or no population.

The State A portion of the utility service area is chosen as a proxy for the entire service area to facilitate comparison with the version of the HOPM used by Nateghi et al. (2014). It is also centrally-located within the service area and lies closest to the track of three of the four hurricanes used to evaluate this version of the HOPM (Figure 3.2). Using just one state also allows for greater efficiency in both the data extraction and modeling processes. Additionally, using the State A portion of the service area facilitates

comparison with prior versions of the model, which evaluated several states, including

State A (Han et al. 2009a; Han et al. 2009b; Nateghi et al. 2014; Quiring et al. 2011).
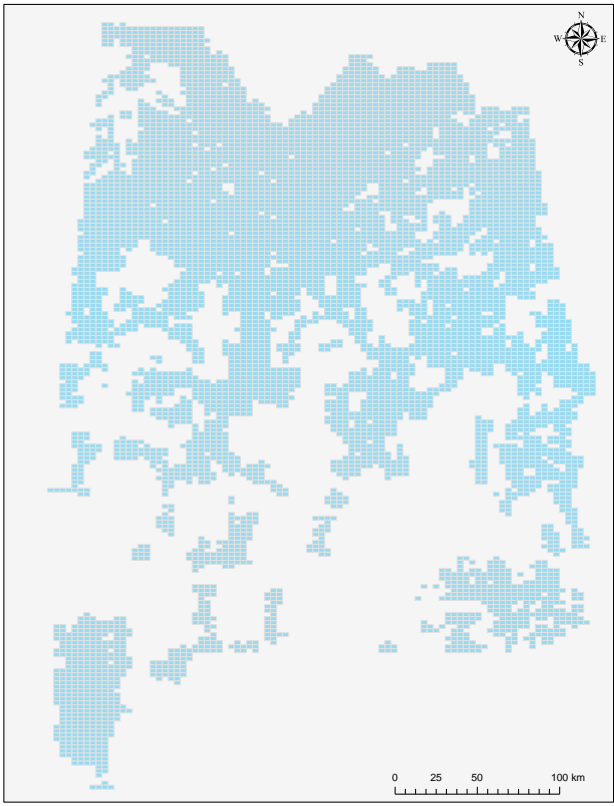
Figure 3.1. State A portion of utility company service area.
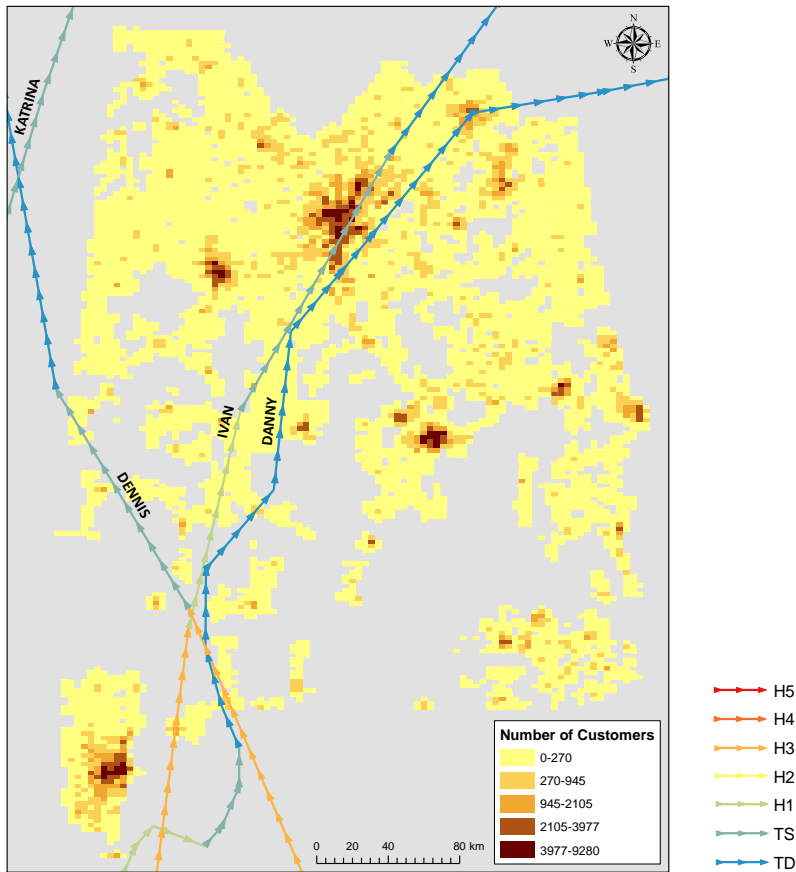
Figure 3.2. State A service area showing hurricane tracks and the number of customers.

3.2 Hurricanes of interest

For consistency with previous versions of the HOPM, this thesis will include data from four hurricanes (Figure 3.2): Danny (1997), Ivan (2004), Dennis (2005), and Katrina (2005). Georges has been used in previous versions of the model, however, it will not be included here due to major inconsistencies with the precipitation data (see section 3.4 for details).

Danny formed near the coast of Louisiana and moved slowly east-northeastward, eventually making landfall near the Mississippi River delta as a category one hurricane. A second landfall, and the point of reference for this study, was made near the mouth of Mobile Bay in State A on July 19th, 1997. Following its second landfall, Danny continued to meander slowly northward over State A for several days, gradually weakening to a tropical depression (Pasch 1997).

Ivan originated off the west coast of Africa and reached category five strength in the Caribbean, but weakened to a category three storm just prior to landfall near Gulf Shores, State A on September 16th, 2004. By September 17th, Ivan was a tropical depression centered over the northeastern part of State A (Stewart 2004).

Dennis also originated off the west coast of Africa and eventually made landfall in Cuba as a category four storm. After departing Cuba, Dennis continued its trek westward and northwestward, ultimately making landfall on Santa Rosa Island, just east of Pensacola on the Florida panhandle, as a category three storm (Beven 2005).

Katrina developed over the Bahamas and made its first landfall on the southeastern Florida peninsula as a category one storm. As Katrina reemerged over the Gulf of Mexico, it rapidly strengthened into a dangerous category five storm as it approached the Central Gulf Coast. A second landfall was made along the Mississippi Delta near Buras, Louisiana on August 29th, 2005 as a strong category three storm, followed by a final landfall near the Louisiana/Mississippi state line, which will be the point of reference for this study. Of all the hurricanes included in the HOPM, Katrina

was by far the strongest and most destructive, with significant impacts extending

hundreds of miles from the center (Knabb et al. 2005).

3.3 Datasets

3.3.1 Tree species

The host-tree species dataset from the 2012 National Insect and Disease Risk

Map (NIDRM) was selected to represent tree species for this thesis. This dataset, in

conjunction with other data layers, is used by the Forest Health Technology Enterprise

Team (FHTET) of the USDA Forest Service to model tree species risk due to insect,

disease, and other hazards (Krist et al. 2012). This particular dataset was chosen for use

in this thesis primarily because of its comprehensive coverage, with very few data gaps

across the study region, even in urban areas (Figure 3.3). The number and grouping of

species classes is also an improvement over the 2008 U.S. Forest Type dataset (Table

3.1; Figures 3.3, 3.4), the only other known and complete tree species dataset with U.S.

coverage.

The 2012 NIDRM data were acquired directly from Jim Ellenwood at FHTET

and were available in raster (ESRI GRID) format, which facilitated the extraction

process. The dataset contains the dominant tree species by grid cell for the entire United

States at a 240-meter resolution. Dominance is determined by selecting the species with

the largest amount of basal area, or amount of square feet of wood per unit area, in each

raster cell and assigning it to that cell (J. Ellenwood 2014, personal communication;

Krist et al. 2012).

Dataset	Source	Number of classes (U.S.)	Number of classes (study area)	Spatial resolution
NIDRM 2012	**Krist et al. (2012)**	**306**	**92**	**240-m or 30-m**
U.S. Forest Type	Ruefenacht et al. (2008)	133	40	250-m

Table 3.1. Tree species dataset comparison.

The 2012 NIDRM host species dataset was derived by predicting tree species

parameters (i.e., basal area, stand density index, and quadratic mean diameter) for

300,000 USDA Forest Service Forest Inventory and Analysis (FIA) plots using a

classification and regression trees (CART) model. The model included FIA plot data

overlaid with a wide variety of geospatial predictor layers, including soils, climate,

terrain, and Landsat satellite reflectance values. Tree species parameter values were

calculated for each midpoint between FIA plots. The 2012 NIDRM is a dramatic

improvement over the 2006 NIDRM which was based solely on FIA data and used a

simple Inverse Distance Weighting (IDW) technique, instead of a regression-based

model, to generate the midpoint values (Krist et al. 2012).
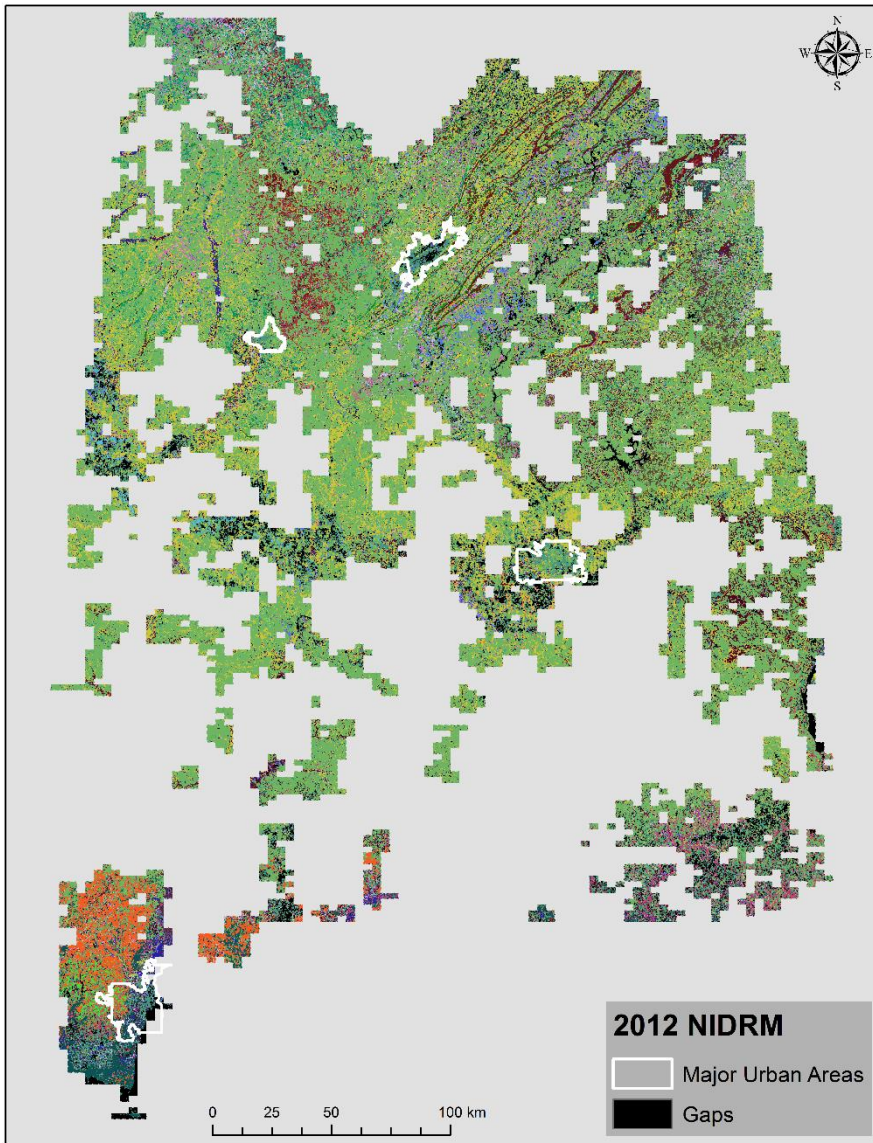
Figure 3.3. Tree species from the 2012 NIDRM, with major urban areas
(https://www.census.gov/geo/maps-data/data/tiger-line.html).
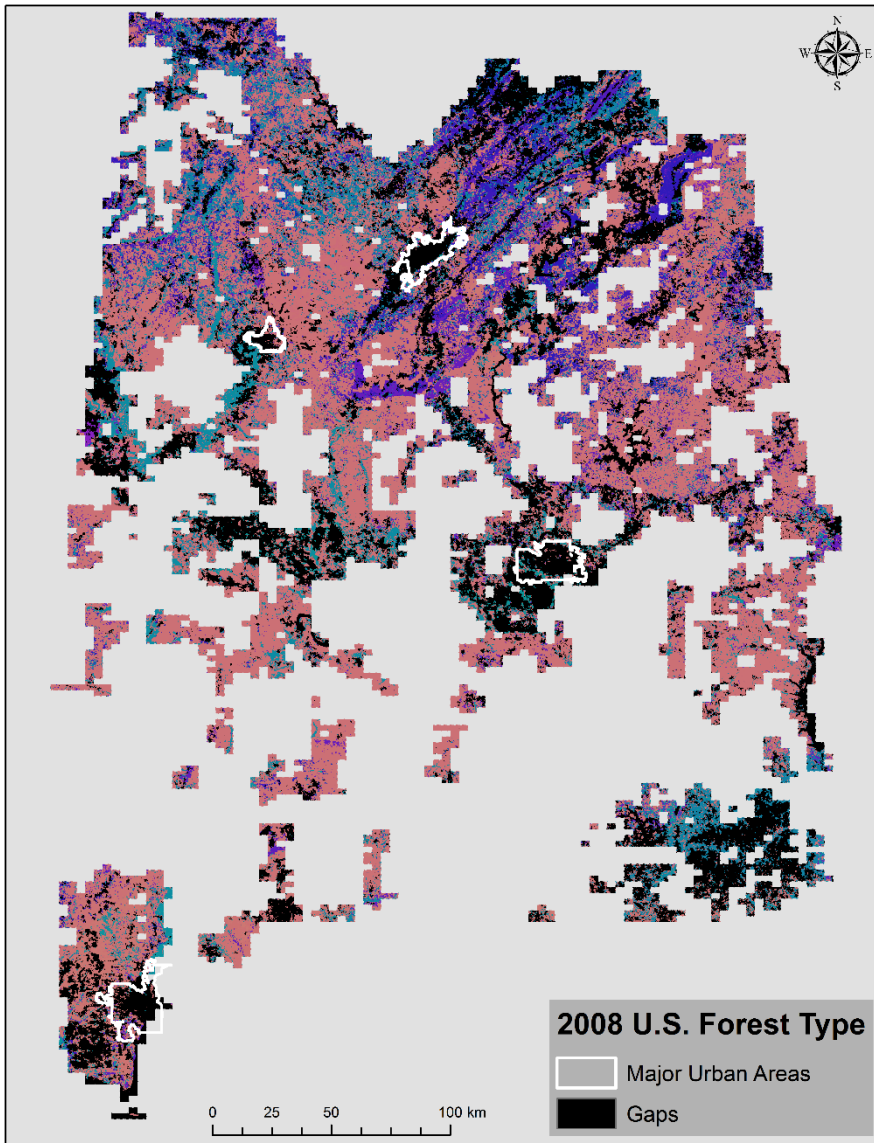
Figure 3.4. Tree species from the 2008 U.S. Forest Type dataset, with major urban areas.

3.3.2 48-hour Precipitation

The National Centers for Environmental Prediction (NCEP) Environmental Modeling Center (EMC) Stage II/Stage IV hourly precipitation dataset was chosen as a proxy for the 48-hour precipitation centered on the time of landfall. These data are currently used at NCEP to verify precipitation forecasts and as input into various environmental models (Lin and Mitchell 2005). The Stage IV data are also widely used by the National Weather Service (NWS) River Forecast Centers (Lin and Mitchell 2005), where they are referred to as Stage III (Lin 2014a). This particular dataset was chosen because of its hourly availability, which allows for greater precision in defining the 48-hour period for precipitation surrounding landfall, and for its "multi-sensor" approach, integrating radar-derived data with the rain gage data (Lin 2001, 2014a; Lin and Mitchell 2005).

The Stage II (http://data.eol.ucar.edu/codiac/dss/id=21.049) and Stage IV (http://data.eol.ucar.edu/codiac/dss/id=21.093) datasets were both obtained from the National Center for Atmospheric Research (NCAR) Earth Observing Laboratory (EOL) website at a 4-kilometer resolution. The data were converted to GeoTIFFs using the NOAA Weather & Climate Toolkit (http://www.ncdc.noaa.gov/wct/). Each file contained hourly precipitation totals in millimeters. The Stage II data were available from 1996 to 2001 and they were used for all hurricanes occurring prior to 2002 (Danny only). The Stage IV data were available beginning in 2002 and were used for the remainder of the storms (Ivan, Dennis, and Katrina).

The Stage II data were derived from NWS hourly digital precipitation (HDP) radar estimates from the WSR-88D (NEXRAD) radar product. These data then underwent a bias correction using three sets of rain gage reports from Automated Surface Observation System (ASOS) and Hydrometeorological Automated Data System (HADS) observations (Lin 2001; Lin and Mitchell 2005). The Stage IV data were created using a similar procedure, only they were derived from the Stage III data available at each of the 12 River Forecast Centers and then mosaicked together (Lin 2014a; Lin and Mitchell 2005). The total accumulated 48-hour precipitation for each storm is shown in Figure 3.5. A significant advantage of the Stage IV (and Stage III) data is the manual quality control (QC), which includes the hourly removal of gages that are malfunctioning and radar artifacts that may appear in the data (NWS 2015).
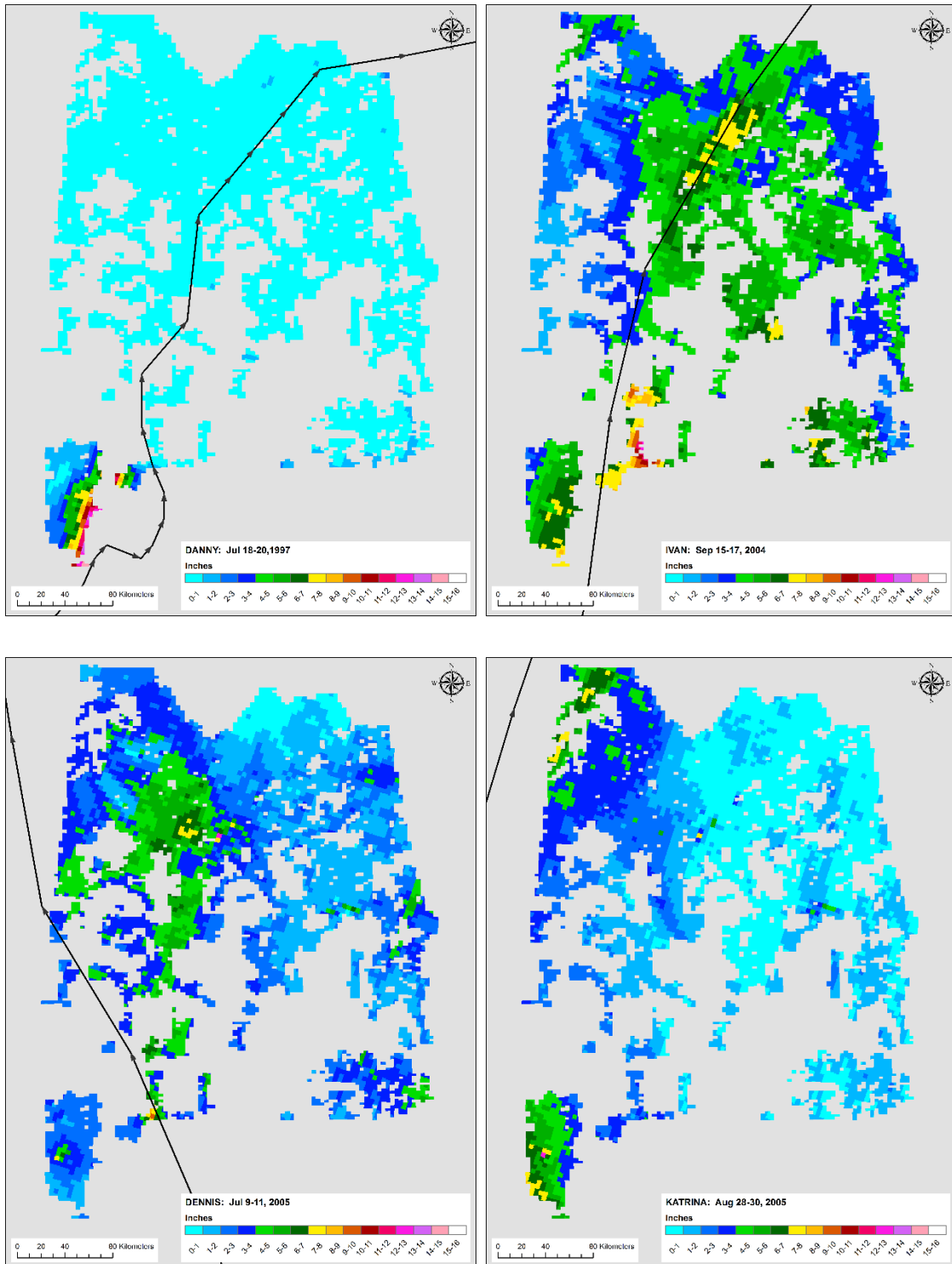
Figure 3.5. Pre-processed total 48-hour precipitation (4-km resolution) from the Stage II (Danny) and Stage IV (Ivan, Dennis, Katrina) datasets for the study region.

