ESTIMATING HURRICANE OUTAGE AND DAMAGE RISK

IN POWER DISTRIBUTION SYSTEMS

A Dissertation

by

SEUNG RYONG HAN

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

DOCTOR OF PHILOSOPHY

August 2008

Major Subject: Civil Engineering

ESTIMATING HURRICANE OUTAGE AND DAMAGE RISK

IN POWER DISTRIBUTION SYSTEMS

A Dissertation

by

SEUNG RYONG HAN

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

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Seth Guikema
Committee Members,	David Rosowsky
	Jose Roesset
	Steven M. Quiring
Head of Department,	David Rosowsky

August 2008

Major Subject: Civil Engineering

ABSTRACT

Estimating Hurricane Outage and Damage Risk in Power Distribution Systems.

(August 2008)

Seung Ryong Han, B.S., KunKuk University at Seoul;

M.S., Korea University at Seoul

Chair of Advisory Committee: Dr. Seth Guikema

Hurricanes have caused severe damage to the electric power system throughout the Gulf coast region of the U.S., and electric power is critical to post-hurricane disaster response as well as to long-term recovery for impacted areas. Managing hurricane risks and properly preparing for post-storm recovery efforts requires rigorous methods for estimating the number and location of power outages, customers without power, and damage to power distribution systems. This dissertation presents a statistical power outage prediction model, a statistical model for predicting the number of customers without power, statistical damage estimation models, and a physical damage estimation model for the gulf coast region of the U.S. The statistical models use negative binomial generalized additive regression models as well as negative binomial generalized linear regression models for estimating the number of power outages, customers without power, damaged poles and damaged transformers in each area of a utility company's service area. The statistical models developed based on transformed data replace hurricane indicator variables, dummy variables, with physically measurable variables, enabling future predictions to be based on only well-understood characteristics of hurricanes. The physical damage estimation model provides reliable predictions of the number of damaged poles for future hurricanes by integrating fragility curves based on structural reliability analysis with observed data through a Bayesian approach. The models were developed using data about power outages during nine hurricanes in three states served by a large, investor-owned utility company in the Gulf Coast region.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Guikema, and my committee members, Dr. Rosowsky, Dr. Roesset, and Dr. Quiring, for their guidance and support throughout the course of this research.

Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience. This study was partially funded by a private utility company that wishes to remain anonymous. This utility also provided the data used in the analysis. I gratefully acknowledge their support.

Finally, thanks to my mother for her encouragement.

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

Page

Page

LIST OF TABLES

1. INTRODUCTION

In recent years, hurricanes have caused severe power interruption throughout the Gulf Coast region of the U.S. For example, the central Gulf Coast region (Louisiana, Alabama, Mississippi, Florida and Georgia) has been significantly impacted recently by Hurricanes Danny (1997), Georges (1998), Hanna (2002), Isidore (2002), Frances (2004), Ivan (2004), Jeanne (2004), Cindy (2005), Dennis (2005), and Katrina (2005). In addition to causing considerable direct repair and restoration costs for utility companies, hurricane-related power outages and damage to power distribution systems may result in loss of services from a number of other critical infrastructure systems leading, in turn, to significant delays in post-storm recovery for the impacted region.

Liu et al. (2005) developed the first rigorous statistical model for estimating power outage risk during hurricanes. They developed a generalized linear regression model for estimating the spatial distribution of power outages during hurricanes using power outage data from past hurricanes in the Carolinas. However, Liu et al. (2005) relied on the use of hurricane indicator variables. These are binary variables, one per hurricane, that indicate which hurricane a given outage was from. Without including these variables in the model, the models of Liu et al. (2005) did not fit the past outage data as well. These types of models can be used to predict the spatial distribution of power outages from a hurricane that is threatening a utility company's service area. However, one must make assumptions about how to include the binary hurricane variables. For example, one could assume that the approaching hurricane is equally likely to be like each of the past hurricanes and thus average the effects of the indicator variables. However, because the hurricane indicator variables are not tied to measurable characteristics of hurricanes, it is difficult to know what aspects of hurricanes they are capturing. System managers may place more confidence in a model based on measurable characteristics of hurricanes, and such a model would help to improve the understanding of the impacts of hurricanes on electric power distribution systems.

This dissertation follows the style of *Journal of Structural Engineering*.