43

3.3 Methods

3.3.1 Objective 1: Identify the most suitable version of the HOPM for evaluating the influence of tree species and 48-hour precipitation

As described in Chapter II, the version of the HOPM used in this thesis is a random forest model, where the final result (predicted outages[4]) is the average of predictions from a large set of regression trees. The number of trees is determined by the *randomForest* function in R, which accepts a response variable (i.e., number of outages) and a set of predictor variables (Table 3.2) as its inputs. In this case, the actual number of outages from each of the four hurricanes (Danny, Dennis, Ivan, and Katrina) will be used for the response variable. For the predictor variables in Table 3.2, hurricane characteristics (i.e., wind speed and duration of strongest winds) are derived from data for each of the four storms, power system data (i.e., transformers, poles, overhead line, switches, customer density, and tree trimming) are acquired directly from the utility company, and environmental data are obtained from a variety of relevant sources. All datasets, except for tree species and 48-hour precipitation, have already been incorporated and used in previous versions of the HOPM.

---

[4] An outage can be defined as a "non-transitory activation of a protective device"; protective devices include fuses, circuit breakers, or automatic circuit reclosers. Quiring, S. M., A. B. Schumacher, and S. D. Guikema, 2014: Incorporating Hurricane Forecast Uncertainty into a Decision-Support Application for Power Outage Modeling. *B Am Meteorol Soc*, **95**, 47-58, doi: 10.1175/Bams-D-12-00012.1. The nearest protective device upstream activates when an agent causes physical damage to the power system, such as when a tree falls on a line. This is done to isolate the damage, and all customers on that portion of the system temporarily lose power. This one incident is then considered to be a single outage and can result in varying numbers of customers losing power. Liu, H., R. A. Davidson, and T. Apanasovich, 2007: Statistical forecasting of electric power restoration times in hurricanes and ice storms. *IEEE Transactions on Power Systems*, **22**, 2270-2279.

Version	Response variable	Predictor variables
All variables	Outages	Elevation, Transformers, Poles, Overhead line, Switches, Time since last landfall, Wind speed (max), Duration of strongest winds, Slope, Aspect, Topography, Land cover type, Mean annual precipitation, Standardized Precipitation Index, Soil moisture (depth 1[5]), Soil moisture (depth 2[6]), Soil type, Soil layers (number), Soil layer depth, Hydrologic group, Customer density, Tree trimming, **Tree species, 48-hour precipitation**
Public	Outages	Elevation, Time since last landfall, Wind speed (max), Duration of strongest winds, Slope, Aspect, Topography, Land cover type, Mean annual precipitation, Standardized Precipitation Index, Soil moisture (depth 1), Soil moisture (depth 2), Soil type, Soil layers (number), Soil layer depth, Hydrologic group, Customer density, Tree trimming, **Tree species, 48-hour precipitation**
Gust	Outages	Wind speed (max), Customer density, **Tree species, 48-hour precipitation**
Wind	Outages	Wind speed (max), Duration of strongest winds, Customer density, **Tree species, 48-hour precipitation**
Reduced	Outages	Wind speed (max), Duration of strongest winds, Soil moisture (depth 1), Soil moisture (depth 2), Customers (number), Tree trimming, **Tree species, 48-hour precipitation**
Mean-only (no model)	Outages	Mean of the training (leftover) sample

Table 3.2. Description of the five model versions. New predictors are in bold.

The first step in model selection is running the HOPM for all model versions with and without tree species and 48-hour precipitation. As mentioned in Chapter II, there are six versions of the HOPM, each with different predictors: All Variables, Public,

---

[5] 0-10 cm depth
[6] 40 cm and below

Gust, Gust and Duration, and Reduced (Table 3.2). The Reduced model is the one used by Nateghi et al. (2014). Each of these models will be run three separate times for all hurricanes combined (26,724 grid cells total), each time with a different combination of the new input datasets: tree species only, 48-hour precipitation only, and tree species and 48-hour precipitation combined. As part of each model run, a holdout analysis will be performed, in which data for 30% (~8,056) of the grid cells will be held out, while the remainder of the data (~18,668 grid cells) will be used to train the model. The holdout analysis will be run for 30 iterations, and the 30% holdout will be randomly generated for each iteration. As in Nateghi et al. (2014), a mean-only, or "no model", scenario will also be generated by taking the mean of the response variable (actual outages) for all of the training (leftover) grid cells. This scenario will serve as a baseline for each of the five models (i.e., it is the no-skill model).

After training is complete, the new version of the model can then be tested. The first part of the testing process will involve generating error statistics to quantify predictive accuracy for each of the held-out grid cells. This will be done in R for each of the five models, with (the three combinations) and without the new variables added. Mean absolute error (MAE), or the mean absolute difference between the actual and predicted values, will be calculated for each of the thirty holdouts using the following formula: $\frac{1}{8056}\sum_{i=1}^{8056}|actual_i - predicted_i|$. Mean square error (MSE), or the mean squared difference between the actual and predicted values, will also be calculated: $\frac{1}{8056}\sum_{i=1}^{8056}(actual_i - predicted_i)^2$. Standard deviation from the mean (SD) will then

46

be computed for MAE and MSE using the *sd* function in R. The percent change in MAE

and MSE, $\frac{\overline{error_b} - \overline{error_a}}{\overline{error_a}} \times 100$ (where a = without new variables added and b = with

new variables added), will also be calculated for each of the five models for tree species

only, 48-hour precipitation only, and both combined. A negative value for percent

change indicates a decrease in error with the new variables added, while a positive value

shows an increase in error. Because MSE gives more weight to the higher residuals

(Guikema and Quiring 2012), MSE is typically a better measure of error for normal

distributions, while MAE is generally preferred for more uniform distributions (Chai and

Draxler 2014). On the contrary, Willmott and Matsuura (2005) argue that MAE should

always be preferred over MSE and RMSE in evaluating average model performance.

This is, in part, because the larger errors tend to have a greater influence on the total

squared error than the smaller errors, resulting in increasingly larger values for MSE and

RMSE (Willmott and Matsuura 2005). For the purposes of this thesis, both MAE and

RMSE will be considered equally.

Using the *predict* function within the *randomForest* package in R, outage

predictions will then be generated for the remainder (leftover portion) of the grid cells.

MAE and MSE will be calculated using the methods described above. Maps showing

outage quantities and the absolute error, or absolute difference in actual and predicted

outages, by grid cell will be generated using ArcGIS and will aid in model selection.

Finally, the error statistics, and their associated standard deviations, will be used

along with the outage maps to determine the optimal model for examining the impact of

tree species and 48-hour precipitation on model performance. The ideal model in this case will have the lowest of the MAE and MSE values, as well as the highest reduction in error from the model without tree species and 48-hour precipitation. It will also show the least spatial bias in absolute error. Once a model is chosen, tree species and 48-hour precipitation can then be evaluated for their performance in the model (Objectives 2 and 3).

3.3.2 Objective 2: Determine if including tree species data in the HOPM will improve model accuracy

The original 2012 NIDRM displays tree species at a 240-meter resolution, so the data need to be resampled to the resolution of the utility company grid cells. The resampling will be done using Zonal Statistics in ArcGIS, and the majority tree species will be assigned to each grid cell. This identifies the most prevalent tree species in each grid cell, and each prevalent species covering at least 0.5% of the total grid cells will be used in the model. This process results in a total of 12 NIDRM classes: 10 are tree species classes, one is an 'other' class, and the final is an empty class representing gaps in the data (Figure 3.6a). For the purposes of this thesis, the 'other' class and the empty class will be combined into one class and interpreted as 'other land cover'. A comparison with the 2011 National Land Cover Dataset (http://www.mrlc.gov/nlcd2011.php) confirms that the areas covered by both of these classes contain little to no tree cover. Typically these are regions dominated by water,

cropland, pastureland, or dense urban development. In total then there are 11 classes that

will be input into the model, 10 of which represent a specific tree species (Figure 3.6b).

Combined these 11 classes account for 86% of the NIDRM raster cells, and over 98% of

the grid cells within the service area (Table 3.3).
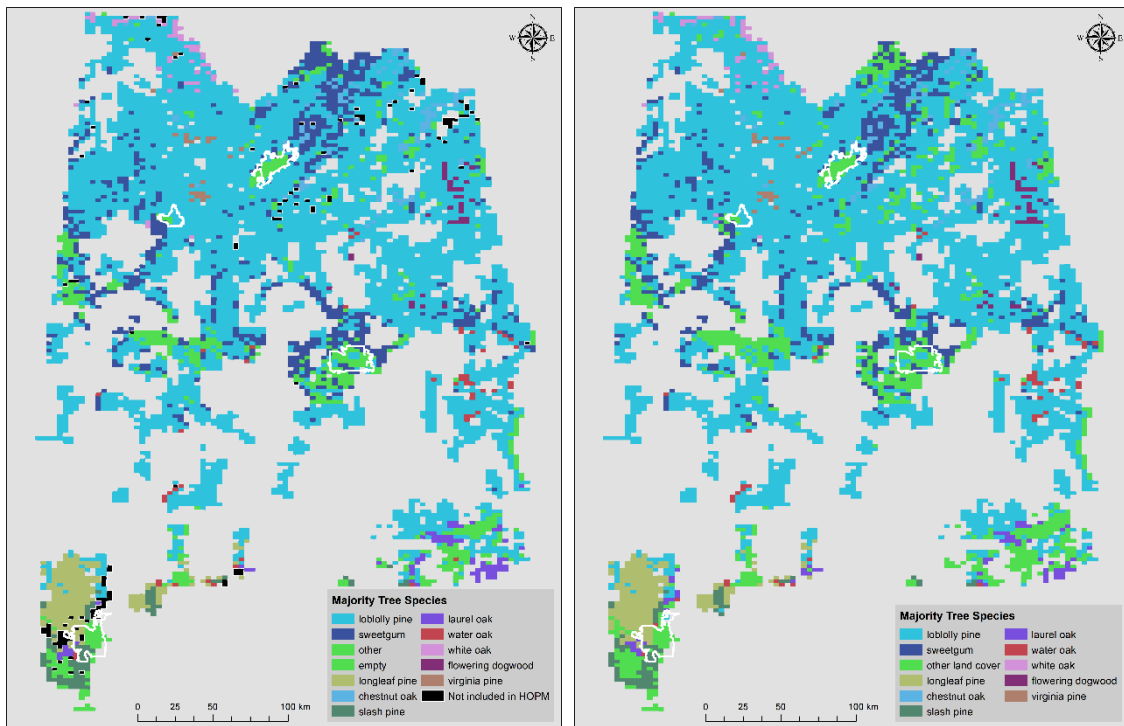
Figure 3.6. a. Tree species (majority) by grid cell, 28 classes, in order of descending prevalence. Urban areas are outlined in white. **b.** Tree species (% coverage) by grid cell, 11 classes, in order of descending prevalence.

Majority (HOPM) species	% total grid cells	Num grid	% total NIDRM cells
loblolly pine	69.20%	4623	41.83%
sweetgum	11.17%	746	14.92%
other	4.74%	317	8.84%
empty	3.64%	243	4.99%
longleaf pine	2.98%	199	1.98%
chestnut oak	1.53%	102	1.90%
slash pine	1.30%	87	1.05%
laurel oak	1.02%	68	1.32%
water oak	0.78%	52	3.62%
white oak	0.73%	49	2.43%
flowering	0.58%	39	1.87%
Virginia pine	0.45%	30	1.20%
TOTAL	98.11%	6555	85.95%
black cherry	0.31%	21	0.70%
swamp tupelo	0.28%	19	0.42%
pignut hickory	0.25%	17	1.20%
green ash	0.25%	17	1.18%
yellow-poplar	0.22%	15	1.61%
scarlet oak	0.13%	9	0.35%
sweetbay	0.10%	7	0.52%
water tupelo	0.09%	6	0.15%
shortleaf pine	0.06%	4	0.91%
eastern redcedar	0.04%	3	0.25%
red maple	0.03%	2	1.27%
post oak	0.03%	2	1.19%
baldcypress	0.01%	1	0.05%
blackgum	0.01%	1	0.42%
sourwood	0.01%	1	0.33%
live oak	0.01%	1	0.03%
TOTAL	100.00%	6681	96.54%

Table 3.3. Coverage for Majority and HOPM species classes (bolded classes are those used in the HOPM).

Prior to model integration, percent coverage will be calculated for each of the 11 tree species classes by grid cell using the Tabulate Area tool in ArcGIS. This tool calculates the areal coverage for each class in square meters. To obtain the percent coverage by grid cell, the areal coverage for each species will be divided by the sum of the areal coverage for all species combined. The most prevalent species by grid cell, according to percent coverage, is shown in Figure 3.6b.

The tree species percentages for each grid cell were aggregated with the other predictors used by the HOPM. The final product is a spreadsheet containing all of the predictors by grid cell for each of the four hurricanes. There is a separate column in the spreadsheet with the name of each storm (e.g., Danny), and for every grid cell that belongs to that storm, a value of '1' is assigned to that column. Thus, the spreadsheet will contain four sets of 6,681 grid cells, one for each hurricane.

After the model is run (Objective 1), several statistical measures will be generated to evaluate the performance of tree species in the HOPM. These measures include variable importance, partial dependence, and local variable importance. Prior to calculation of these measures, maps will be generated that display the polarized differences in the absolute error (AE) by grid cell. These maps will highlight areas where the model has improved the most with the addition of the new variables as well as areas where the model has experienced a decline in performance. The polarized differences of AE are calculated by subtracting the AE for the respective model (without the new variables added) from the same model (with the new variables added).

51

3.3.2.1 Variable importance

First, measures of variable importance are automatically generated as part of running the random forest model in R. A brief overview of how a random forest functions is useful for explaining how variable importance is calculated. The random forest function creates subsets of the training data, with each subset consisting of randomly selected grid cells (Benyamin 2012). A decision tree is then grown for each subset, and each tree predicts the number of outages for each grid cell in the holdout dataset. The average prediction for a grid cell across all trees then becomes the predicted number of outages for that grid cell (Liaw and Wiener 2002). Within each tree, a series of splits are made, with each split usually being referred to as a node. At each node, several predictors are randomly selected from the full set and assessed for their ability to split the data (Figure 3.7; Benyamin 2012; Stewart 2014; Touw et al. 2012). Ultimately, the predictor resulting in the lowest residual sum of squares (RSS) will be selected to define the split at that node (Kühnlein et al. 2014; Touw et al. 2012). This process of node splitting continues until each end node reaches purity (RSS = 0), or until a certain minimum sample size (e.g., five grid cells in a node) is reached (Guikema et al. 2010).
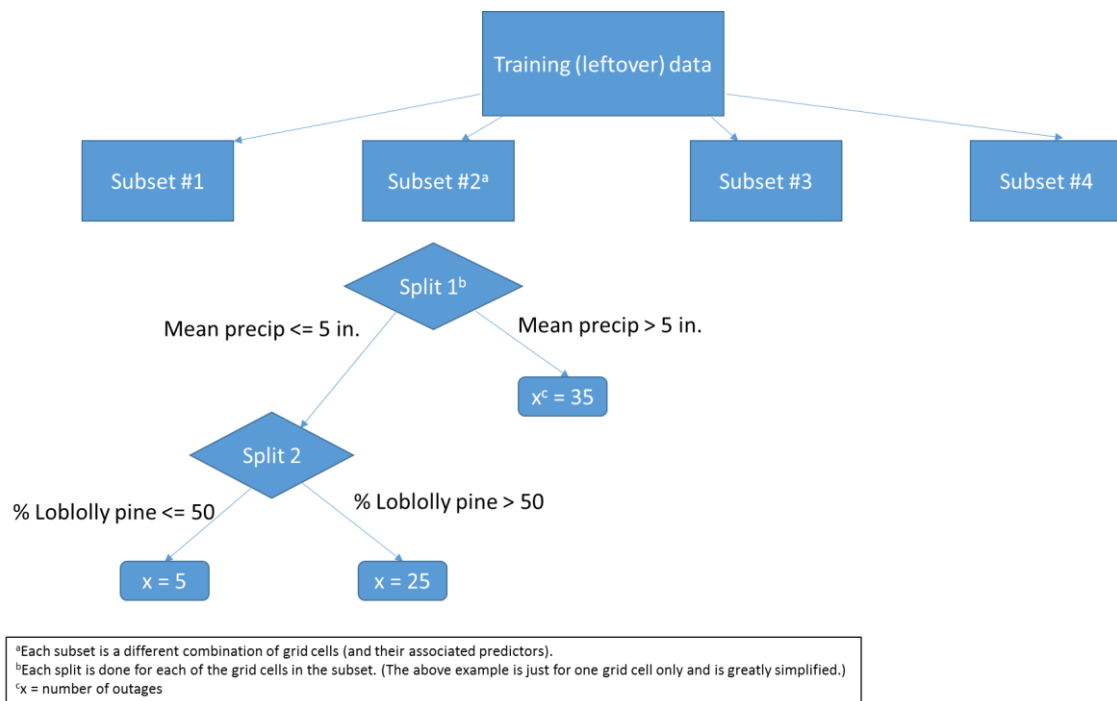
Figure 3.7. Sample random forest tree.

To calculate variable importance, the decrease in RSS is summed for every node where that predictor variable is used to split the data (Touw et al. 2012), and then normalized by the total number of trees (Horvath et al. 2007). This value is the decrease in node impurity, which is the same as the increase in node purity (IncNodePurity) reported by R. Typically a higher node purity for a predictor indicates a better relationship with the actual number of outages (Stewart 2014). After the increase in node purity is generated, the importance of each variable will be normalized so that the most important variable has a value of 100 (Quiring et al. 2011). The variables will then be ranked in order of decreasing importance. Due to random sampling, outage predictions and rankings of variable importance vary slightly from one model run to the next. The

variables that are ranked highest are the ones that have the greatest node purity across all runs.

3.3.2.2 Partial dependence

Next, partial dependence plots will be generated for tree species and 48-hour precipitation using the method from Nateghi et al. (2014). Partial dependence is a measure of the influence of a particular predictor variable on the response variable (in this case, outages) while the effects of the other predictors are held constant (Nateghi et al. 2014). Partial dependence plots are important because they indicate which variables may have the most predictive power and they can also depict thresholds where a variable may have even greater influence. For example, wind gust has a strong (non-horizontal) relationship with outages right around 25 m/s, after which the curve flattens and becomes more horizontal, indicating less of a relationship (Figure 3.8). Thus, it could be said that an increase in wind gust speed from 20 to 30 m/s has a much greater impact on the number of outages than an increase from 10 to 20 m/s.
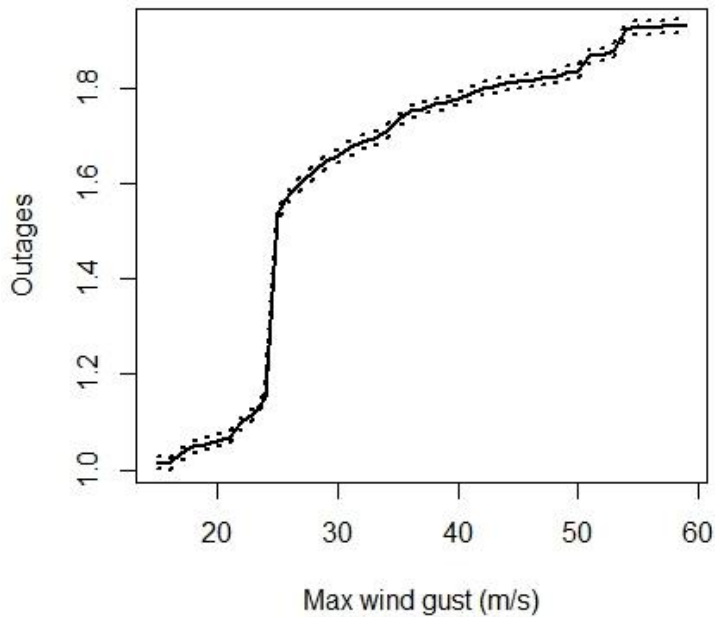
Figure 3.8. Partial dependence plot for 3-second wind gust speed.

3.3.2.3 Local variable importance

To assist in the interpretation of partial dependence, maps showing local variable importance will be generated in ArcGIS and used in conjunction with the wind resistance and survival rankings from several of the studies mentioned in Chapter II. For example, Duryea et al. (2007) examined the response of over eighty tree species in Florida to winds from nine hurricanes including Dennis, Ivan, and Katrina. The local variable importance scores may help explain why certain species are more important.

According to Touw et al. (2012), local variable importance in a random forest can describe the link between variables and samples. Each local variable importance score will be calculated from the out-of-bag (OOB) data from all trees in the random

forest. The OOB data comes from the sample and is the portion (usually 1/3) of the validation data that is not used to train each tree. To calculate local variable importance, a portion of the OOB data is first permuted, or altered; then, the MSE for the permuted OOB data and the original OOB data are calculated. The permuted MSE is then subtracted from the original MSE to get a local importance score for each variable in each grid cell. The score is a measure of the predictive accuracy of the sample, and in this case, it is a measure of the impact on the correct prediction of outages. A negative score indicates a negative impact on the prediction of outages, or in other words, the variable did not improve the prediction. A 0 score is neutral, and a positive score indicates that the variable had a positive impact on, or improved, the prediction of outages. A higher score indicates a more positive impact on outage prediction (Grömping 2009; Touw et al. 2012).

To calculate local variable importance, thirty iterations of the HOPM will be run for the model version selected in Objective 1. Local variable importance (LVI) scores will then be generated for each variable by grid cell. As in Tonn et al. (2014), LVI scores will be calculated for each grid cell by taking the mean importance of all four storms for the selected variable and then dividing that value by the sum of the mean importance scores for all of the variables. The result is a percentage importance value that indicates the importance, or contribution, of the variable in relation to the other variables in the model. LVI scores will only be calculated for tree species having a significant relationship with outages, as indicated by the partial dependence plots and variable

importance scores. An LVI map will also be generated to show the variable with the highest importance for each grid cell.

3.3.3 Objective 3: Determine if including precipitation data in the HOPM will improve model accuracy

Using the NCEP data, hourly precipitation totals will be acquired for the 24 hours prior to and following the time of landfall for each hurricane. For storms with multiple landfalls, the point of landfall closest to the service area will be used (Table 3.4); for Danny, Dennis, and Katrina, this also corresponds to the final point of landfall[7]. A similar method is used by Zhu and Quiring (2013) and Zhu et al. (2013) to estimate daily tropical cyclone precipitation (TCP) in Texas. In both of those studies, hurricane positions (based on center of circulation) are obtained for every six-hour period, and buffers based on the radius of the outer closed isobar (ROCI) are generated for each position. All of the buffers for a given day are then merged into one buffer, and precipitation measurements from any weather station located within that buffer are attributed to the storm total precipitation for that day (Zhu and Quiring 2013).

---

[7] After passing over State A, Ivan moved northeastward toward the Delmarva Peninsula and out into the Atlantic. Several days later, the storm began moving south and then southwestward, back over Florida and into the Gulf of Mexico once again. A second landfall was made in southwestern Louisiana on September 24, 2004. Stewart, S. R., 2004: Tropical Cyclone Report: Hurricane Ivan, 2-24 September 2004. National Hurricane Center, 44 pp, [Available online at http://www.nhc.noaa.gov/data/tcr/AL092004_Ivan.pdf.]

Hurricane	Landfall date	Landfall time (UTC)	Landfall coordinates	Wind speed at landfall	48-hour precip
Danny	7/19/1997	18:00	30.4° N, 87.9° W	33 m/s (75 mph)	18:00 7/18-7/20
Ivan	9/16/2004	6:50	30.2° N, 87.9° W	54 m/s (121 mph)	7:00 9/15-9/17
Dennis	7/10/2005	19:30	30.4° N, 87.1° W	54 m/s (121 mph)	20:00 7/09-7/11
Katrina	8/29/2005	14:45	30.2° N, 89.6° W	54 m/s (121 mph)	15:00 8/28-8/30

Table 3.4. Landfall data for the four hurricanes (Beven 2005; Knabb et al. 2005; Pasch 1997; Stewart 2004). The last column is the 48-hour timespan used for data extraction.

To apply this model in real-time, it will be necessary to forecast precipitation. This is difficult to do many days prior to landfall. Therefore, the 48-hour window centered on landfall may not be an appropriate timescale to employ for real-time applications. It is used in this study because it typically corresponds with the highest precipitation amounts for most storms.

One of the downsides to using the Stage II data is that it contains negative (-1) values. This may be due to radar malfunction (Ware 2005) and/or missing gage data (Lin 2014b). For the purposes of this analysis, these negative values are assumed to indicate missing data (NoData), as no information was found on these values in the metadata.

Before converting to inches and calculating the 48-hour sum for each storm, each pixel containing a -1 value was set to zero. Since Danny is the only storm that contains -1 values, the impact should be minimal (Figure 3.9). In addition, a maximum of only 6 out of 48 hours have -1 values for Danny, and most of these are confined to a very small area in the northwest corner of the service area (Figure 3.9).
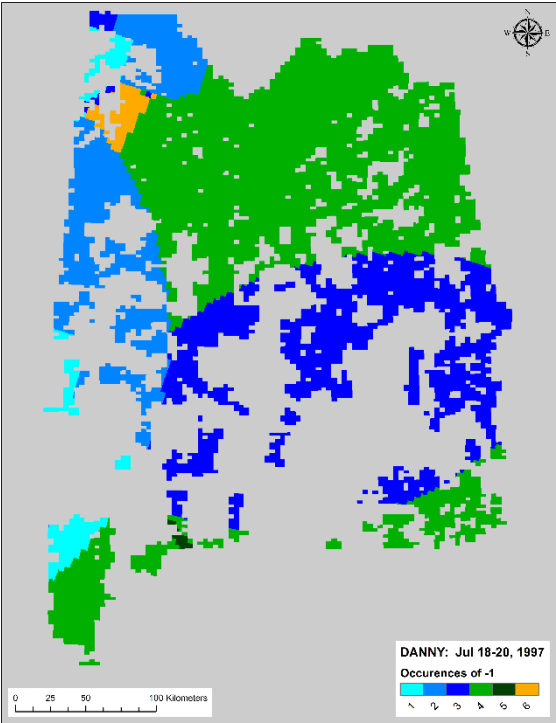
Figure 3.9. Locations of missing (-1) values for Danny.

The precipitation data are at a 4-kilometer resolution, so they were resampled to the resolution of the utility company grid cells in ArcGIS using Zonal Statistics. The

59

mean, minimum, and maximum precipitation value was assigned to each grid cell
(Figure 3.10).
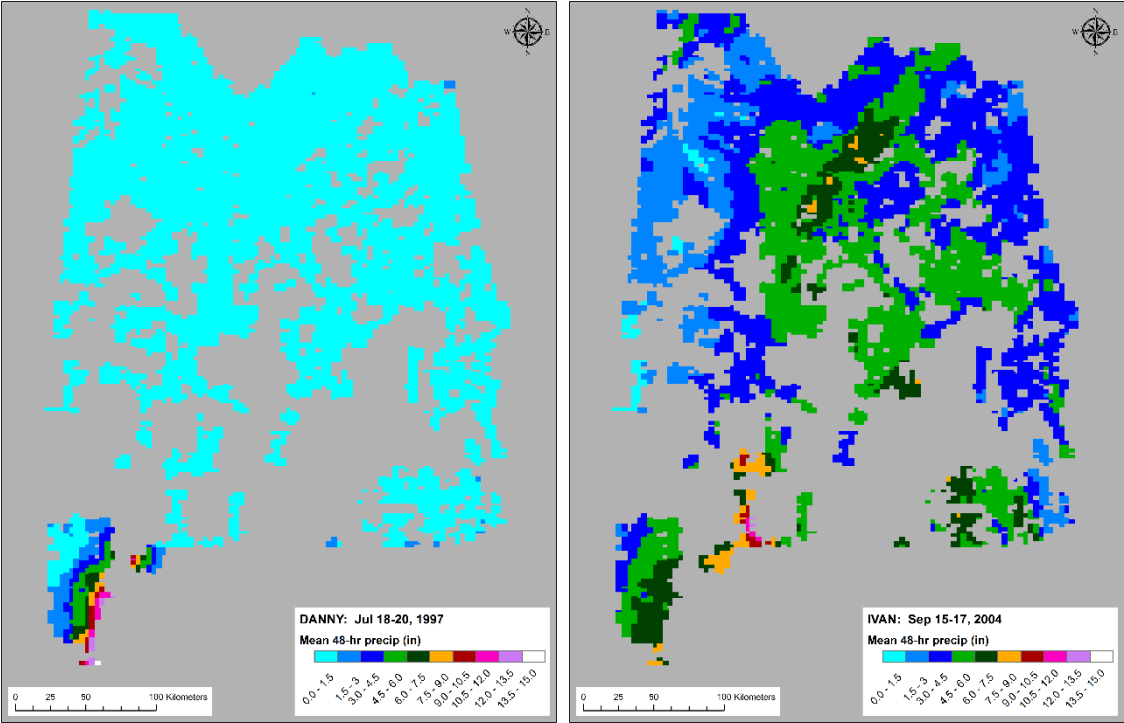
Figure 3.10. Post-processed 48-hour precipitation (mean) by grid cell for each storm
(Danny, Ivan, Dennis, and Katrina).
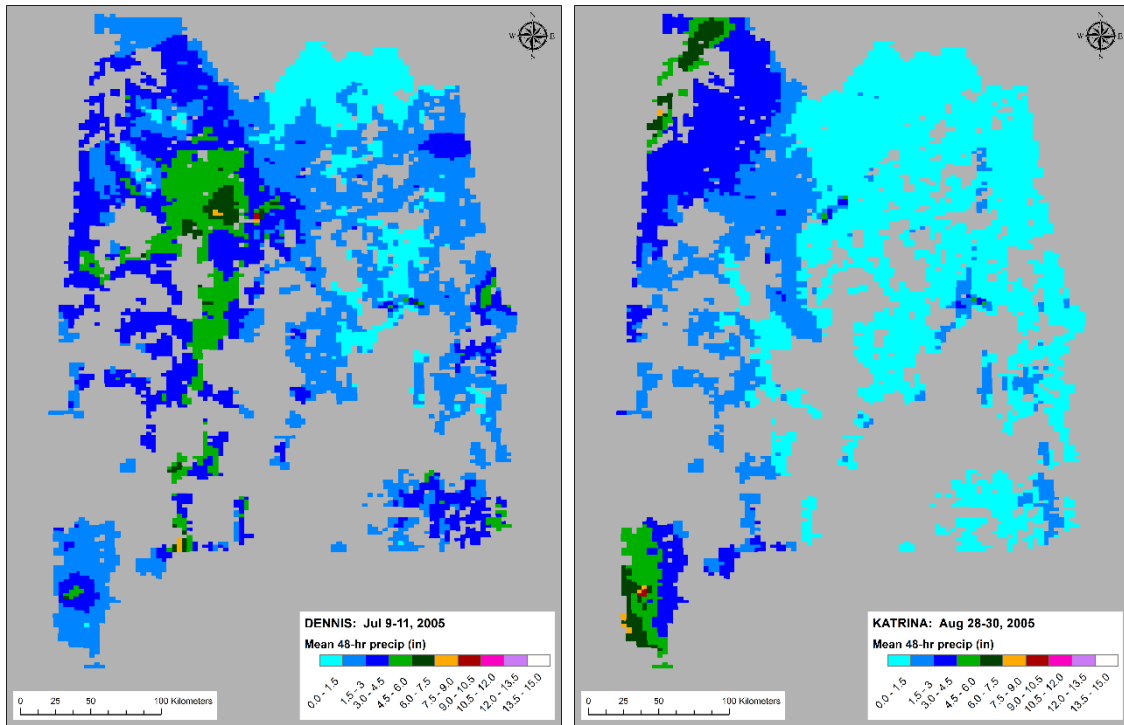
Figure 3.10 Continued.

After the model is run (Objective 1), similar methods will be used to evaluate the addition of 48-hour precipitation on model performance. These include calculation of the polarized differences in AE as well as measures of variable importance, partial dependence, and local variable importance, which will be calculated using the same methodology described in Objective 2.

The issues with the data from Georges led us to assess differences between the Stage II and Stage IV datasets. For hurricanes where both the Stage II and the Stage IV data are available (Ivan, and Katrina), 48-hour precipitation totals from the Stage II data are subtracted from the Stage IV totals, and maps displaying the differences are

61

generated in ArcGIS. Observed differences are as high as 9.3 inches for Ivan and 7.4 inches for Katrina (Figure 3.11). This rather drastic result raises some concern in using the Stage II data; however, it is the only option for Danny.
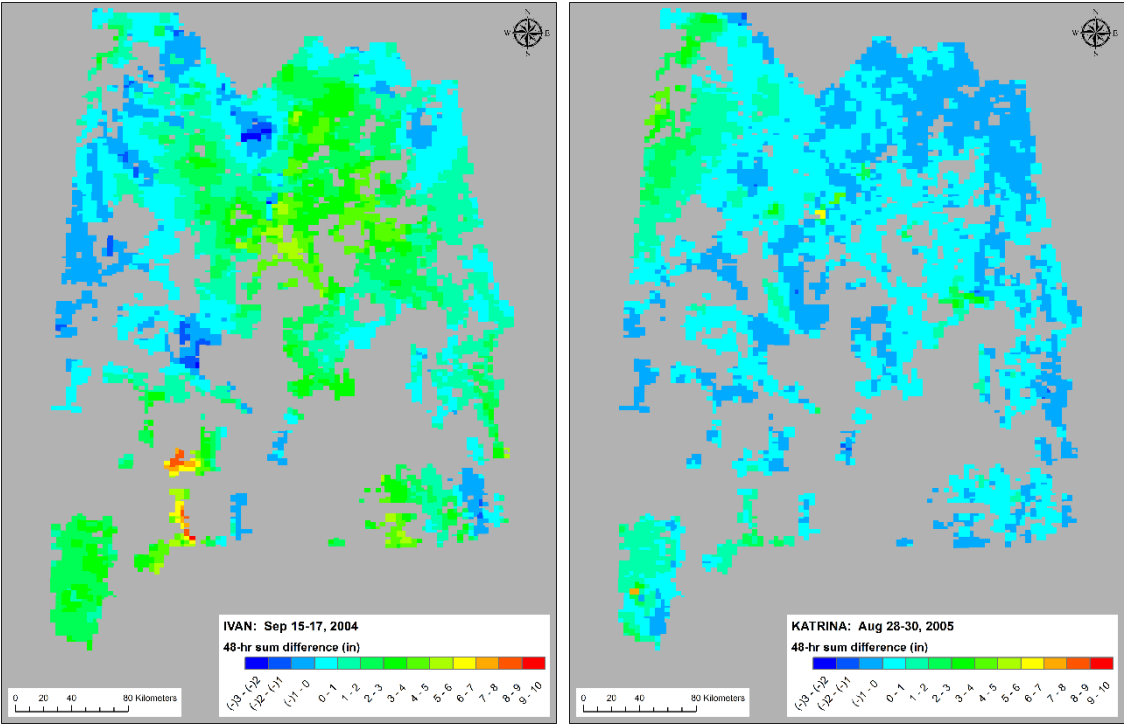
Figure 3.11. Stage IV-Stage II difference maps for Ivan and Katrina (inches).

CHAPTER IV

MODEL EVALUATION AND SELECTION

4.1 Model error

The main goal of Objective 1 is to determine the most appropriate model for analyzing the effect of tree species and 48-hour precipitation. This is done by calculating error statistics for each model using the three combinations of variables as input (tree species only, 48-hour precipitation only, and both combined). This process is done in multiple stages, each one filtering out less accurate or less appropriate models.

During the first stage, MAE and MSE are calculated for each model. The Gust model has the highest MAE and MSE, while the Reduced and All Variables models have the lowest errors (Tables 4.1 and 4.2). Therefore, these two models are the most suitable for assessing the importance of new variables and they will be used in the subsequent analyses. This finding agrees with Nateghi et al. (2014). Including a larger number of predictors and/or more power system predictors typically increases model accuracy.

Variable	Reduced	All Variables	Public	Gust+ Duration	Gust	Mean only	Mean MAE	Mean SD
Tree only	0.747	0.755	0.773	0.830	0.849	1.605	0.791	0.014
Tree and precip	0.754	0.762	0.780	0.812	0.824	1.612	0.787	0.017
Precip only	0.757	0.762	0.781	0.827	0.906	1.606	0.806	0.021
Model mean	0.753	0.760	0.778	0.823	0.860	1.608	0.795	0.017

Table 4.1. Mean Absolute Error (MAE) for the 5 models that were evaluated: Reduced, All Variables, Public, Gust + Duration, and Gust. Mean only (no model) is shown for comparison only and it is not included in calculation of mean MAE and mean SD.

Variable	All Variables	Reduced	Public	Gust+ Duration	Gust	Mean only	Mean MSE	Mean SD
Tree only	4.084	4.114	4.293	5.002	5.307	14.56	4.560	0.491
Tree and precip	4.447	4.414	4.646	5.281	5.474	15.29	4.852	0.602
Precip only	4.391	4.491	4.646	5.569	6.436	15.09	5.107	0.684
Model mean	4.307	4.340	4.528	5.284	5.739	14.98	4.840	0.592

Table 4.2. Mean Square Error (MSE) for the 5 models that were evaluated: Reduced, All Variables, Public, Gust + Duration, and Gust. Mean only (no model) is shown for comparison only and it is not included in calculation of mean MSE and mean SD.

The next stage involves looking at the change in model error with the new variables added. The percent change in MAE and MSE is calculated for the Reduced and All Variables models for each of the three combinations of new variables (Table 4.3). The Reduced model shows the greatest improvement with the addition of the new variables. The addition of tree species or both tree species and 48-hour precipitation variables has the greatest influence in terms of reducing model error. Adding the tree species and precipitation variables to the All Variables model has little influence on model performance. This is likely because the All Variables model already has a large

number of predictors, some of which may serve as proxies for the new variables being added. The Reduced model has significantly fewer predictors and therefore benefits the most from the addition of new variables.

Model	MAE (new)	MAE (old)	MAE (%Δ)	MSE (new)	MSE (old)	MSE (%Δ)
Reduced, combined	0.754	0.762	-1.050	4.414	4.764	-7.347
Reduced, tree only	0.747	0.754	-0.928	4.114	4.406	-6.627
Reduced, precip only	0.757	0.760	-0.395	4.491	4.703	-4.508
Mean	**0.753**	**0.759**	**-0.791**	**4.340**	**4.624**	**-6.161**
All Variables, combined	0.762	0.763	-0.131	4.447	4.452	-0.112
All Variables, tree only	0.755	0.757	-0.264	4.084	4.089	-0.122
All Variables, precip only	0.762	0.761	0.131	4.391	4.389	0.046
Mean	**0.760**	**0.760**	**-0.088**	**4.307**	**4.310**	**-0.063**

Table 4.3. Change in error (%, new MAE-old MAE). A negative (-) % change in error indicates a decrease in error with the new variables added.