Past research such as Liu et al. (2005) also focused on modeling only power outages, where a power outage is defined as the activation of a protective device. A single outage could affect few customers or it could affect hundreds of customers. However, the number of customers without power is more aligned with the methods utility companies use for pre-hurricane deployment of repair crews and materials. Also, it would be helpful to have direct estimates of the amount of actual damage (e.g., broken poles and transformers) to power distribution systems during hurricanes. Accurate and reliable customer outage predictions and damage predictions can help utility companies better manage the effects of hurricanes by providing estimates of the number of customers without power at a spatially detailed level and the amount of damage to poles and transformers in the distribution system at a spatially detailed level rather than estimates of only the number of power outages. This thesis develops, tests, and demonstrates models for estimating the spatial distribution of not only electric power outages but also the number of customers without power and the amount of damage during hurricanes using only measurable characteristics of hurricanes, the power system, local geography, and local climate.

One other researcher took a different, i.e., non-regression, approach to estimating risk to power systems during hurricanes. The Caribbean Disaster Mitigation Project (1996) developed structural reliability models to estimate damage to power distribution system poles. The Caribbean Disaster Mitigation Project (1996) included hurricane simulation modeling together with a structural analysis of the poles in the power distribution system to account for the effects of hurricane-related wind. However, this study considered only flexural damage to poles under wind loads in their structural reliability model, not foundation failure. In this thesis, fragility curves for power distribution system poles considering foundation failure are developed. In addition, this thesis combined the information provided by structural reliability methods with the information contained in actual failure data through a Bayesian approach.

This study developed statistical models for predicting the number of power outages, customers without power, damaged poles and damaged transformers for 3.66

km (12,000 foot) by 2.44 km (8,000 foot) grid cells covering a company's service area for an approaching hurricane while relying only on information that is measurable prior to the hurricane making landfall. These models were based on information about the hurricane, the power system, and the local climatology and geography. The data was supplied by a large, investor-owned utility company serving the Gulf Coast region. I used generalized linear models (GLMs) and generalized additive models (GAMs), a type of model appropriate for regression analysis of count data. However, GLMs and GAMs are based on the assumption that the explanatory variables are statistically independent of each other. Regression modeling based on highly correlated input data (i.e., collinear data) can lead to poor estimation of regression parameters, and the input data analyzed in this study are highly correlated. To avoid the collinearity problem, the data was transformed through principal components analysis (PCA) as will be discussed in detail below. The resulting models provide predictions of the number of outages, customers without power, damaged poles, and damaged transformers that can help a utility company better manage the effects of hurricanes by pre-positioning and deploying repair personnel and materials prior to a hurricane making landfall.

2. BACKGROUND[†]

2.1 Generalized Linear Models

A standard model for count data such as power outages is the Poisson generalized linear regression model. Let the vector of the *n* explanatory variables for grid cell *i* (*i* = 1,…,*m*) be given by $\vec{x}_i^{'} = [x_{1i},...,x_{ni}]$ and the number of power outages in grid cell *i* be given by y_i. A regression model based on the Poisson distribution for the counts conditional on the observed values of the explanatory variables specifies that the conditional mean of the counts is given by a continuous function, $\mu(\vec{\beta},\vec{x}_i)$, of the covariate values as specified in equation (2.1), where $\vec{\beta}$ is the *n* x 1 vector of regression parameters (e.g., Cameron and Trivedi 1998).

$$E[y_i \mid \vec{x}_i] = \mu(\vec{\beta},\vec{x}_i) \tag{2.1}$$

Conditional on $\vec{x}_i^{'}$, the probability density function assumed for y_i in a Poisson regression model is given, for non-negative integers y_i, by:

$$f(y_i \mid \vec{x}_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} \tag{2.2}$$

We use the standard log link function to specify the conditional mean. That is, we assume that $E[y_i \mid \vec{x}_i] = \exp(\vec{x}_i^{'}\vec{\beta})$. This model is called a Poisson Generalized Linear Model (GLM) because it generalizes standard multivariate linear regression to incorporate a different conditional likelihood function for Poisson-distributed count data. It is a convenient and widely used model, but it is based on the assumption that the conditional mean and the conditional variance, given by ω_i, of the count data are equal:

$$\mu_i = \omega_i = \exp(\vec{x}_i^{'}\vec{\beta}) \tag{2.3}$$

This strong assumption of a conditional variance equal to the conditional mean is not a valid assumption for some count data sets, including the outage data used in this study.

[†] This material is adapted from Han et al. (2008a, 2008b, 2008c) where this material is presented in a similar form.

In many cases, the data is overdispersed relative to the Poisson model, meaning that the conditional variance is greater than the conditional mean.