Error graphs can also give a little more perspective to changes in error among the models (Figure 4.1). Both the All Variables and the Public models show little change in error with the addition of the new tree species and 48-hour precipitation variables. These two models already contain many variables and so the addition of new variables has little influence. The addition of the new variables in the other three models (Gust, Gust+Duration, and Reduced) decreases error, with the most significant reduction being in the Gust only model. It is also apparent when looking at the MAE graphs that the Reduced model has a comparable amount of error to that of the All Variables or Public models. The MSE graphs also show that, among those three models, the Reduced model

65

experiences the largest decrease in error with the addition of the new variables. These observations confirm what was stated previously and provide further evidence for using the Reduced model for this analysis.
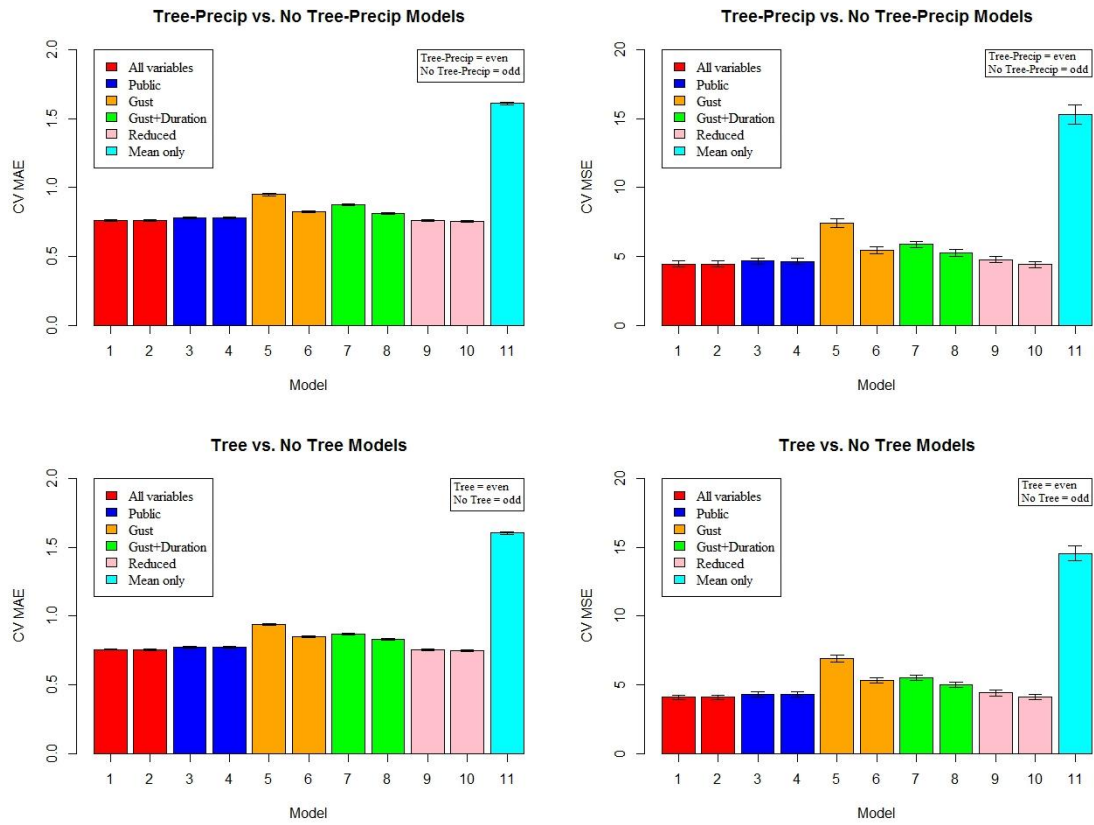
Figure 4.1. MAE/MSE error graphs for all model versions.
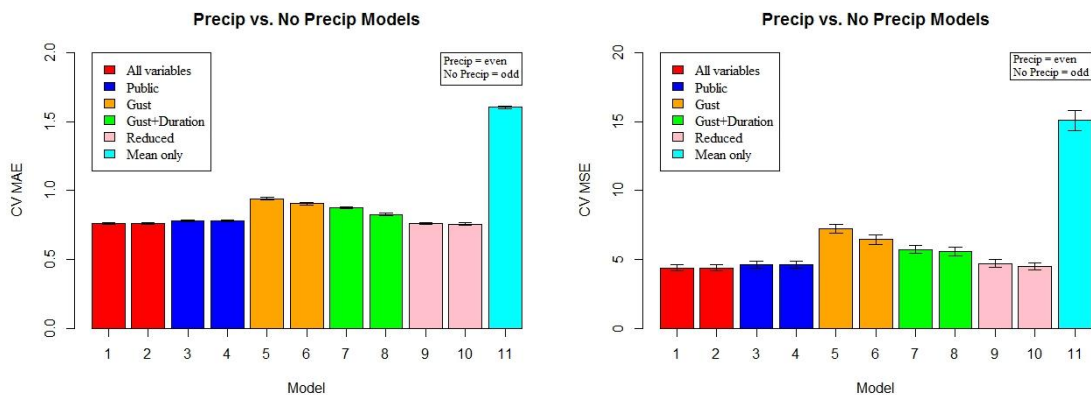
Figure 4.1 Continued.

4.2 Outage prediction performance

One way to quantify outage prediction performance is to show the spatial pattern of differences in the absolute error (AE). AE is generally preferred over squared error (SE) because it is less influenced by outliers (Chai and Draxler 2014; Guikema and Quiring 2012). Displaying the error in this way can aid in identifying spatial patterns where model performance may be particularly high or low. This was done previously by Quiring et al. (2011) to evaluate the influence of new variables on a prior version of the HOPM. Here the polarized differences of AE are calculated by grid cell for the two model candidates, the Reduced and the All Variables models (Figure 4.2). The polarized differences in AE are generated by subtracting the AE for the respective model (without the new variables added) from the same model (with the new variables added). Differences that are greater than 1.5 standard deviations below the mean are shown in green and indicate that the model improved with the addition of the new variables.

Differences greater than 1.5 standard deviations above the mean are shown in brown and

indicate that the model performance declined with the addition of the new variables

(Figure 4.2).
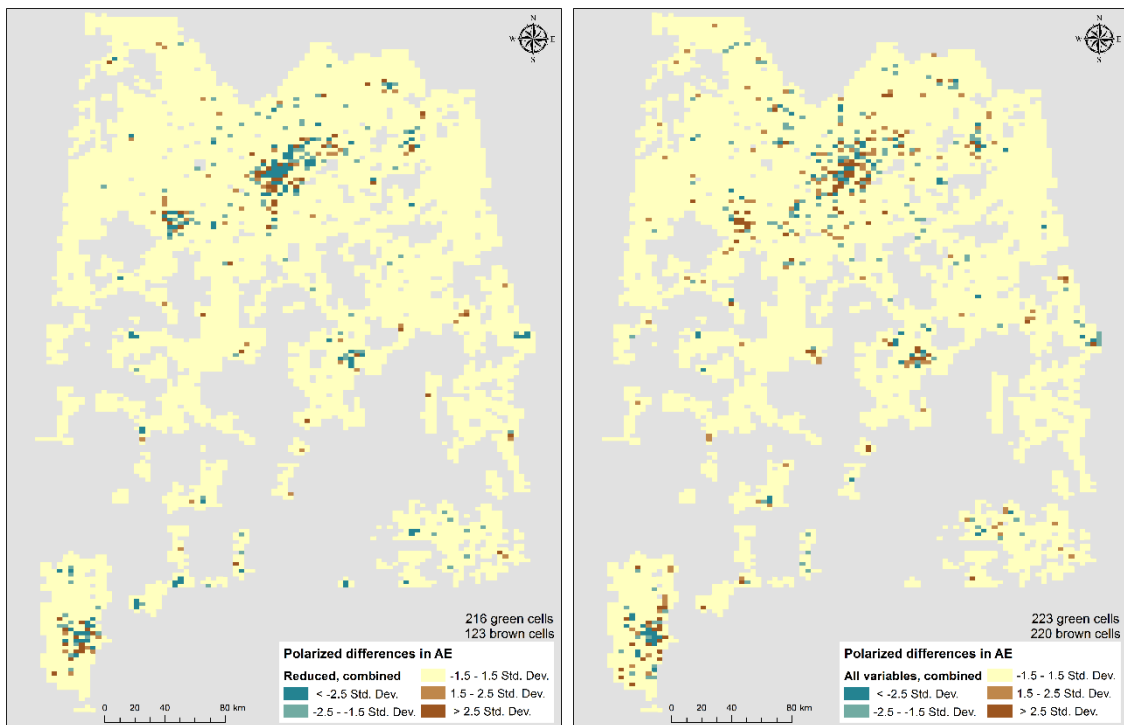
Figure 4.2. Polarized differences in AE (Model 1 − Model 2) for the Reduced and All
Variables models. Model 1 = without the new variables added; Model 2 = with the new
variables added. Differences > 1.5 SD below the mean (green) indicate significant model
improvement, while differences > 1.5 SD above the mean (brown) indicate a significant
decline in performance.

The polarized differences of AE show that the Reduced model, with tree species

and 48-hour precipitation variables added, performs better (by 93 grid cells) than the

original Reduced model (without the added variables), particularly in urban areas. The

All Variables model was more accurate in 223 grid cells, and less accurate in 220 grid cells, with the addition of the new variables. Therefore, overall there was not a consistent pattern of model improvement for the All Variables model.

Hot spot maps can provide an even clearer depiction of model performance by identifying statistically significant clusters of positive (red) and negative (blue) AE differences using the Getis-Ord Gi statistic (Figure 4.3). These maps highlight the spatial patterns in Figure 4.2, with the Reduced model showing many more areas of improvement (blue) spatially than the All Variables model, which actually shows one large area of decline in accuracy (red). Again, the areas of greatest change in accuracy tend to be centered on the major cities.

Another way to assess model performance is to divide the grid cells into bins based on the number of outages and then calculate the MAE for each bin. This method was used previously by Han et al. (2009a) to compare the accuracy of two prior versions of the HOPM. Four outage groups were determined using the Natural Breaks (Jenks) classification scheme in ArcGIS. The change in MAE for each of these groups again shows that the Reduced model experiences a larger decrease in error with the addition of the new variables than the All Variables model (Table 4.4). In fact, the All Variables model actually experiences a slight increase in average MAE with the addition of the new variables. The MAE for both models is consistently lower with fewer outages, which is to be expected due to the larger sample size. The Reduced model experiences its greatest improvement in the grid cells with the highest number of outages (> 21),

while the All Variables model shows more improvement, though very minimal, in grid

cells with lower numbers of outages.
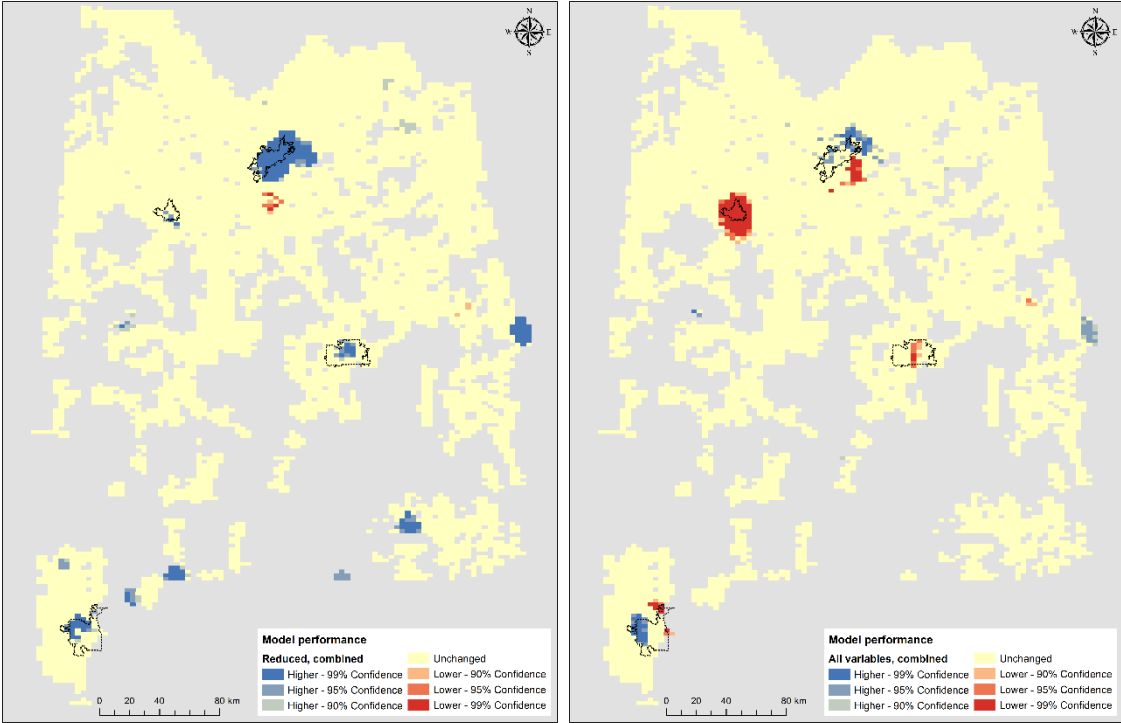
Figure 4.3. Statistically significant clusters of positive (red) and negative (blue) polarized AE differences, calculated using the Getis-Ord Gi statistic. More negative AE differences (blue) indicate higher model performance, while more positive AE differences (red) indicate lower model performance.

		Reduced, combined			All Variables, combined		
Actual outages	Num cells	MAE (new)	MAE (old)	MAE (Δ)	MAE (new)	MAE (old)	MAE (Δ)
0 ~ 2	5805	0.261	0.272	-0.011	0.258	0.259	-0.001
2 ~ 8	718	1.234	1.321	-0.087	1.202	1.210	-0.008
8 ~ 21	134	3.162	3.291	-0.129	3.139	3.119	0.020
> 21	24	8.187	8.809	-0.622	7.739	7.686	0.053
Mean	-	3.211	3.423	-0.212	3.085	3.069	0.016

Table 4.4. Change in MAE ($MAE_{new} - MAE_{old}$) by outage group (actual outages are for all four storms combined).

4.3 Model selection

In summary, of the five models, the Reduced model is the one that shows the greatest improvement with the addition of the new variables. As indicated by Figures 4.2 and 4.3 above, there is a spatial pattern to where the Reduced model performs better with the addition of the new variables. This may help identify predictors in the All Variables model that are highly correlated with the new variables. In addition to showing improvement in model accuracy, the Reduced model is also the version of the HOPM used by Nateghi et al. (2014), and therefore it will serve as a good basis for comparison in this thesis.

CHAPTER V

EVALUATION OF TREE SPECIES

5.1 Error and outage prediction performance

The addition of tree species results in the lowest MAE and MSE for the Reduced model (Tables 4.1 and 4.2). This suggests that tree species variables are useful for improving model performance. Error decreases significantly when tree species is added to the model (Table 4.3). Tree species and precipitation combined result in the most error reduction, followed by tree species alone, and lastly 48-hour precipitation.

Another method for assessing the influence of the new predictors is to calculate the polarized difference in AE by service area grid cell, as was done for the Reduced and All Variables models in Chapter IV. The polarized difference is computed by simply subtracting the AE of the original model (without the added variables) from the AE of the new model (with the added variables). Thus, a negative (-) polarized difference value would indicate an improvement in AE for that grid cell with the addition of the new variables, while a positive difference value would indicate a decline in AE. The spatial pattern for the polarized difference in AE is fairly similar for both tree species and 48-hour precipitation (Figure 5.1), with greater differences in error generally occurring near urban areas.
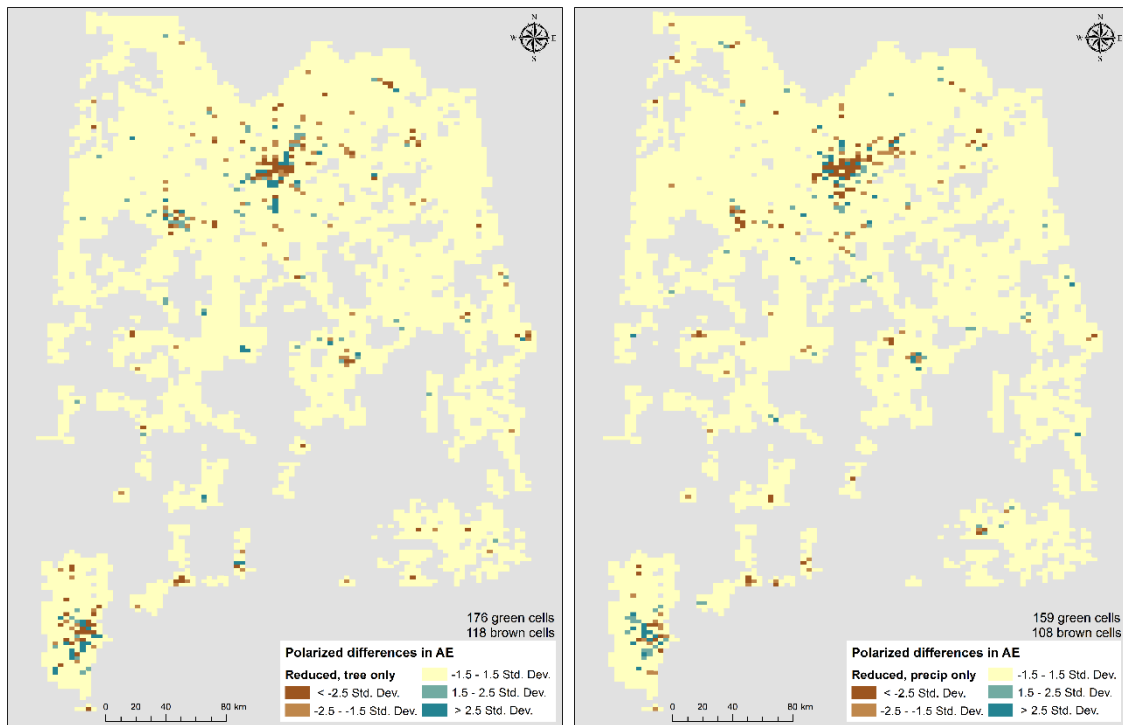
Figure 5.1. Polarized differences in AE (Model 1 − Model 2) for the Reduced model with tree species and precipitation added. Model 1 = without the new variables added; Model 2 = with the new variables added. Differences > 1.5 SD below the mean (green) indicate significant model improvement, while differences > 1.5 SD above the mean (brown) indicate a significant decline in performance.

To facilitate comparison between the two sets of variables, the polarized difference data are summarized in Table 5.1. The number of grid cells showing an improvement (higher) or a decline (lower) in model performance are grouped by three ranges of SD for each of the three variable sets. The percent differences between the number of improving and declining grid cells for each SD range are also shown. A positive percent difference value indicates that the number of improving grid cells are greater than the number of declining grid cells, while a negative value signifies a greater

number of declining grid cells. The results indicate that the addition of tree species and 48-hour precipitation both result in similar improvements to model performance, with 48-hour precipitation resulting in the greatest overall improvement in model performance, particularly at lower SD's. This trend then suggests that tree species may play an important role in grid cells showing average-to-high levels of improvement, while 48-hour precipitation plays a fairly important role across all grid cells.

Model	+/-0.5 to +/-1.5 SD	+/-1.5 SD to +/-2.5 SD	> +/-2.5 SD	Average
Tree only	7	29	37	24
Precip only	13	25	40	26
Combined	-1	42	44	28
Average	**6**	**33**	**41**	26

Table 5.1. Percent differences in model performance, based on polarized differences in AE. Values indicate the percent difference between improving (-SD) and declining (+SD) grid cells for each range of SD's.

"Hot spot" maps are also generated in ArcGIS using the Getis-Ord Gi statistic, which uses spatial autocorrelation to identify statistically significant high and low clusters in the data (Figure 5.2). In this case, areas in blue are where the model performs better (lower difference in AE), while the areas in red are where the model performs worse (higher difference in AE). The addition of tree species variables seems to result in multiple areas of model improvement scattered throughout the service area, with the most improvement occurring in urban areas. There are also a few areas where the new model performs worse, most notably in the north-central portion of the service area on

the outskirts of a large city. Overall, tree species seems to result in model improvement

over a broad spatial area, covering multiple regions of the service area, while 48-hour

precipitation focuses almost strictly on the major cities.
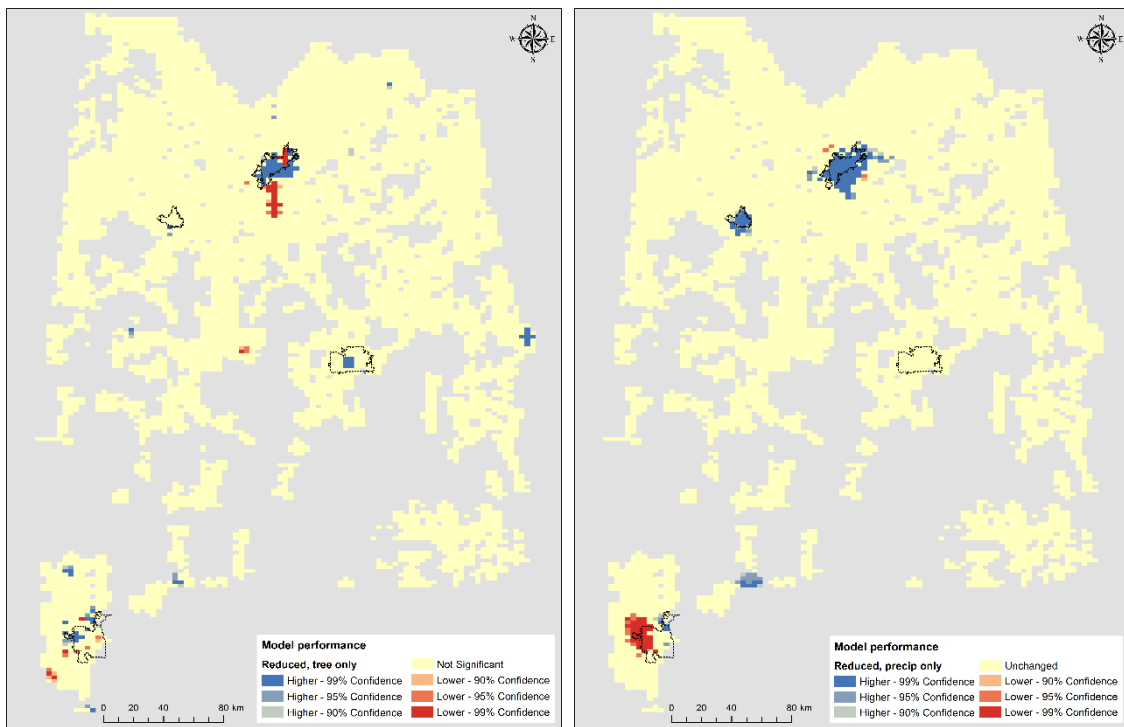
Figure 5.2. Statistically significant clusters of positive (red) and negative (blue) polarized AE differences (with tree species and with precipitation added), calculated using the Getis-Ord Gi statistic. More negative AE differences (blue) indicate higher model performance, while more positive AE differences (red) indicate lower model performance. The black outline denotes major urban areas.

Lastly, MAE for tree species is calculated for four outage groups (Table 5.2).

These are the same groupings used in Chapter IV and are determined using the same

methodology. A negative change in MAE signifies an improvement in model

performance, while a positive change indicates a decline in model performance. The change in MAE with the addition of both variable sets becomes more negative with an increase in the number of outages (and a resulting smaller sample size of grid cells), which suggests that tree species and precipitation are more accurate predictors for grid cells having larger numbers of outages. For the majority of the grid cells, tree species appears to perform slightly better; although, it does seem that 48-hour precipitation may be a better predictor for grid cells with higher numbers of outages.

		Tree species			48-hour precipitation		
Actual outages	Num cells	MAE (new)	MAE (old)	MAE (Δ)	MAE (new)	MAE (old)	MAE (Δ)
0 ~ 2	5805	0.264	0.274	-0.010	0.265	0.272	-0.007
2 ~ 8	718	1.270	1.337	-0.067	1.279	1.33	-0.051
8 ~ 21	134	3.245	3.333	-0.088	3.331	3.429	-0.098
> 21	24	7.648	7.782	-0.134	7.917	8.519	-0.602
Mean	-	3.107	3.182	-0.075	3.198	3.388	-0.190

Table 5.2. Change in MAE by outage group (actual outages are for all four storms combined).

In summary, the addition of tree species results in general model improvement, while improvement from 48-hour precipitation is almost strictly limited to major urban areas (where in some cases it may outperform tree species). Overall, tree species may have a slightly greater effect on improving model performance than 48-hour precipitation, but combined is definitely where they hold the greatest predictive power.

5.2 Variable importance

To assess the influence of each individual covariate within the tree species set, variable importance (the increase in node purity) is calculated using the random forest function in R. As in Quiring et al. (2011), variable importance values are normalized, with the most important covariate having a value of 100.

The importance graph in Figure 5.3 shows the variable importance, ranked by increase in node purity, for all covariates in the Reduced model. The number of customers and the tree trimming factor rank the highest, followed by wind duration and speed, soil moisture, and 48-hour precipitation. The tree species covariates rank lowest on this scale of variable importance. However, as is evidenced in the previous section, this alone does not mean they are insignificant predictors, and combined they can result in considerable model improvement.
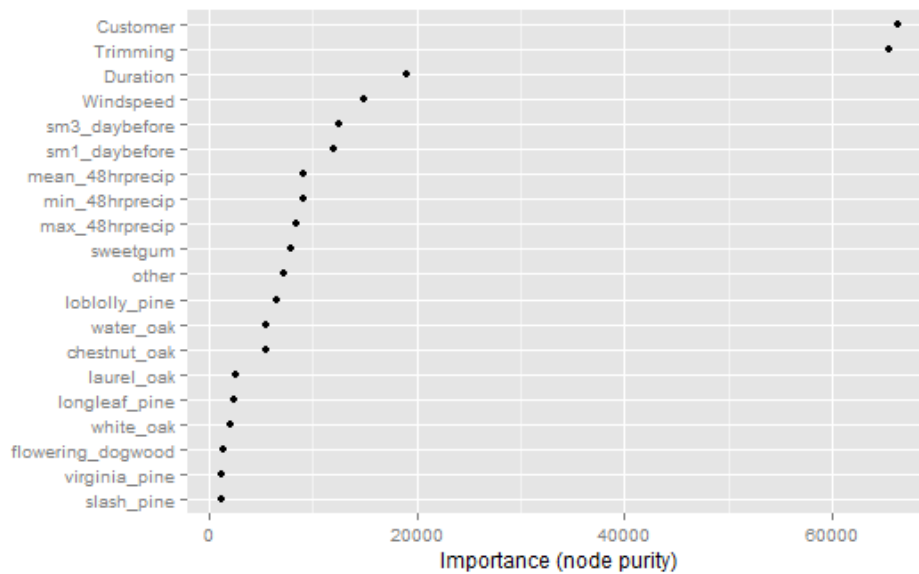
Figure 5.3. Variable importance graph for the Reduced model.

Within the tree species covariates, several clusters can be identified by grouping covariates with similar normalized importance scores (Table 5.3). This results in four groups total for the tree species covariates. The first group, showing the most importance, includes sweetgum, other, and loblolly pine. The three 48-hour precipitation covariates are also assigned to this same group. The second group consists of water oak and chestnut oak, followed by the third group containing laurel oak, longleaf pine, and white oak; and the fourth group with flowering dogwood, Virginia pine, and slash pine. Each of the tree species covariates will be discussed in more detail in the next section.

Predictor	Normalized importance	Normalized importance (rounded)
Mean 48-hr precip	13.78	14
Min 48-hr precip	13.74	14
Max 48-hr precip	12.74	13
Sweetgum	11.86	12
Other	10.92	11
Loblolly pine	9.89	10
Water oak	8.47	8
Chestnut oak	8.44	8
Laurel oak	3.97	4
Longleaf pine	3.65	4
White oak	3.11	3
Flowering dogwood	2.26	2
Virginia pine	1.93	2
Slash pine	1.88	2

Table 5.3. Variable importance, normalized (only the added variables are shown) and grouped by clusters.

5.3 Partial dependence

The generation of partial dependence plots can give a much better idea of a covariate's relationship with outages than variable importance alone. Instead of assigning a numerical ranking, partial dependence associates the percent coverage of a tree species with a certain number of outages. This pairing of prevalence with number of outages can give meaningful insights into the role a certain species may play in outage prediction. It can also be helpful in determining which species may result in higher numbers of outages.

5.3.1 Group one (Highest importance)

Of all 11 tree species classes, sweetgum appears to have one of the most significant relationships with power outages. As seen in the partial dependence plot (Figure 5.4), a constant, nearly linear, increase in outages persists until about 70% prevalence. This nearly linear relationship with outages is unique to sweetgum. After about 70% prevalence, sweetgum shows very little association with outages. A lesser relationship with outages also occurs between about 50 and 60% prevalence. This general trend of a decline in outage relationship with increasing prevalence indicates that, in grid cells covered by mostly sweetgum, variables other than sweetgum are affecting the number of outages. This also seems to support the finding from Duryea et al. (2007) that sweetgum is one of the most wind-resistant species. In areas where sweetgum is most prevalent (>70%), outages may be less likely to happen, and if they do occur, it will likely be due to some other variable.
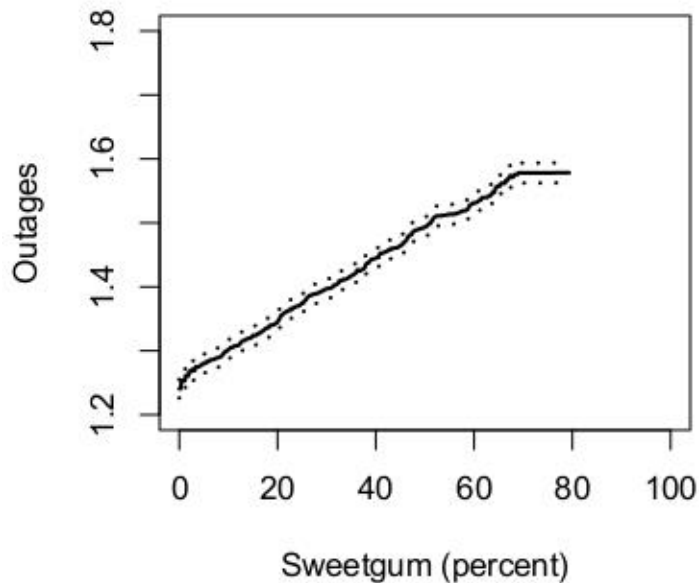
Figure 5.4. Percentage of sweetgum (per grid cell) vs. number of outages.

The Other covariate is associated with areas containing fewer trees, such as farmland, water, or urban areas. The trend of the Other covariate partial dependence plot (Figure 5.5) is drastically different from sweetgum and many of the other tree species. At lower levels of prevalence, the Other covariate appears to have a negative association with outages, indicating that lesser amounts of the Other covariate in a grid cell may mean fewer outages. At higher levels of prevalence, however, and especially above about 80% prevalence, the Other covariate has a very positive association with outages, indicating that higher amounts of the Other covariate probably will lead to greater numbers of outages. This makes sense, especially if the Other covariate is used as a proxy for urban or more developed regions.
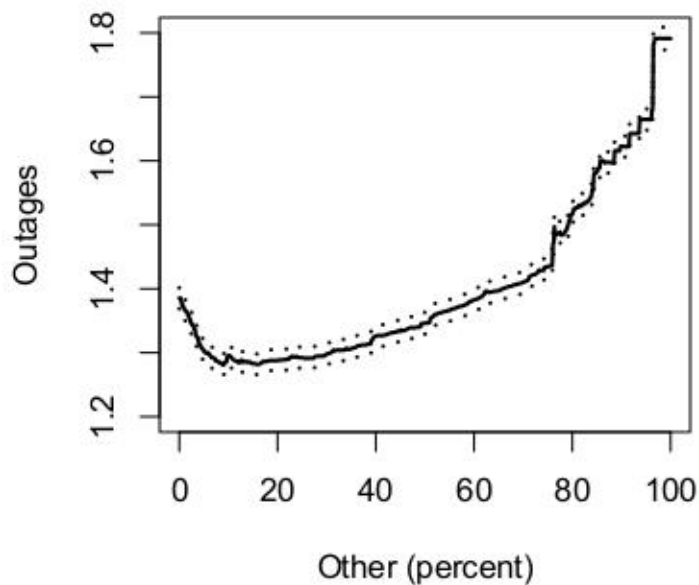
Figure 5.5. Percentage of the Other covariate (per grid cell) vs. number of outages.

Loblolly pine shows a trend very similar to the Other covariate, with a negative relationship at a lower percent prevalence (<10%) and a positive relationship at higher levels of percent prevalence (>80%) (Figure 5.6). This pattern indicates that a higher proportion of loblolly pine may lead to more outages, while the absence of loblolly pine may result in fewer outages. This finding also agrees with Duryea et al. (2007) and other studies that loblolly pine may be a less wind-resistant species, having about a 60% survival rate (Duryea et al. 2007).
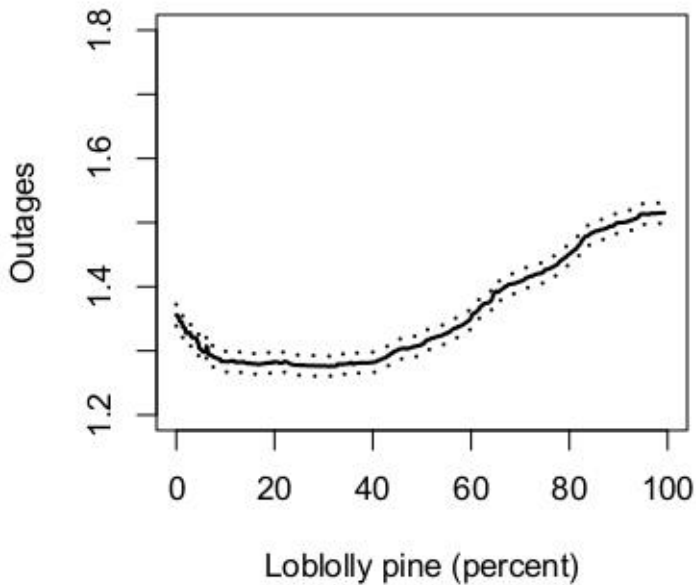
Figure 5.6. Percentage of loblolly pine (per grid cell) vs. number of outages.

5.3.2 Group two (Medium-high importance)

Unlike the previously mentioned tree species classes, changing the proportion of water oak does not have a major impact on the number of outages (Figure 5.7). This could be due to its lack of coverage throughout the service area. At very low levels of percent prevalence, water oak may contribute to lower numbers of outages or have very little relationship with outages at all. However, between about 5% and 30% coverage, water oak does seem to have a positive relationship with the number of outages. Above 30% prevalence, there seems to be little or no relationship. It is difficult to identify the exact association that water oak may have with outages, and a larger sample size of grid cells containing water oak may aid in identifying the nature of the relationship.
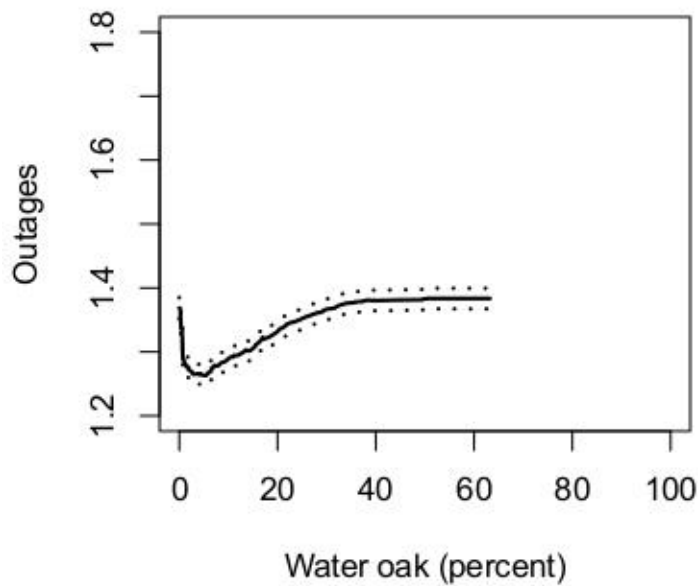
Figure 5.7. Percentage of water oak (per grid cell) vs. number of outages.

Chestnut oak has a strong positive relationship with outages up to about 15% prevalence, after which there is little or no relationship with the number of outages. The shape of this curve is most similar to that of sweetgum. It suggests that as the prevalence of chestnut oak increases, the number of outages remain relatively constant. This agrees with Duryea et al. (2007), which ranks chestnut oak, or swamp chestnut, as having a medium-high wind resistance, just below that of sweetgum.
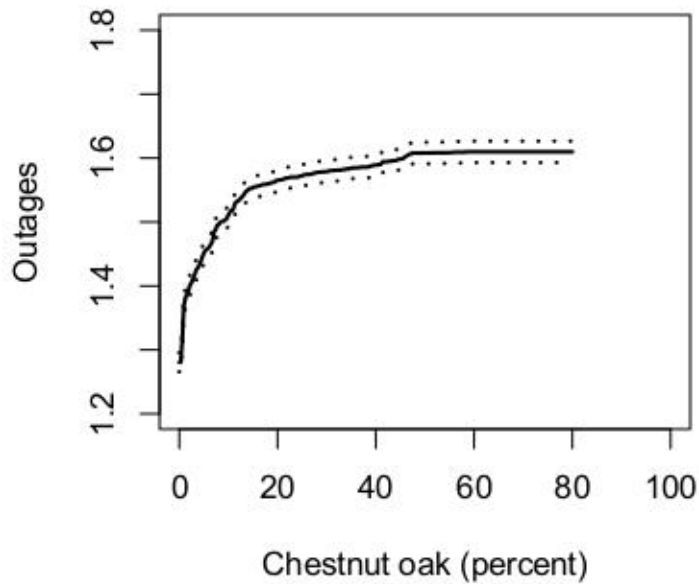
Figure 5.8. Percentage of chestnut oak (per grid cell) vs. number of outages.

5.3.3 Group three (Medium-low importance)

Unlike most of the tree species in groups one and two, the majority of those in group three show little or no relationship with outages regardless of their prevalence. This appears to be the case for both laurel oak and longleaf pine. Laurel oak does show a slight negative relationship with outages at its lowest levels of percent prevalence and a slight positive relationship with outages between 20 and 30 percent prevalence (Figure 5.9). The plot for longleaf pine (Figure 5.10) shows a small positive relationship with outages at lower percentages, but again this is negligible compared with that of the other tree species. As with water oak, a larger sample size of grid cells containing these two tree species may assist in identifying stronger relationships.
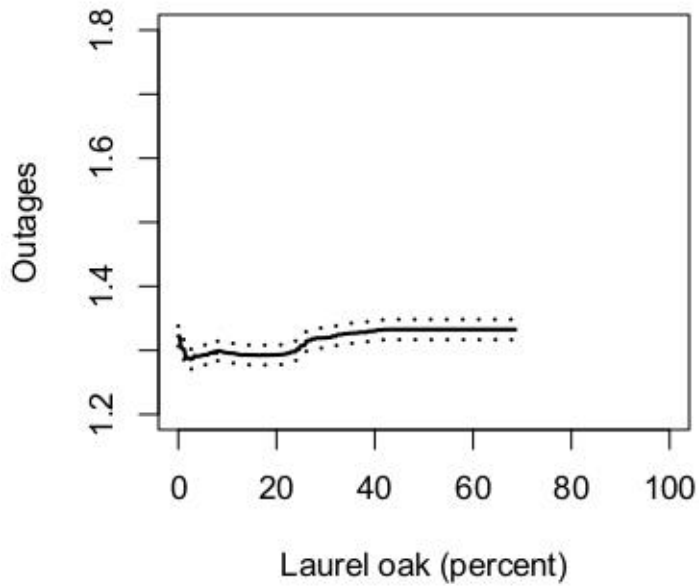
Figure 5.9. Percentage of laurel oak (per grid cell) vs. number of outages.
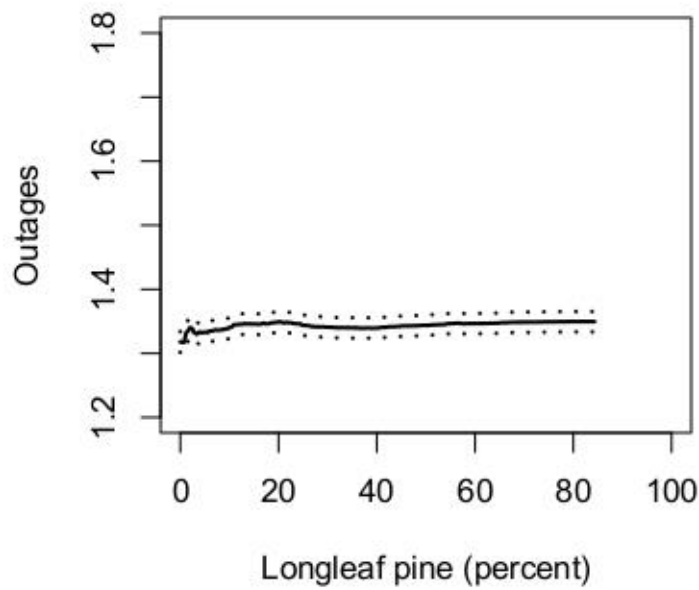
Figure 5.10. Percentage of longleaf pine (per grid cell) vs. number of outages.

White oak has a strong positive relationship with outages up until about 50% prevalence (Figure 5.11). Unlike most other tree species, white oak shows an apparent peak in outages between 40 and 50 percent prevalence. This indicates that a higher presence of white oak may, to some degree, result in a higher number of outages. Above 50% prevalence, white oak shows little or no relationship with the number of outages, which suggests that white oak may prevent the number of outages from increasing in grid cells where it is more widespread. The results from Duryea et al. (2007) and other related studies also seem to be mixed whether white oak is a wind-resistant species. Thus, white oak likely falls somewhere in the middle, between sweetgum and loblolly pine, in terms of wind resistance and its relationship with outages.
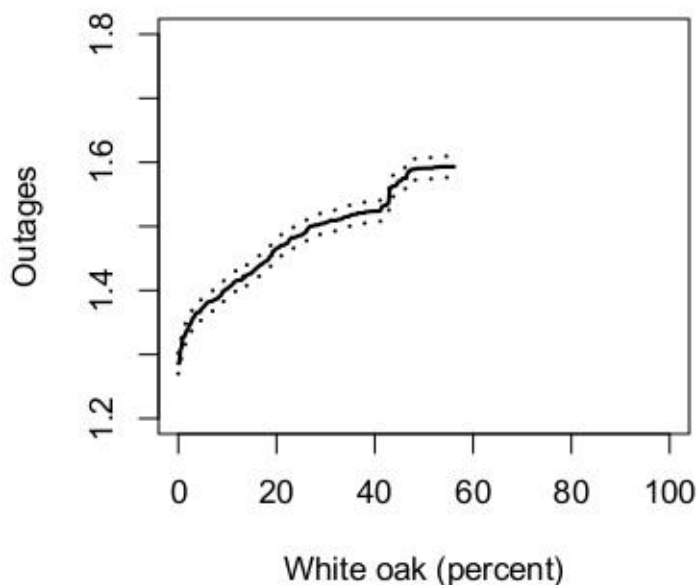
Figure 5.11. Percentage of white oak (per grid cell) vs. number of outages.