One method for modeling overdispersed data is to use a negative binomial GLM. With a negative binomial GLM, the count data are assumed to follow a negative binomial probability density function conditional on $\vec{x}_i^{'}$ and α, the overdispersion parameter, as shown in equation (2.4)

$$f(y_i \mid \vec{x}_i, \alpha) = \frac{\Gamma(y_i + \alpha)}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu} \right)^{\alpha^{-1}} \left(\frac{\mu}{\alpha^{-1} + \mu} \right)^{y_i} \quad (2.4)$$

where $\mu_i = \exp(\vec{x}_i^{'} \vec{\beta})$ as for the Poisson GLM (Cameron and Trivedi 1998). The variance of the count data under a negative binomial model is $\omega_i = \mu_i + \alpha\mu_i^2$ (e.g., Cameron and Trivedi 1998). This model can be derived in a number of ways, one of which is by starting with a Poisson GLM and adding a gamma-distributed random term with mean 1 and variance α to the link function (Cameron and Trivedi 1998). This type of model was used in estimating power outages from hurricanes in the southeastern U.S. by Liu et al. (2005). Liu et al. (2008) extended this approach by using a Generalized Linear Mixed Model (GLMM) to examine the importance of spatial correlation in statistical power outage estimation models. Because Liu et al. (2008) showed that including spatial correlation through the GLMM framework did not significantly improve model fit, I used the simpler GLM modeling framework in this study.

2.2 *Generalized Additive Models*

As with a GLM, a GAM is composed of a random component, an additive component, and a link function. A GAM is different from a GLM in that an additive predictor replaces the linear predictor. That is, the linear form $\alpha + \sum_j x_j \beta_j$ is replaced with the additive form $\alpha + \sum_j f_j(x_j)$ where $f_i(x_i)$ is a function that smoothes the jth component of **X**. More specifically, a GAM generally assumes that the response Y has a distribution with the mean $\mu = E[Y | X_1, \cdots\cdots, X_p]$ linked to the predictor via a link function

$$g(\mu) = \alpha + \sum_{j=1}^{p} f_j(X_j) \tag{2.5}$$

where each f_j is a smoothing function of a specified class of functions estimated non-parametrically (Hastie and Tibshirani 1990). While the nonparametric form of f_j makes the model more flexible, the additivity is retained and allows one to fit the model in much the same way as GLMs. This approach allows the form of the relationship between the explanatory variables and the measure of interest, here power outages during hurricanes, to be estimated directly from the data.

2.3 Model Fitting and Measuring Goodness of Fit

I used three different methods to compare fitted models for a data set. The first is the deviance of the fitted models, defined as (Cameron and Trivedi 1998):

$$deviance = -2\left(\log L_{\max} - \log L_{fitted}\right) \tag{2.6}$$

where $\log L_{\max}$ is the maximum log-likelihood achievable and $\log L_{\text{fitted}}$ is the log-likelihood of the fitted model. In comparing models, a lower deviance is preferred. A formal hypothesis test for comparing two models can also be defined based on the deviances of the models. A likelihood ratio test is a formal hypothesis test using the difference in deviance between two nested models. This difference in deviance is approximately χ^2 distributed with the degrees of freedom equal to the number of parameters by which the models differ (e.g., Cameron and Trivedi 1998, Agresti 2002). While this provides a formal comparison of the models, it is only valid when the set of covariates, also referred to as explanatory variables, used in one model is a subset of the covariates included in the other model.

The second and third methods used for comparing different models are based on pseudo-R^2, measures of the fit of a GLM that are meant to provide similar insights as R^2 does in linear regression. There are different definitions of pseudo-R^2, depending on what one wishes to measure. One common psedo-R^2 is R^2_{dev}, a deviance-based pseudo-R^2. R^2_{dev} is defined as (Cameron and Trivedi 1998):

$$R^2_{dev} = 1 - \frac{D(y,\hat{\mu})}{D(y,\bar{y})} \tag{2.7}$$

where $D(y,\hat{\mu})$ is the deviance of the fitted model and $D(y,\bar{y})$ is the deviance of the intercept-only model. This pseudo-R^2 thus measures the reduction in deviance achieved by including regression parameters. An alternate pseudo-R^2 can be defined based on α, the overdispersion parameter of the model (e.g. Liu et al. 2005). This pseudo-R^2, defined in equation (2.8), measures the reduction in variability above the Poisson model (i.e., the amount of variability not due to Poisson variability about the mean) due to the inclusion of regression parameters.

$$R^2_{dev} = 1 - \frac{\alpha}{\alpha_0} \tag{2.8}$$

In equation (2.8), α is the overdispersion parameter for the fitted model and α_0 is the overdispersion parameter for the intercept-only model.