5.3.4 Group four (Lowest importance)

Flowering dogwood ranks as one of the least important covariates. As the prevalence of flowering dogwood increases up to 40%, there is a slight increase in the number of outages (Figure 5.12). After about 50% prevalence, the partial dependence plot levels off. Similar patterns were observed with sweetgum, chestnut oak, and white oak, all of which showed moderate-to-high levels of wind resistance (Duryea et al. 2007).
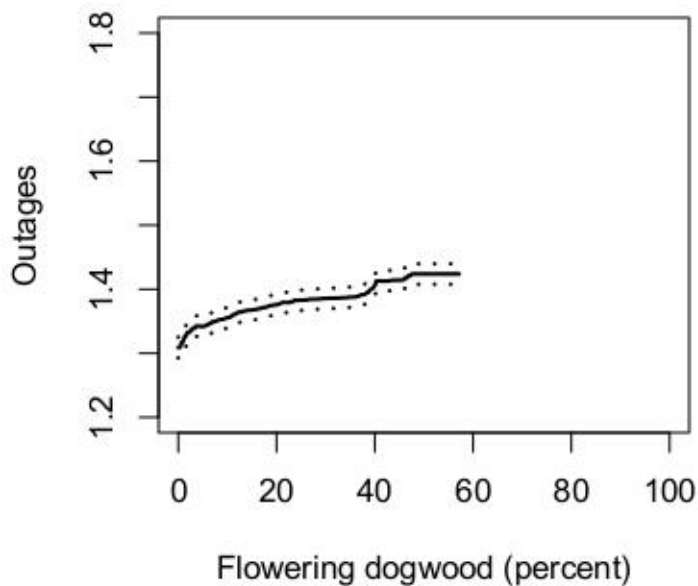
Figure 5.12. Percentage of flowering dogwood (per grid cell) vs. number of outages.

The partial dependence plot for Virginia pine closely resembles that of white oak. Two areas with higher slopes are visible below 30% prevalence, with little to no

relationship with outages thereafter (Figure 5.13). This pattern suggests that other

covariates may be contributing to the number of outages in grid cells where Virginia

pine is most prevalent.
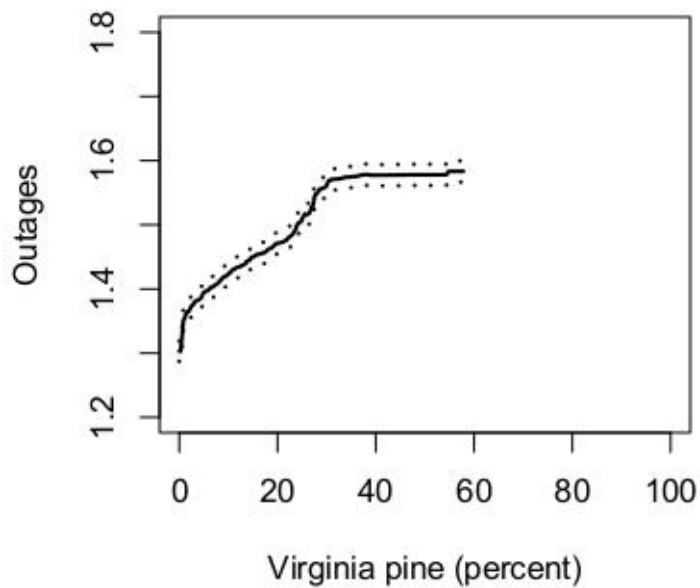
Figure 5.13. Percentage of Virginia pine (per grid cell) vs. number of outages.

Similar to laurel oak and longleaf pine, slash pine shows little or no relationship

with outages across all levels of prevalence (Figure 5.14). This could again be due to a

lack of grid cells containing slash pine.
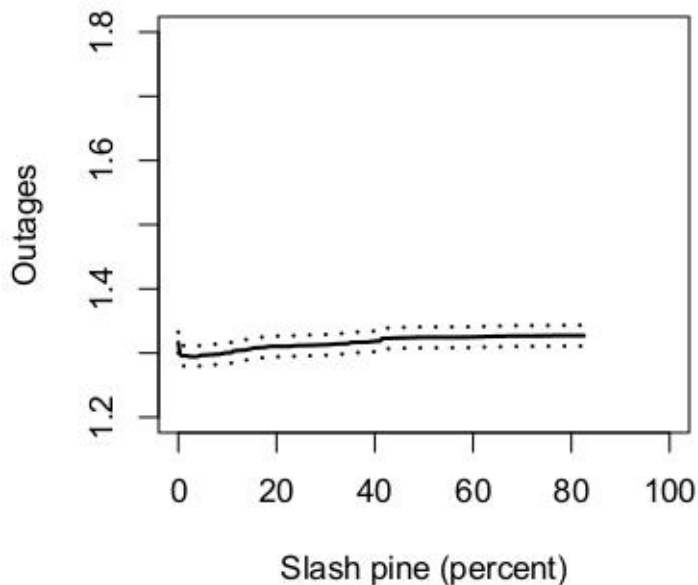
Figure 5.14. Percentage of slash pine (per grid cell) vs. number of outages.

5.3.5 Discussion

According to the variable importance and partial dependence results, tree species covariates contributing most significantly to the model include sweetgum, other, and loblolly pine. Chestnut oak, white oak, and Virginia pine also show strong relationships to outages in the partial dependence plots and may be useful predictors. Also, there appears to be a relatively strong spatial relationship between tree species and outages, with species in central and northern portions of the service area (i.e., sweetgum, loblolly pine, chestnut oak, and white oak) showing a significantly stronger relationship with outages than species in southern portions of the service area (i.e., laurel oak, longleaf pine, and slash pine) (Figure 5.15). Many other confounding factors may also contribute

to such a relationship, including proximity to coastline, precipitation, and storm track and intensity.
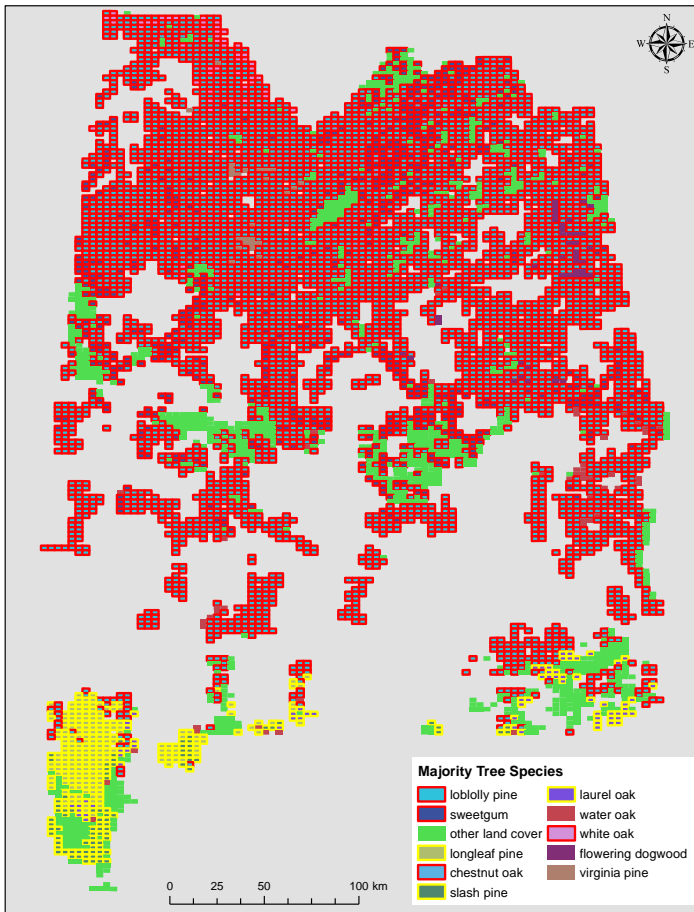
Figure 5.15. Outages and tree species. Species with a stronger relationship to outages are outlined in red. Species with little to no outage relationship are outlined in yellow.

To aid in further identification of relationships between individual tree species covariates and outages, a comparison between percent prevalence and outage relationship is done for each tree species (Table 5.4). A high negative relationship with

outages indicates that the species, at that level of prevalence, results in a decrease in outages. A high positive relationship shows that the species, at that level of prevalence, may lead to an increase in outages. Species showing either a negative relationship or no relationship at greater than 50% prevalence and a positive relationship below 50% prevalence are highlighted in light blue. They include sweetgum, chestnut oak, white oak, and flowering dogwood. These species do not contribute to an increase in outages at high levels of prevalence, which may be indicative of a higher wind resistance. At lower levels of prevalence, they do show a positive relationship with outages, but this is likely due to the influence of other covariates or other outside factors. Species showing a positive relationship with outages at greater than 50% prevalence and a negative relationship below 50% prevalence are highlighted in red. These species include the Other covariate, representative of farmland and developed areas, and loblolly pine. These are covariates that may contribute to an increase in outages at higher levels of prevalence. These findings seem to agree with those of past studies, including Duryea et al. (2007).

Predictor	Highest (-) relationship, ↓outages	Little/no relationship	Highest (+) relationship, ↑outages	Sig (Δ outages)
Sweetgum	-	> 70%	0-50%	~0.4
Other	0-10%	10-30%	80-100%	~0.5
Loblolly pine	0-10%	10-40%	50-90%	~0.2
Water oak	0-5%	40-60%	5-30%	~0.1
Chestnut oak	-	50-80%	0-10%	~0.3
Laurel oak	0-3%	40-70%	-	~0.0
Longleaf pine	-	3-90%	0-3%	~0.0
White oak	-	50-60%	0-30%, 40-50%	~0.3
Flowering dogwood	-	50-60%	0-20%, 40-50%	~0.1
Virginia pine	-	30-60%	0-30%	~0.3
Slash pine	0-3%	3-80%	-	~0.0

Table 5.4. Percent prevalence and outage relationship comparison. A higher (-) relationship corresponds to a decrease in outages, while a higher (+) relationship signifies an increase in outages. Red species covariates may be associated with more outages, blue species covariates may be related to fewer outages, and gray species covariates have little or no influence on outages.

5.4 Local variable importance

The local variable importance is calculated for each of the more significant tree species predictors (highlighted in Table 5.4). LVI may provide more details about the spatial distribution of importance for the selected covariates. Scores are calculated for each grid cell by taking the mean importance of all four storms for the selected covariate and then dividing that value by the sum of the mean importance scores for all of the covariates. The result is a percentage importance value that indicates the importance, or contribution, of the covariate in relation to the other covariates in the model. A negative percent LVI value indicates a lack of importance for that grid cell and shows where the covariate may have resulted in a decrease in model improvement. A zero value indicates

little or no importance, while an increasingly positive value specifies that the covariate may be a reasonable predictor for that particular grid cell and result in some model improvement.

In Figure 5.16 the percent LVI for each of the six tree species is displayed in terms of standard deviation from the mean. This classification should help identify areas of most significant importance. Of the six species, sweetgum and chestnut oak show the most cohesive clusters of significant importance. For sweetgum, an area of relatively high importance is concentrated within a major urban area in the southwest portion of the service area. For chestnut oak, several significant areas of high importance are visible. The main center is located in the southwest portion of the service area and includes a major city; another is situated just to the northeast of that region, and yet another is focused along a band extending through the center of the service area.

The four other tree species show less significant areas of importance. However, the white oak and the Other covariates both appear to have an area of higher importance extending to the south from the northernmost urban area. Loblolly pine and flowering dogwood also show some areas of slightly higher importance on the northwestern outskirts of a major urban area in the southwest corner of the service area.
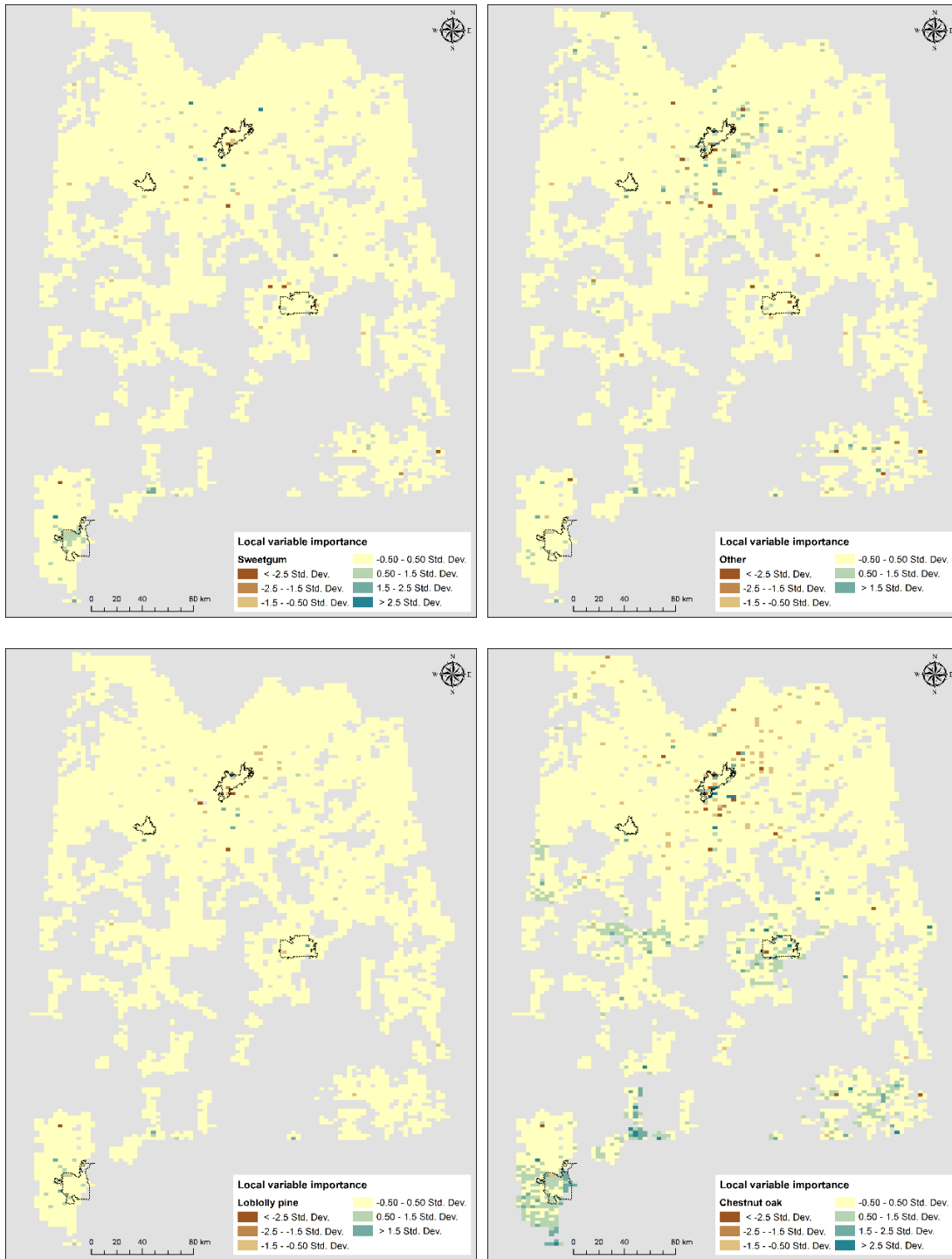
Figure 5.16. Local variable importance (LVI) for selected tree species. LVI is calculated as the % LVI of a covariate out of the total LVI for all covariates (old and new) in the model. Major urban areas are outlined in black.
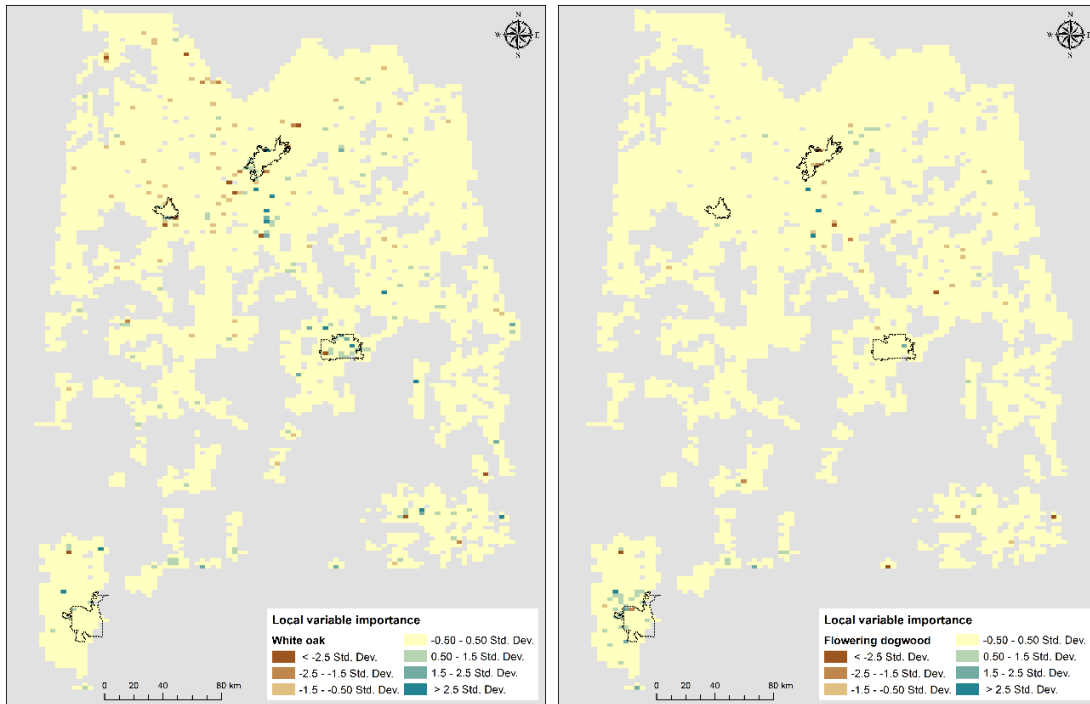
Figure 5.16 Continued.

To better illustrate the relationship of tree species importance with the importance of the other covariates, maps are generated that compare the highest percent LVI values for tree species with the percent LVI values for all new covariates (Figure 5.17a) and for all covariates (old and new) combined (Figure 5.17b). In both cases, only covariates with the highest percent LVI value are shown for each grid cell. Chestnut oak, loblolly pine, sweetgum, and the Other covariate all seem to have the most widespread coverage of higher LVI values (Figure 5.17b), and this trend agrees with the partial dependence and LVI plots above. Also, the majority of species with the highest importance tend to be concentrated in the southwest corner of the service area, near the

coast (Figure 5.17a). The only major exceptions to this are chestnut oak and the Other covariate.
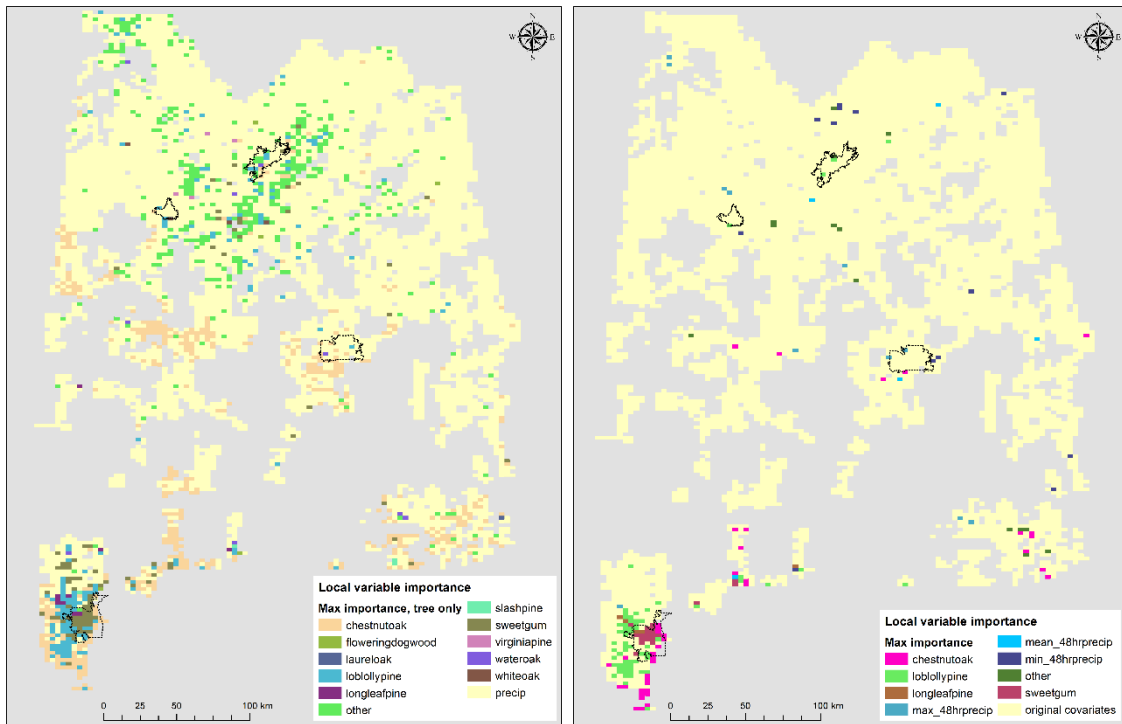
Figure 5.17. a. Tree species with the highest % LVI (out of all new covariates) by grid cell. **b.** New covariates with the highest % LVI out of all covariates (old and new). Major urban areas are outline in black.

In summary, chestnut oak, loblolly pine, and sweetgum clearly show more significant clusters of higher importance, which indicates that these species may have a significant effect on outages in these areas. For sweetgum, the areas of highest importance appear to be urban regions, while for chestnut oak they are more widespread, affecting urban and rural regions almost equally. The presence of chestnut oak may also

97

be a significant factor in outages near the coast. The lack of clustering of the other covariates does not necessarily indicate these covariates are of lesser importance in the model, but that their importance and contribution to the model may be more uniform across the service area. Also, tree species in general appears to have a significantly higher importance in the southwest corner of the service area, where it is also a more important predictor of outages than almost all of the other covariates (old and new).

CHAPTER VI

EVALUATION OF 48-HOUR PRECIPITATION

6.1 Error and outage prediction performance

The addition of 48-hour precipitation to the Reduced model results in some improvement, but only about half as much as with tree species or tree species and precipitation combined (Tables 4.1 and 4.2). This suggests that 48-hour precipitation may play a role in improving model performance, particularly when accompanied by tree species.

An assessment of the percent differences between the higher and lower performing grid cells shows strong positive values for 48-hour precipitation across all SD's (Table 5.1). Positive percent difference values indicate that the model is improving with the addition of the new variable. The results also show that 48-hour precipitation contributes to the greatest overall improvement in model performance, particularly at lower SD's, or in grid cells showing lower levels of improvement. This suggests that the addition of 48-hour precipitation may serve to fill in some of the gaps and improve accuracy in grid cells where tree species does not.

The "hot spot" map in Figure 5.2 identifies clusters where the model performs better (lower polarized difference in AE) or worse (higher polarized difference in AE) with 48-hour precipitation added. Four main areas of improvement are apparent,

primarily occurring in or near major cities. This trend makes sense as the vast majority

of power system equipment is located in areas with more customers, or higher

population densities, and as a result greater improvements are likely to occur within

those grid cells. However, the addition of 48-hour precipitation also seems to result in

more improvement in larger cities than the addition of tree species. This suggests that a

higher number of customers may not be the only factor influencing improvement, but

that 48-hour precipitation may actually play a significant role in outage prediction for the

major cities within the service area.

In addition to the areas of improvement, one rather large area of declining

performance is also apparent on the western side of a major city in the southwest corner

of the service area. A variety of factors could contribute to such a decline, including the

relatively small sample of hurricanes used to train the model. Inclusion of more storms,

especially ones that track closest to the regions of highest error, may help to reduce or

even eliminate areas such as this in the future.

Classification of the change in MAE by outage group (Table 5.2) confirms many

of the relationships found in Figure 5.2. In areas with a greater number of outages, the

addition of 48-hour precipitation tends to result in the greatest model improvement;

these would typically be areas with a larger number of customers, or major cities.

However, even in grid cells with lower numbers of outages, 48-hour precipitation still

contributes significantly to model improvement, falling just behind tree species in terms

of negative change in MAE.

In summary, the addition of 48-hour precipitation results in less overall model improvement than tree species. However, it does perform significantly better than tree species in major urban areas, and this may serve to fill in some of the gaps that tree species may have missed. A greater sample set of storms for the training dataset will also likely improve performance of 48-hour precipitation in the model.

6.2 Variable importance

The variable importance graph (Figure 5.3) below shows that all three 48-hour precipitation covariates rank higher than the tree species covariates, but lower than all of the original covariates. Mean and minimum 48-hour precipitation have nearly the same importance scores, while maximum 48-hour precipitation seems to play a slightly less significant role in outage prediction for the model. The 48-hour precipitation covariates also fall in close proximity to the two soil moisture covariates, indicating that some relationship between 48-hour precipitation and soil moisture probably exists. As noted in the previous section, 48-hour precipitation still results in significant model improvement, even with soil moisture present.

The normalized importance scores in Table 6.1 show the similarity of the mean and minimum 48-hour precipitation covariates, with the maximum 48-hour precipitation covariate falling well below the other two. This suggests that when adding 48-hour precipitation to future versions of the HOPM, only one 48-hour precipitation covariate

would need to be included, with the mean or minimum being preferred. Also, the two soil moisture covariates are clearly in their own group.

Predictor	Normalized importance	Normalized importance (rounded)
Soil moisture 3	19.02	19
Soil moisture 1	18.16	18
Mean 48-hr precip	13.78	14
Min 48-hr precip	13.74	14
Max 48-hr precip	12.74	13
Sweetgum	11.86	12
Other	10.92	11
Loblolly pine	9.89	10

Table 6.1. Variable importance, normalized and grouped by clusters.

6.3 Partial dependence

6.3.1 Comparison

All three 48-hour precipitation covariates have a significant positive relationship with outages below ~7 inches (Figures 6.1, 6.2, and 6.3). This trend shows that low to moderate amounts of precipitation falling over the course of a hurricane in the service area contribute the most to outages. Above ~7 inches of rain, however, the precipitation does not appear to be strongly related to outages. Thus, there appears to be somewhat of an upper limit on the influence of precipitation. Minimum 48-hour precipitation seems to result in the most significant change in outages (Figure 6.2), followed by mean 48-hour precipitation (Figure 6.1), with maximum 48-hour precipitation resulting in the least

amount of change (Figure 6.3). Minimum 48-hour precipitation has the steepest curve, which is indicative of a strong relationship with outages.
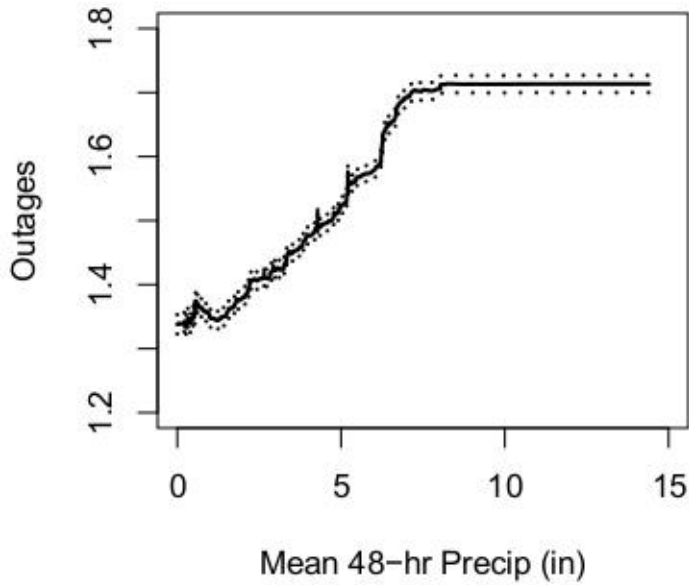
Figure 6.1. Mean 48-hour precipitation (by grid cell) vs. number of outages.
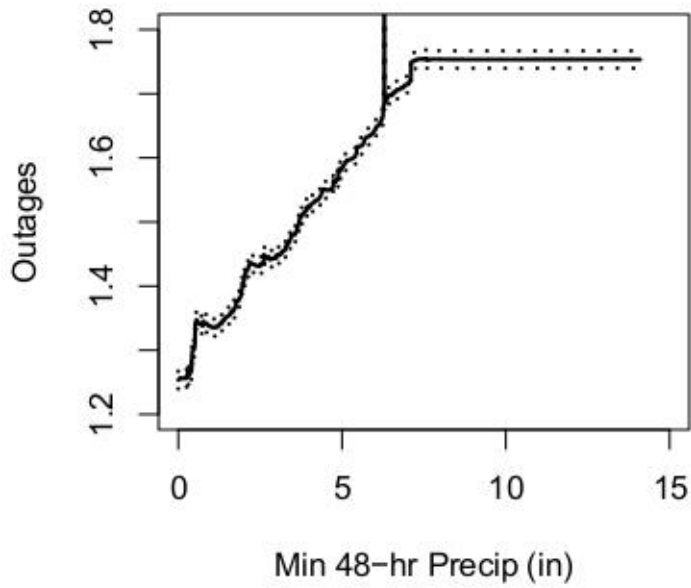
Figure 6.2. Minimum 48-hour precipitation (by grid cell) vs. number of outages.
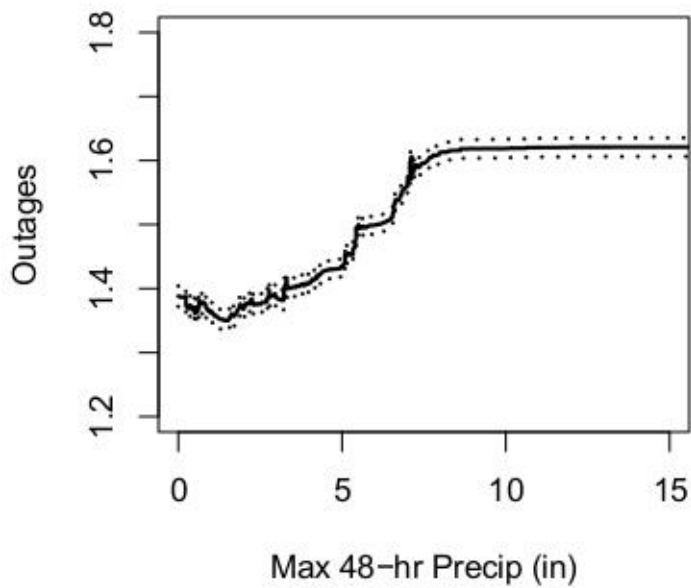
Figure 6.3. Maximum 48-hour precipitation (by grid cell) vs. number of outages.

6.3.2 Discussion

Based on the above analysis, precipitation can significantly increase outages, but only up to ~7 inches, after which there is little effect. As the three precipitation covariates are strongly related, only one would need to be used in the HOPM. This would probably be either the mean or the minimum, as they both rank highest in variable importance for the model. The mean may be the best representative of 48-hour precipitation and the ideal covariate to include in the model, especially since it ranks highest of all the 48-hour precipitation covariates in importance.

6.4 Local variable importance

Local variable importance (LVI) may provide insights regarding specific locations in the service area where 48-hour precipitation has the greatest impact on outage prediction. It is calculated for each of the three 48-hour precipitation covariates using the same method described previously for tree species. In Figure 6.4, percent LVI is displayed in terms of standard deviation from the mean. This classification will hopefully serve to highlight areas of greatest and least importance. A positive percent LVI value indicates that the covariate has greater importance in that particular grid cell, while a negative percent LVI value shows a lack of importance. A value close to zero is indicative of little or no importance. Covariates with higher LVI values in a grid cell may be a significant predictors of outages for that grid cell.

Of the three 48-hour precipitation covariates, the minimum 48-hour precipitation covariate appears to show a more cohesive spatial pattern of importance (Figure 6.4). Areas of greatest importance for this covariate tend to occur within major urban areas and over the northern portion of the service area. Mean and maximum 48-hour precipitation both show a similar pattern, though not as pronounced and with fewer significant clusters in the northern portion of the service area.
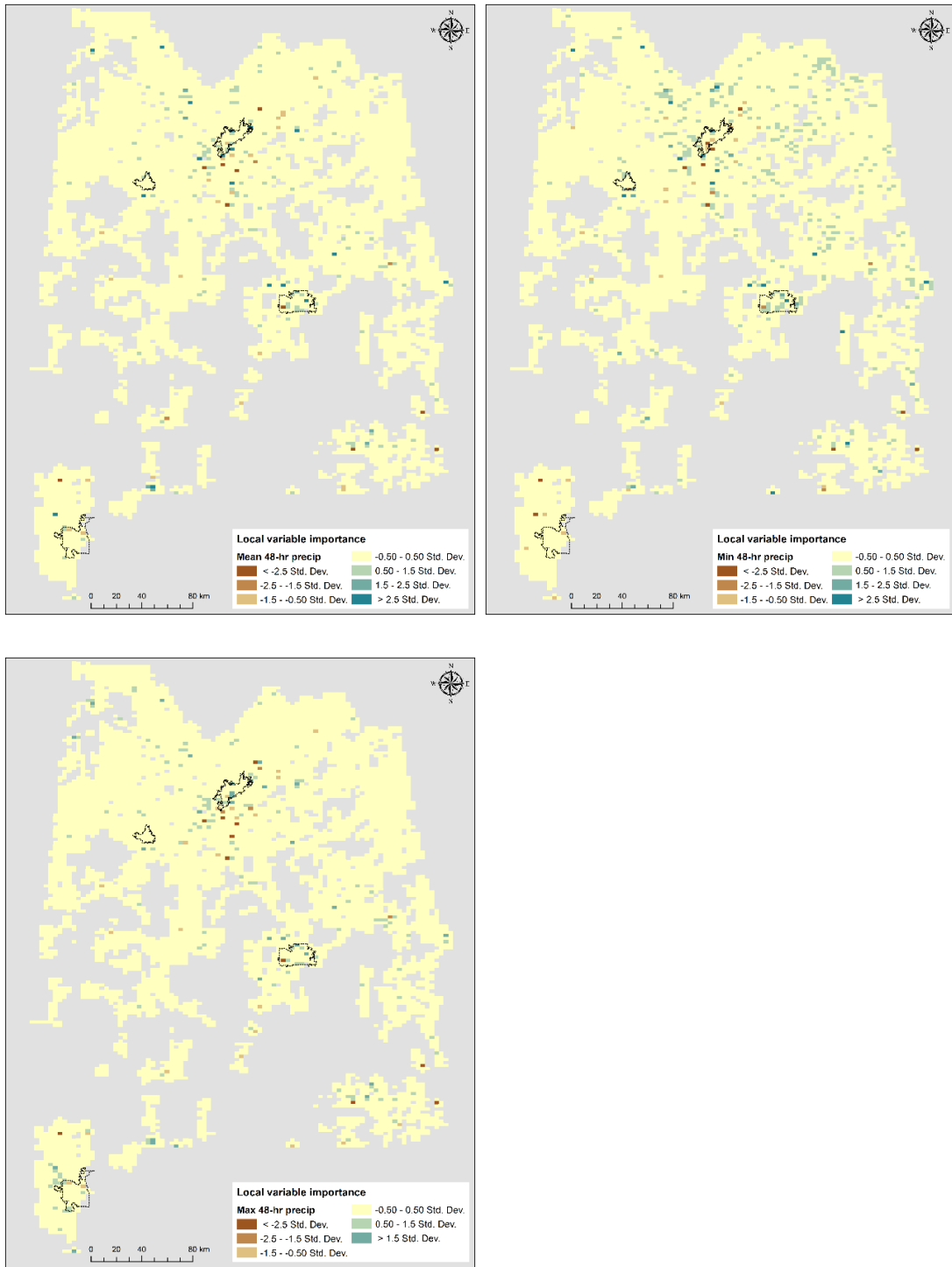
Figure 6.4. Local variable importance (LVI) for 48-hour precipitation. LVI is calculated as the % LVI of a covariate out of the total LVI for all covariates (old and new) in the model. Major urban areas are outlined in black.

To further examine the relationship of 48-hour precipitation importance with the importance of the other covariates, maps displaying 48-hour precipitation with the highest percent LVI values out of all new covariates (Figure 6.5a) and the highest percent LVI values out of all covariates (Figure 6.5b) are produced. A clear pattern is evident where the minimum 48-hour precipitation covariate holds greater importance over northeastern sections of the service area, while the maximum 48-hour precipitation covariate seems to have more influence in central portions. Importance for mean 48-hour precipitation appears to be dispersed fairly equally throughout the service area. All precipitation covariates show the least importance in the far southwest corner of the service area, which is also where tree species tends to have the greatest influence on outage prediction.

In summary, 48-hour precipitation may be a better predictor of outages in or near major cities as well as in central and northern portions of the service area. Also, the minimum 48-hour precipitation may have a greater influence, and be a better predictor of outages, over the northern tier of the service area. 48-hour precipitation tends to have less of an impact on outage prediction in the southwest corner of the service area, which is where tree species holds the highest importance for the model.
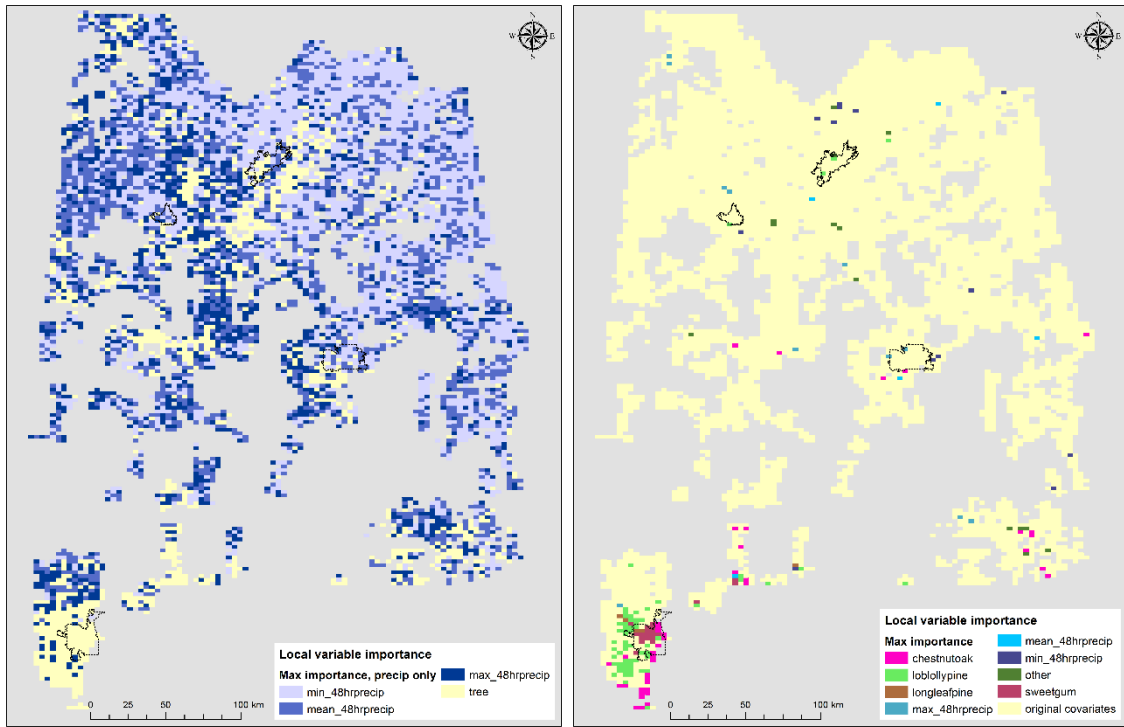
Figure 6.5. a. 48-hour precipitation with the highest % LVI (out of all new covariates) by grid cell. **b.** New covariates with the highest % LVI out of all covariates (old and new). Major urban areas are outline in black.

CHAPTER VII

CONCLUSION

7.1 Summary

This thesis has examined the importance of two new variables, tree species and storm-derived (48-hour) precipitation, on HOPM prediction of hurricane-induced power outages. In Objective 1, the new variables were added to several different versions of the HOPM. An optimal model was then selected for use in the analysis, which was the Reduced (six-variable) model used by Nateghi et al. (2014). With the new variables added, this model performed as well as the All Variables model. This may be useful in determining which version of the HOPM to use in the future.

Objective 2 focused on tree species and assessed how adding new tree variables influenced model performance. Comparisons were made with and without tree species added to see its effect on the model. In general, tree species tended to improve the model more than 48-hour precipitation, especially in non-urban areas and in the southwest corner of the service area. To assess the impact of specific tree species on outages, partial dependence plots and maps of variable importance were also generated. Using these results, it was possible to determine how individual tree species influenced power outages. According to variable importance and partial dependence, sweetgum and non-treed areas (Other) had the greatest overall impact on outages, while laurel oak, longleaf pine, and slash pine had the least impact. Species that had the greatest impact on outages

110

were typically more prevalent in central and northern portions of the service area, while species with little or no impact on outages were more prevalent near the coast in the southern portion of the study region. Certain tree species, such as sweetgum, chestnut oak, white oak, and flowering dogwood, typically have a higher wind-resistance and so there may be fewer outages in locations where these species are most common. Other tree species (i.e., loblolly pine) tend to have a lower wind-resistance and may result in more outages in locations where they are highly concentrated.

Objective 3 examined how the total precipitation for the 48-hour period surrounding landfall (pre- and post-landfall) influenced outages. Unlike tree species, 48-hour precipitation appeared to have a greater impact in urban areas, particularly in central and northern portions of the service area, and model improvements were almost exclusively limited to these locations. Based on the partial dependence plot, 48-hour precipitation was most strongly related to outages when rainfall was less than 7 inches. The slope of the relationship was relatively flat when there was more than 7 inches of rainfall which suggests that there is a threshold above which additional rainfall does not lead to additional outages. The results also showed that the mean or minimum 48-hour precipitation are better predictors of outages than the maximum 48-hour precipitation.

7.2 Implications

The prediction of power outages in advance of a land-falling hurricane has many benefits for utility companies, and potentially emergency managers and the general

public as well. Utility companies generally lack sufficient personnel to efficiently restore power after a major event such as a hurricane, and they often must call on other neighboring utility providers for assistance. Prior to doing so, the utility company requesting assistance must first estimate the number of personnel and the amount of equipment that may be required to restore power quickly and effectively. A delicate balance must be maintained, as requesting more assistance than is needed could result in unnecessary expenditures, while requesting too little assistance may result in longer power restoration times. Thus, an accurate power outage prediction model, along with a knowledge of the key drivers of outages, will be beneficial for utility planning prior to such storm events. This study has sought to meet both of those requirements and, along with an improved version of the HOPM, provide an analysis of how tree species and precipitation influence outages.

First, the improved version of the HOPM developed from this thesis contains significantly fewer covariates than the All Variables model. This will benefit utility companies by reducing runtime, while providing more accurate outage predictions.