2.4 Principal Components Analysis

One of the problems often encountered when fitting regression models to data is that the covariates may be correlated, violating one of the assumptions underlying regression modeling. High degrees of correlation lead to unstable estimates of regression parameters with standard regression approaches. This means that the parameter estimates are highly sensitive to the particular set of data used to fit the model, leading to potential problems with the predictive ability of the fitted model. There are two main approaches for overcoming this difficulty, changing the model used or transforming the data to remove correlation problems. In this study I used a data transformation method called Principal Components Analysis (PCA).

A Principal Component Analysis (PCA) is a mathematical procedure that transforms the data set to a new orthogonal coordinate system such that the transformed data are mutually orthogonal. This means that the transformed data are not correlated.

The transformation can be done by decomposing the data matrix, $\bar{x}$, into its eigenvalues and eigenvectors. The eigenvalues are a measure of the variance of each of the elements of $\bar{x}$, and the eigenvectors are used to transform the data into orthogonal vectors. The results of a PCA are a vector of the eigenvalues, a matrix of the eigenvectors, and a matrix of the transformed data. The transformed data can then be used for fitting regression models.

The PCA was done in the program R using the "prcomp" command which is done by a singular value decomposition of the standardized data to obtain principal components for the covariance matrix. The commands history and the results of the PCA are given in Appendix A where the eigenvectors are referred to as loadings. These loadings would are used to transform data into the principal components by taking a weighed linear combination of the original data, where the weights are given by the eigenvectors. This approach allowed me to overcome the problem of high degrees of correlation in the input data for my models.

3. DATA DESCRIPTION[‡]

The models developed in this thesis are based on data provided by a large, investor-owned utility company in the Gulf Coast region. This company serves much of the central Gulf Coast region, and the statistical models in this thesis are based on covering this service area with 3.66 km (12,000 foot) by 2.44 km (8,000 foot) grid cells. I have data from the utility's service area in three Gulf Coast states, which I will refer to as States A, B, and C in order to protect the identity of the data provider. There are 6,681 grid cells for State A, 602 grid cells for State B, and 7,330 grid cells for State C. I used data on outages during 5 hurricanes (Danny, Dennis, Georges, Ivan, and Katrina) in State A, during 3 hurricanes in State B (Dennis, Ivan, and Katrina), and during 8 hurricanes in State C (Cindy, Dennis, Frances, Hanna, Isidore, Ivan, Jeanne, and Katrina).

3.1 Hurricane Characteristic Data

In order to capture the characteristics of the wind field during a given hurricane, I used estimates of the maximum 3-second gust wind speed and the length of time that the winds were above 20 m/s (44.7 miles per hour) for each grid cell based on the hurricane wind field model developed by Huang et al. (2001), the same model that was used in an earlier study of power outages during hurricanes in North and South Carolina (Liu et al. 2005). In this hurricane wind field model, reconnaissance flight data is used to develop a gradient-level wind estimate model based on Georgiou's wind field model (Georgiou 1985) and the hurricane decay model of Vickery and Twisdale (1995). This model produces an estimate of the gradient-level wind speed throughout the duration of a hurricane at the center of each grid cell. This estimated wind speed was then converted to a "surface wind speed", the wind speed estimated at a height of 10 m in an assumed open exposure location, by using a multiplicative gradient-to-surface conversion factor.

[‡] The data used in this thesis is the same as that used in Han et al. (2008a, 2008b, 2008c). The description of the data given in this section is adapted from a combination of the data description sections of these three papers.

The gradient-to-surface conversion factor was taken to be 0.72 for sites more than 10 km from the coast, 0.80 for sites within 10 km from the coast, and 0.90 for sites adjacent to the sea as suggested by Rosowsky et al. (1999). I did not attempt to use different conversion factors based on records of local land cover types. I also did not correct for local topography effects because I did not have enough detailed information to include this in the model. Figure 3.1 shows the surface wind speeds on two sites as an example of comparison between estimated wind speeds by using the hurricane wind field model and measured wind speeds. The wind speeds of the left plot represent the wind on the site located right on the track of hurricanes, showing a vortex shape of hurricanes. The wind speeds of the right plot shows typical pattern of wind speeds during hurricanes, indicating when the hurricane made landfall.

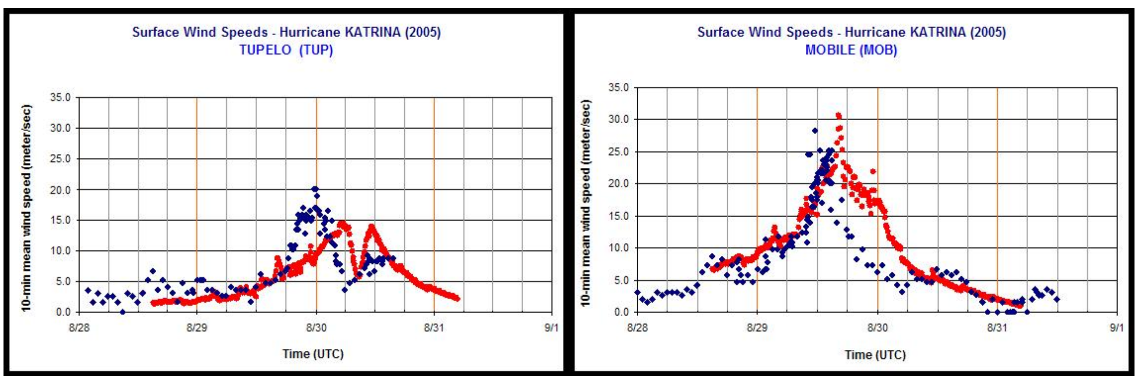

Figure 3.1. Surface wind speed comparison in State A for Hurricane Katrina.

Based on the results of Liu et al. (2005), I initially included hurricane indicator variables in my statistical models. These variables are binary variables in the regression model signifying which hurricane a given outage is from, and these variables may capture additional features of the hurricane not captured with the wind speed variables. However, as discussed above, it would be preferable to be able to use measurable characteristics of hurricanes rather than binary hurricane indicator variables. One of the main advances in the model presented in this section is that it uses input variables that are measurable prior to a hurricane making landfall rather than hurricane indicator variables while still providing fits to the outage data that equal or exceed those of a model that includes hurricane indicator variables.

3.2 Fractional Soil Moisture Anomalies[§]

I included additional variables that help to explain the variability of outages across a service area and between different hurricanes. One of these variables dealt with soil moisture levels. Soil moisture is thought to impact the stability of poles and trees, with highly saturated soil potentially increasing both the likelihood of poles being blown over and the likelihood of trees being blown onto poles and power lines during hurricanes. To account for this, I calculated fractional soil moisture anomalies at the time of hurricane landfall to represent the degree of soil saturation at different depths in the soil and included this information in the statistical model.

Soil moisture was simulated for each of the grid cells using the Variable Infiltration Capacity (VIC) model. The VIC model is a semi-distributed hydrological model that is capable of representing subgrid-scale variations in vegetation, available water holding capacity, and infiltration capacity (Liang et al. 1994, 1996a, 1996b). The influence of variations in soil properties, topography, and vegetation within each grid cell are accounted for statistically by using a spatially varying infiltration capacity. VIC utilizes a soil-vegetation-atmosphere transfer scheme that accounts for the influence of vegetation and soil moisture on land-atmosphere moisture and energy fluxes and these fluxes are balanced over each grid cell (Andreadis et al. 2005). The model has been utilized in basin-scale hydrological modeling (Abdulla et al. 1996, Cherkauer and Lettenmaier 1999, Nijssen et al. 1997, and Wood et al. 1997), continental-scale simulations associated with the North American Land Data Assimilation System (NLDAS) (Maurer et al. 2002), and global-scale applications Nijissen et al. (2001). A thorough evaluation of VIC was undertaken as part of NLDAS and the results indicated that soil moisture is generally well simulated by the VIC model (Robock et al. 2003).

[§] The soil moisture data used in this study was provided by Dr. Steven Quiring and his students from the Department of Geography. Creating this input to the statistical model was not part of the author's Ph.D. research.

These findings are supported by a recent soil moisture model evaluation which demonstrated that the VIC model accurately simulated the wetting and drying of the soil (Meng and Quiring 2007).

The VIC model was forced using station-based measurements of daily maximum and minimum temperature and precipitation. Daily 10 m wind speeds from NCEP/NCAR reanalysis were also used. Additional meteorological and radiative forcings such as vapor pressure, shortwave radiation, and net longwave radiation were derived using established relationships with maximum and minimum temperatures, daily temperature range, and precipitation. Soil characteristics were extracted from the Natural Resource Conservation Service's State Soil Geographic Database (STATSGO). Land cover and vegetation parameters were derived using the global vegetation classification developed by Hansen et al. (2000).