Next, the variable importance and partial dependence plots helped to identify tree species which have more of an impact on outages. These species can then be included in future versions of the HOPM, thus reducing the number of tree species covariates needed to run the model while maintaining a high level of accuracy. Also, the identification of locations in the service area where a highly susceptible tree species is located may aid in pinpointing specific areas where further preventative action, such as additional tree trimming, may be needed. Likewise, if a location contains a tree species

that generally has greater wind resistance then less frequent tree trimming may be required.

Lastly, an improved understanding of how storm-derived (48-hour) precipitation influences outages may be helpful for pre-storm planning and mitigation efforts. Increases in rainfall were associated with increases in outages. Therefore, additional crews may need to be dispatched to areas forecast to receive greater amounts of precipitation. It also appears that there are certain regions within the service area where storm-derived precipitation may be more of a factor than others. These locations are useful to know for assisting with preparedness and mitigation efforts.

7.3 Future improvements and research

As with any research study, there are a variety of alterations that could be made to improve both the methodology and the results. Most of the improvements mentioned below pertain to either the two input datasets or the methods used to extract these datasets.

For tree species, a slightly different approach would be to generate a new covariate for each species along roads. It would be fairly easy to obtain a detailed road dataset for the service area, and then in ArcGIS create appropriate-sized buffers around each of the road centerlines. The tree species dataset could then be clipped to the buffers, so that only percent tree species near roads would be included for each grid cell. A 30-meter resolution version of the tree species dataset is also available from FHTET, and

113

this would allow for even greater accuracy when clipping out the tree species. The premise behind this new predictor would be that distribution lines and poles are typically found near roads to facilitate access by power crews. Thus, the percent coverage of tree species near roads would likely be a better predictor of outages than the percent coverage across the entire grid cell. Also, the addition of more tree species covariates or the inclusion of other tree-related predictors (i.e., age, health, height, root structure, and DBH) may also serve to enhance model performance.

For storm-derived precipitation, perhaps a less-biased and more spatially consistent dataset, such as PRISM (http://www.prism.oregonstate.edu), could be used instead of the NCEP Stage II/IV data. Though PRISM is only available on a daily basis, the quality of the data appears to be considerably improved over the NCEP data, with fewer gaps and interpolated regions, and daily data should still give a sufficiently accurate representation of total precipitation surrounding the time of landfall. The data are also available to order at an 800-meter resolution as opposed to a 4-kilometer resolution. PRISM data are available back to 1981, so the use of this dataset may also allow for the inclusion of more storms in the model. This would likely result in greater accuracy, not just for precipitation, but for all hurricane-related variables.

Further model improvement may also result from the inclusion of data from neighboring states in the utility company service area. This will most likely produce a greater number of tree species covariates for analysis, in addition to adding more diversity to the other predictors in the model.

7.4 Conclusion

The accurate and timely prediction of power outages prior to hurricane landfall can have a significant and far-reaching impact on utility companies as well as the general public. Tree species and storm-derived (48-hour) precipitation are both influential predictors of power outages within this utility company's service area, and the addition of these variables to the HOPM can significantly improve model performance. Also, certain regions of the service area may be more susceptible to outages due to the influence of these two variables during a hurricane.

REFERENCES

Balijepalli, N., S. S. Venkata, C. W. Richter Jr, R. D. Christie, and V. J. Longo, 2005: Distribution system reliability assessment due to lightning storms. *IEEE Transactions on Power Delivery*, **20**, 2153-2159.

Barry, P. J., C. Doggett, R. L. Anderson, and K. M. Swain Sr, 1993: How to evaluate and manage storm-damaged forest areas. USDA Forest Service Region 8. Management Bulletin R8-MB-63, [Available online at http://www.forestpests.org/storm/.]

Benyamin, D., 2012: A Gentle Introduction to Random Forests, Ensembles, and Performance Metrics in a Commercial System. Accessed 10 Nov 2014. [Available online at https://citizennet.com/blog/2012/11/10/random-forests-ensembles-and-performance-metrics/.]

Beven, J., 2005: Tropical Cyclone Report: Hurricane Dennis, 4-13 July 2005. National Hurricane Center, 25 pp, [Available online at http://www.nhc.noaa.gov/data/tcr/AL042005_Dennis.pdf.]

Broström, E., and L. Söder, 2007: Ice storm impact on power system reliability. *12th International Workshop on Atmospheric Icing on Structures*, Yokohama, Japan. [Available online at http://www.diva-portal.org/smash/get/diva2:467566/FULLTEXT01.pdf.]

Brown, R., S. Gupta, R. Christie, S. Venkata, and R. Fletcher, 1997: Distribution system reliability assessment: momentary interruptions and storms. *IEEE Transactions on Power Delivery*, **12**, 1569-1575.

Cerruti, B. J., and S. G. Decker, 2011: A Statistical Forecast Model of Weather-Related Damage to a Major Electric Utility. *Journal of Applied Meteorology and Climatology*, **51**, 191-204, doi: 10.1175/JAMC-D-11-09.1.

Chai, T., and R. R. Draxler, 2014: Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, **7**, 1247-1250.

Conner, W. H., I. Mihalia, and J. Wolfe, 2002: Tree community structure and changes from 1987 to 1999 in three Louisiana and three South Carolina forested wetlands. *Wetlands*, **22**, 58-70, doi: 10.1672/0277-5212(2002)022[0058:TCSACF]2.0.CO;2.

Davidson, R. A., H. Liu, I. K. Sarpong, P. Sparks, and D. V. Rosowsky, 2003: Electric power distribution system performance in Carolina hurricanes. *Natural Hazards Review*, **4**, 36-45.

DeGaetano, A. T., B. N. Belcher, and P. L. Spier, 2008: Short-term ice accretion forecasts for electric utilities using the weather research and forecasting model and a modified precipitation-type algorithm. *Weather and Forecasting*, **23**, 838-853.

Duryea, M. L., 1997: Wind and trees: Surveys of tree damage in the Florida Panhandle after Hurricanes Erin and Opal. University of Florida Cooperative Extension Service. Circular 1183, 7 pp, [Available online at http://purl.fcla.edu/UF/lib/FR010.]

Duryea, M. L., and E. Kampf, 2007: Wind and Trees: Lessons Learned from Hurricanes. Urban Forest Hurricane Recovery Program Series. Publ. 118, 17 pp, [Available online at http://edis.ifas.ufl.edu/fr173.]

Duryea, M. L., E. Kampf, and R. C. Littell, 2007: Hurricanes and the urban forest: I. Effects on southeastern United States coastal plain tree species. *Arboriculture and Urban Forestry*, **33**, 83-97.

Emanuel, K. A., 2005: *Divine wind : the history and science of hurricanes.* Oxford University Press, 285 pp.

Fredericksen, T., R. Hedden, and S. Williams, 1993: Testing loblolly pine wind firmness with simulated wind stress. *Canadian Journal of Forest Research*, **23**, 1760-1765.

Garrigues, M. W., Z. Fan, D. L. Evans, S. D. Roberts, and W. H. Cooke, 2012: Geospatial Relationships of Tree Species Damage Caused by Hurricane Katrina in South Mississippi. *Proceedings of the 16th Biennial Southern Silvicultural Research Conference*, Charleston, SC, USDA Forest Service Southern Research Station, 127-132. [Available online at http://www.srs.fs.usda.gov/pubs/41423.]

Glitzenstein, J. S., and P. A. Harcombe, 1988: Effects of the December 1983 tornado on forest vegetation of the Big Thicket, southeast Texas, USA. *Forest Ecology and Management*, **25**, 269-290.

Gresham, C. A., T. M. Williams, and D. J. Lipscomb, 1991: Hurricane Hugo wind damage to southeastern U.S. coastal forest tree species. *Biotropica*, **23**, 420-426.

Grömping, U., 2009: Variable importance assessment in regression: linear regression versus random forest. *The American Statistician*, **63**, 308-319, doi: 10.1198/tast.2009.08199.

Guikema, S. D., and S. M. Quiring, 2012: Hybrid data mining-regression for infrastructure risk assessment based on zero-inflated data. *Reliability Engineering & System Safety*, **99**, 178-182, doi: 10.1016/j.ress.2011.10.012.

Guikema, S. D., R. A. Davidson, and H. Liu, 2006: Statistical models of the effects of tree trimming on power system outages. *IEEE Transactions on Power Delivery*, **21**, 1549-1557.

Guikema, S. D., S. M. Quiring, and S. R. Han, 2010: Prestorm estimation of hurricane damage to electric power distribution systems. *Risk analysis : an official publication of the Society for Risk Analysis*, **30**, 1744-1752, doi: 10.1111/j.1539-6924.2010.01510.x.

Han, S. R., S. D. Guikema, and S. M. Quiring, 2009a: Improving the predictive accuracy of hurricane power outage forecasts using generalized additive models. *Risk analysis : an official publication of the Society for Risk Analysis*, **29**, 1443-1453, doi: 10.1111/j.1539-6924.2009.01280.x.

Han, S. R., S. D. Guikema, S. M. Quiring, K. H. Lee, D. Rosowsky, and R. A. Davidson, 2009b: Estimating the spatial distribution of power outages during hurricanes in the Gulf coast region. *Reliability Engineering & System Safety*, **94**, 199-210, doi: 10.1016/j.ress.2008.02.018.

Hook, D. D., M. A. Buford, and T. M. Williams, 1991: Impact of Hurricane Hugo on the South Carolina coastal plain forest. *Journal of Coastal Research*, **8**, 291-300.

Horvath, S., T. Shi, R. M. Shai, C. Chen, and S. Nelson, 2007: Statistical Methods Supplement and R software tutorial: Gene Filtering with a Random Forest Predictor. Accessed 15 Feb 2015. [Available online at http://labs.genetics.ucla.edu/horvath/RandomForestScreening/SupplementRandomForests.pdf.]

Johnsen, K. H., J. R. Butnor, J. S. Kush, R. C. Schmidtling, and C. D. Nelson, 2009: Hurricane Katrina winds damaged longleaf pine less than loblolly pine. *Southern Journal of Applied Forestry*, **33**, 178-181.

Jones, K. F., 1996: Ice accretion in freezing rain. U.S. Army Corps of Engineers. CRREL Report 96-2, 23 pp, [Available online at http://www.dtic.mil/dtic/tr/fulltext/u2/a310659.pdf.]

Knabb, R. D., J. R. Rhome, and D. P. Brown, 2005: Tropical Cyclone Report: Hurricane Katrina, 23-30 August 2005. National Hurricane Center, 43 pp, [Available online at http://www.nhc.noaa.gov/data/tcr/AL122005_Katrina.pdf.]

Krishnamurthy, V., and A. Kwasinski, 2013: Characterization of power system outages caused by hurricanes through localized intensity indices. *Power and Energy Society General Meeting*, Vancouver, B.C., Institute of Electrical and Electronics Engineers, 1-5, doi: 10.1109/IGARSS.2006.493.

Krist, F. J., and Coauthors, 2012: 2013-2027 National Insect and Disease Forest Risk Assessment. USDA Forest Service, Forest Health Technology Enterprise Team. FHTET-14-01, 209 pp, [Available online at http://www.fs.fed.us/foresthealth/technology/pdfs/2012_RiskMap_Report_web.pdf.]

Kühnlein, M., T. Appelhans, B. Thies, and T. Nauss, 2014: Improving the accuracy of rainfall rates from optical satellite sensors with machine learning — A random forests-based approach applied to MSG SEVIRI. *Remote Sensing of Environment*, **141**, 129-143, doi: 10.1016/j.rse.2013.10.026.

Kupfer, J. A., A. T. Myers, S. E. McLane, and G. N. Melton, 2008: Patterns of forest damage in a southern Mississippi landscape caused by Hurricane Katrina. *Ecosystems*, **11**, 45-60.

Kwasinski, A., W. W. Weaver, P. L. Chapman, and P. T. Krein, 2009: Telecommunications power plant damage assessment for Hurricane Katrina–site survey and follow-up results. *IEEE Systems Journal*, **3**, 277-287.

Li, H., L. Treinish, and J. R. Hosking, 2010: A statistical model for risk management of electric outage forecasts. *IBM Journal of Research and Development*, **54**, 8: 1-8: 11.

Liaw, A., and M. Wiener, 2002: Classification and regression by randomForest. *R news*, **2**, 18-22.

Lin, Y., 2001: GCIP/EOP Surface: Precipitation NCEP/EMC 4KM Gridded Data (GRIB) Multi-Sensor Analysis (MUL) <-2001: dataset description. National Center for Atmospheric Research (NCAR). Accessed 1 Feb 2014. [Available online at http://data.eol.ucar.edu/codiac/dss/id=21.049.]

Lin, Y., 2014a: GCIP/EOP Surface: Precipitation NCEP/EMC 4KM Gridded Data (GRIB) Stage IV Data: dataset description. National Center for Atmospheric Research (NCAR). Accessed 1 Feb 2014. [Available online at http://data.eol.ucar.edu/codiac/dss/id=21.093.]

Lin, Y., 2014b: Q&A about NCEP Stage II/Stage IV. NOAA. Accessed 15 Feb 2014. [Available online at http://www.emc.ncep.noaa.gov/mmb/ylin/pcpanl/QandA/.]

Lin, Y., and K. E. Mitchell, 2005: The NCEP Stage II/IV Hourly Precipitation Analyses: Development and Applications. *19th Conference on Hydrology*, San Diego, CA, Amer. Meteor. Soc. [Available online at https://ams.confex.com/ams/Annual2005/techprogram/paper_83847.htm.]

Liu, H., R. A. Davidson, and T. Apanasovich, 2007: Statistical forecasting of electric power restoration times in hurricanes and ice storms. *IEEE Transactions on Power Systems*, **22**, 2270-2279.

Liu, H., R. A. Davidson, and T. V. Apanasovich, 2008: Spatial generalized linear mixed models of electric power outages due to hurricanes and ice storms. *Reliability Engineering & System Safety*, **93**, 897-912, doi: 10.1016/j.ress.2007.03.038.

Liu, H. B., R. A. Davidson, D. V. Rosowsky, and J. R. Stedinger, 2005: Negative Binomial Regression of Electric Power Outages in Hurricanes. *J Infrastruct Syst*, **11**, 258-267, doi: 10.1061/(Asce)1076-0342(2005)11:4(258).

Maliszewski, P. J., E. K. Larson, and C. Perrings, 2012: Environmental determinants of unscheduled residential outages in the electrical power distribution of Phoenix, Arizona. *Reliability Engineering & System Safety*, **99**, 161-171.

Merry, K., P. Bettinger, and J. Hepinstall, 2009: Physical and biological responses of forests to tropical cyclones affecting the United States Atlantic Ocean and Gulf of Mexico coasts. *American Journal of Environmental Sciences*, **5**, 16.

Nateghi, R., S. D. Guikema, and S. M. Quiring, 2011: Comparison and validation of statistical methods for predicting power outage durations in the event of hurricanes. *Risk analysis : an official publication of the Society for Risk Analysis*, **31**, 1897-1906, doi: 10.1111/j.1539-6924.2011.01618.x.

Nateghi, R., S. Guikema, and S. M. Quiring, 2014: Power outage estimation for tropical cyclones: improved accuracy with simpler models. *Risk analysis : an official publication of the Society for Risk Analysis*, **34**, 1069-1078, doi: 10.1111/risa.12131.

NWS, 2015: About the Hourly Precipitation Analysis Pages. National Weather Service (NWS) Southern Region Headquarters. Accessed 15 Mar 2015. [Available online at http://www.srh.noaa.gov/ridge2/RFC_Precip/.]

Ouyang, M., and L. Dueñas-Osorio, 2014: Multi-dimensional hurricane resilience assessment of electric power systems. *Structural Safety*, **48**, 15-24.

Pasch, R. J., 1997: Preliminary Report: Hurricane Danny, 16-26 July 1997. National Hurricane Center, 13 pp, [Available online at http://www.nhc.noaa.gov/data/tcr/AL051997_Danny.pdf.]

Quiring, S. M., L. Y. Zhu, and S. D. Guikema, 2011: Importance of soil and elevation characteristics for modeling hurricane-induced power outages. *Nat Hazards*, **58**, 365-390, doi: 10.1007/s11069-010-9672-9.

Quiring, S. M., A. B. Schumacher, and S. D. Guikema, 2014: Incorporating Hurricane Forecast Uncertainty into a Decision-Support Application for Power Outage Modeling. *B Am Meteorol Soc*, **95**, 47-58, doi: 10.1175/Bams-D-12-00012.1.

Reed, D. A., 2008: Electric utility distribution analysis for extreme winds. *Journal of Wind Engineering and Industrial Aerodynamics*, **96**, 123-140.

Reed, D. A., M. D. Powell, and J. M. Westerman, 2010: Energy infrastructure damage analysis for hurricane Rita. *Natural Hazards Review*, **11**, 102-109.

Ruefenacht, B., and Coauthors, 2008: USDA Forest Service FSGeodata Clearinghouse - Forest Types of the United States. Accessed 1 Feb 2014. [Available online at http://data.fs.usda.gov/geodata/rastergateway/forest_type/conus_forest_type_metadata.php.]

Shen, S., and D. Koval, 1999: Modelling extreme-weather-related transmission line outages. *1999 IEEE Canadian Conference on Electrical and Computer Engineering*, Edmonton, Alberta, Canada, Institute of Electrical and Electronics Engineers, 1271-1276, doi: 10.1109/CCECE.1999.804874.

Simpson, P., and R. Van Bossuyt, 1996: Tree-caused electric outages. *Journal of Arboriculture*, **22**, 117-121.

Stanturf, J. A., S. L. Goodrick, and K. W. Outcalt, 2007: Disturbance and coastal forests: a strategic approach to forest management in hurricane impact zones. *Forest Ecology and Management*, **250**, 119-135.

Stewart, B. A., 2014: The Failed State: Can it be Predicted? An Application of Random Forest for Variable Model Selection. M.S. Thesis, Dept. of Political Science, Utah State University, 138 pp, [Available online at http://digitalcommons.usu.edu/gradreports/383.]

Stewart, S. R., 2004: Tropical Cyclone Report: Hurricane Ivan, 2-24 September 2004. National Hurricane Center, 44 pp, [Available online at http://www.nhc.noaa.gov/data/tcr/AL092004_Ivan.pdf.]

Tilt, K., J. Eakes, D. West, J. Miles, and W. East, 2006: Restoring Our Hurricane-Ravaged Urban Tree Canopy: Best Management Practices for Site Evaluation, Tree Selection, Establishment, and Maintenance. Alabama Cooperative Extension System. ANR-1295, 20 pp, [Available online at http://www.aces.edu/pubs/docs/A/ANR-1295/ANR-1295.pdf.]

Tonn, G., S. Guikema, C. Ferreira, and S. Quiring, 2014: A Longitudinal Analysis of the Drivers of Power Outages During Hurricanes: A Case Study with Hurricane Isaac. *Probabilistic Safety Assessment and Management*, Honolulu, Hawaii, International Association for Probabilistic Safety Assessment and Management. [Available online at http://psam12.org/proceedings/paper/paper_532_1.pdf.]

Touw, W. G., J. R. Bayjanov, L. Overmars, L. Backus, J. Boekhorst, M. Wels, and S. A. van Hijum, 2012: Data mining in the Life Sciences with Random Forest: a walk in the park or lost in the jungle? *Briefings in Bioinformatics*, **14**, 315–326, doi: 10.1093/bib/bbs034.

Ware, E., 2005: Corrections to radar-estimated precipitation using observed rain gauge data. M.S. Thesis, Dept. of Earth and Atmospheric Sciences, Cornell University, 96 pp, [Available online at http://hdl.handle.net/1813/2115.]

Willmott, C. J., and K. Matsuura, 2005: Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, **30**, 79.

Winkler, J., L. Duenas-Osorio, R. Stein, and D. Subramanian, 2010: Performance assessment of topologically diverse power systems subjected to hurricane events. *Reliability Engineering & System Safety*, **95**, 323-336.

Xi, W., and R. K. Peet, 2008: Hurricane effects on the Piedmont forests: Patterns and implications. *Ecological Restoration*, **26**, 295-298.

Zhou, Y., A. Pahwa, and S.-S. Yang, 2006: Modeling weather-related failures of overhead distribution lines. *IEEE Transactions on Power Systems*, **21**, 1683-1690.

Zhu, D., D. Cheng, R. P. Broadwater, and C. Scirbona, 2007: Storm modeling for prediction of power distribution system outages. *Electric power systems research*, **77**, 973-979.

Zhu, L., and S. M. Quiring, 2013: Variations in tropical cyclone precipitation in Texas (1950 to 2009). *Journal of Geophysical Research: Atmospheres*, **118**, 3085-3096, doi: 10.1029/2012JD018554.

Zhu, L., O. W. Frauenfeld, and S. M. Quiring, 2013: Seasonal tropical cyclone precipitation in Texas: A statistical modeling approach based on a 60 year climatology. *Journal of Geophysical Research: Atmospheres*, **118**, 8842-8856, doi: 10.1002/jgrd.50663.