Soil moisture was simulated by VIC in three layers. In this study, the depth of the first soil layer is 10 cm, the depth of the second soil layer varied from 30 to 50 cm and the third soil layer varied from 40 to 60 cm. Total soil depth (sum of the three layers) was 1 m at all grid cells. Modeled soil moisture data were initially reported as a depth (mm) and then were converted to a percentage of total capacity (fractional soil moisture) for each layer. One advantage of expressing soil moisture as a fraction of total capacity is that it controls for spatial differences in layer depth, bulk density, particle density, and soil porosity, and allows soil moisture from different locations to be directly compared. VIC was run at 1/2 degree (latitude/longitude) resolution and then downscaled to the resolution of the utility company grid (12,000 ft by 8,000 ft) using an Inverse-Distance Weighting (IDW) algorithm (radius of influence = 100 km). For each hurricane, fractional soil moisture was calculated for the 7 days before landfall.

3.3 Precipitation

Long-term precipitation is one of the drivers in the distribution of plant communities over an area, and some types of plant communities may pose higher risks to power distribution lines during hurricanes. For example, some types of trees such as pines may be more susceptible to being blown onto power lines during a hurricane than others, potentially increasing the risk of power outages. Unfortunately, geographically detailed data about the distribution of plant communities is not available for the three states under consideration. To help account for this source of spatial heterogeneity in outage risk associated with precipitation, I included two measures of long term precipitation – mean annual precipitation and a Standardized Precipitation Index.

Mean annual precipitation (mm) was calculated for each of the grid cells based on daily precipitation data from 1915–2004. Daily precipitation data was acquired from the National Oceanic and Atmospheric Administration (NOAA) Cooperative Observer (COOP) network. Mean annual precipitation was calculated at each 1/2 degree grid cell and then downscaled to the utility company grid using an Inverse-Distance Weighting (IDW) algorithm (radius of influence = 100 km). Mean annual precipitation is thought to be related to the types of plants that would tend to grow in a given area.

The Standardized Precipitation Index (SPI) provides a simple and versatile method for quantifying antecedent precipitation (McKee et al. 1993 and 1995). The SPI is a statistical measure of the deviation of precipitation from normal levels and it can be calculated for any time period of interest. The SPI is spatially invariant, meaning that the definition of SPI does not depend on spatial location, (Guttman 1998, Heim 2002, and Wu et al. 2007) and so values of the SPI can readily be compared across time and space. The SPI is influenced by the normalization procedure (e.g., a probability density function) that is used. The National Drought Mitigation Center (NDMC), Western Regional Climate Center (WRCC), and National Agricultural Decision Support System (NADSS) all use the two-parameter gamma probability density function (PDF) to calculate SPI. However, there is little consensus about what normalization procedure is best. Guttman (1999) analyzed six different PDFs and determined that the Pearson Type III was the most appropriate PDF for calculating SPI. Therefore, this PDF was used to generate the SPI values for this study.

SPI was calculated using monthly precipitation data (1915-2005) at each of the 1/2 degree (latitude/longitude) grid cells described in the previous section. The SPI was calculated for six different time periods (1, 2, 3, 6, 12, and 24-months). This provides a means to account for antecedent moisture conditions for a variety of pre-storm time frames, each based on deviations from long-term precipitation patterns. The SPI data was downscaled to the utility company grid using an Inverse-Distance Weighting (IDW) algorithm (radius of influence = 100 km). The SPI data was only utilized for the months during which hurricanes occurred.

3.4 Land Cover

I also used information about land cover and land use in out statistical outage models in order to try to capture differences in outage rates for different land uses. For example, commercial areas may have different outage rates than rural areas, even given equal values for the other explanatory variables. The land cover data I used is publicly available in the National Land Cover Database (NLCD) 2001 (NLCD 2001), which is available from the United States Geological Survey (USGS) Seamless website (http://seamless.usgs.gov/). The NLCD 2001 provides data with a resolution of 1 arc-second (approximately 30 m) for each of 21 land cover classes. I categorized the 21 land cover classes into 8 aggregated classes according to starting numbers of the original 21 classes. This yielded 8 coherent land covers types: water, developed (including residential, commercial, and industrial), barren, forest, scrub, grass, pasture, and wetland. Land cover and land use were obtained by using the program "ArcView". One hundred points were generated in each grid cell and then matched with the land cover data available in USGS with ArcView using the "Join" command. Finally, I got land cover and land use percentage in each grid cell.

3.5 Power System Data

In addition to information discussed above, I included information about the power system obtained from the utility companies. This includes the number of transformers, poles, switches, customers, and the miles of overhead in each grid cell. In addition, I was provided the miles of underground line in each grid cell for the State A and the number of poles in each grid cell for the States A and C.

3.6 Summary of Data

The explanatory variables used in my statistical model are as follows:

- $y_{i,Outages}$: Number of outages in grid cell i

(State A : mean = 0.92, standard deviation = 3.60, minimum = 0, maximum = 156

State B : mean = 12.78, standard deviation = 32.66, minimum = 0, maximum = 461

State C : mean = 0.13, standard deviation = 0.85, minimum = 0, maximum = 32)

- $y_{i,Customers}$: Number of customers without power in grid cell i

(State A : mean = 92.5, standard deviation = 537.69, minimum = 0, maximum = 21,321

State B : mean = 980, standard deviation = 3505.35, minimum = 0, maximum = 40,725

State C : mean = 0.54, standard deviation = 16.68, minimum = 0, maximum = 2,133)

- $x_{i,t}$: Number of transformers in grid cell i

(State A : mean = 87.63, standard deviation = 145.69, minimum = 0, maximum = 1,525

State B : mean = 197.6, standard deviation = 271.87, minimum = 0, maximum = 1,428

State C : mean = 82.61, standard deviation = 175.94, minimum = 0, maximum = 1,440)

- $x_{i,p}$: Number of poles in grid cell i

(State A : mean = 234.9, standard deviation = 373.62, minimum = 1, maximum = 4,311

State C : mean = 170.8, standard deviation = 327.16, minimum = 0, maximum = 3,852)

- $x_{i,o}$: Length of overhead line in grid cell i (in miles)

(State A : mean = 8.58, standard deviation = 8.89, minimum = 0, maximum = 98.88

State B : mean = 12.69, standard deviation = 14.76, minimum = 0.12, maximum = 86.14

State C : mean = 20.52, standard deviation = 31.57, minimum = 0, maximum = 231.23)

- $x_{i,u}$: Length of underground line in grid cell i (in miles)

(State A : mean = 0.82, standard deviation = 3.67, minimum = 0, maximum = 58.29)

- $x_{i,s}$: Number of switches in grid cell i

(State A : mean = 13.16, standard deviation = 28.42, minimum = 0, maximum = 447

State B : mean = 45.07, standard deviation = 67.19, minimum = 0, maximum = 438

State C : mean = 17.22, standard deviation = 39.12, minimum = 0, maximum = 482)

- $x_{i,c}$: Number of customers in grid cell i

(State A : mean = 181.9, standard deviation = 559.62, minimum = 0, maximum = 9,659

State B : mean = 588.3, standard deviation = 1,026.42, minimum = 0, maximum = 6,253

State C : mean = 283.3, standard deviation = 869.96, minimum = 0, maximum = 15,281)

- $x_{i,m}$: Maximum 3-second gust wind speed in m/s

(State A : mean = 21.52, standard deviation = 12.28, minimum = 5.04, maximum = 52.56

State B : mean = 35.41, standard deviation = 9.63, minimum = 17.14, maximum = 57.51

State C : mean = 15.85, standard deviation = 6.88, minimum = 6.48, maximum = 50.85)

- $x_{i,d}$: Duration of strong winds (length of time the wind speed was above 20 m/s) in minutes

(State A : mean = 8.78, standard deviation = 8.83, minimum = 0, maximum = 41.83

State B : mean = 15.8, standard deviation = 7.63, minimum = 0, maximum = 29.83

State C : mean = 2.53, standard deviation = 5.41, minimum = 0, maximum = 26)

- $x_{i,Cindy}$, $x_{i,Dennis}$, $x_{i,Frances}$, $x_{i,Hanna}$, $x_{i,Isidore}$, x_{iIvan}, $x_{i,Jeanne}$: Hurricane indicator variables that equal one if the outages occurred during the given hurricane and zero otherwise. Note that for outages occurring during Hurricane Katrina, all of the hurricane indicator variables are zero.
- $x_{i\ Time}$: Time since the last hurricane landfall in months

(State A : mean = 23.6, standard deviation = 25.05, minimum = 1, maximum = 72

State B : mean = 27.67, standard deviation = 31.57, minimum = 1, maximum = 72

State C : mean = 10.38, standard deviation = 16.28, minimum = 0, maximum = 48)

- $x_{i\ Pressure}$: Central pressure deficit (ΔP) in mb where ΔP = 1013 – Pc with Pc being the central pressure when the hurricane makes landfall

(State A : mean = 60, standard deviation = 22.82, minimum = 24, maximum = 93

State B : mean = 75.67, standard deviation = 12.26, minimum = 67, maximum = 93

State C : mean = 50.5, standard deviation = 26.05, minimum = 10, maximum = 93)

- $x_{i\ RMW}$: Radius of maximum winds in km

(State A : mean = 37.25, standard deviation = 4.98, minimum = 28.59, maximum = 43.11

State B : mean = 34.03, standard deviation = 3.85, minimum = 28.59, maximum = 36.9

State C : mean = 37.51, standard deviation = 4.94, minimum = 28.59, maximum = 43.82)

- $x_{i\ FSM1}$: Fractional soil moisture anomalies at a depth of 0 cm to 10 cm

(State A : mean = 0.12, standard deviation = 0.04, minimum = -0.11, maximum = 0.13

State B : mean = 0.04, standard deviation = 0.02, minimum = -0.02, maximum = 0.1

State C : mean = 0.02, standard deviation = 0.04, minimum = -0.15, maximum = 0.16)

- $x_{i\ FSM2}$: Fractional soil moisture anomalies at a depth of 10 cm to 40 cm

(State A : mean = 0.01, standard deviation = 0.05, minimum = -0.18, maximum = 0.13

State B : mean = 0.02, standard deviation = 0.03, minimum = -0.05, maximum = 0.09

State C : mean = 0.03, standard deviation = 0.05, minimum = -0.16, maximum = 0.15)

- $x_{i\ FSM3}$: Fractional soil moisture anomalies at a depth of 40 cm to 140 cm

(State A : mean = 0.01, standard deviation = 0.05, minimum = -0.12, maximum = 0.31

State B : mean = 0.76, standard deviation = 0.03, minimum = -0.06, maximum = 0.08

State C : mean = 0.05, standard deviation = 0.09, minimum = -0.16, maximum = 0.4)

- $x_{i\ MAP}$: Mean annual precipitation in mm

(State A : mean = 1,436, standard deviation = 63.82, minimum = 1,300, maximum = 1,648

State B : mean = 1,601, standard deviation = 63.76, minimum = 1,424, maximum = 1,666

State C : mean = 1,296, standard deviation = 94.73, minimum = 1,151, maximum = 1,686)

- $x_{i\ SPI1}$: Standardized Precipitation Index (1 month)

(State A : mean = 0.61, standard deviation = 0.87, minimum = -1.9, maximum = 2.94

State B : mean = 0.66, standard deviation = 0.27, minimum = 0.03, maximum = 1.31

State C : mean = 1.28, standard deviation = 0.7, minimum = -0.76, maximum = 3.07)

- $x_{i\ SPI2}$: Standardized Precipitation Index (2 months)

(State A : mean = 0.92, standard deviation = 0.74, minimum = -2.01, maximum = 2.7

State B : mean = 0.77, standard deviation = 0.29, minimum = 0.12, maximum = 1.24

State C : mean = 1.23, standard deviation = 0.68, minimum = -1.08, maximum = 3.07)

- $x_{i\ SPI3}$: Standardized Precipitation Index (3 months)

(State A : mean = 0.88, standard deviation = 0.66, minimum = -1.62, maximum = 2.33

State B : mean = 0.76, standard deviation = 0.34, minimum = -0.1, maximum = 1.3

State C : mean = 0.91, standard deviation = 0.65, minimum = -1.67, maximum = 2.69)

- $x_{i\ SPI6}$: Standardized Precipitation Index (6 months)

(State A : mean = 0.64, standard deviation = 0.53, minimum = -0.88, maximum = 2.02

State B : mean = 1.19, standard deviation = 0.56, minimum = 0.15, maximum = 1.93

State C : mean = 0.62, standard deviation = 0.75, minimum = -1.42, maximum = 2.18)

- $x_{i\ SPI12}$: Standardized Precipitation Index (12 months)

(State A : mean = 0.64, standard deviation = 0.64, minimum = -1.09, maximum = 1.93

State B : mean = 1.08, standard deviation = 0.81, minimum = -0.22, maximum = 2.25

State C : mean = 0.06, standard deviation = 0.97, minimum = -1.83, maximum = 2.14)

- $x_{i\ SPI24}$: Standardized Precipitation Index (24 months)

(State A : mean = 0.58, standard deviation = 0.31, minimum = -0.45, maximum = 1.48

State B : mean = 0.87, standard deviation = 0.14, minimum = 0.41, maximum = 1.18

State C : mean = 0.18, standard deviation = 0.71, minimum = -2.13, maximum = 1.51)

- $x_{i\ LC1}$: Percentage of land cover in grid cell that is water

(State A : mean = 2.35, standard deviation = 7.56, minimum = 0, maximum = 100

State B : mean = 15.1, standard deviation = 27.35, minimum = 0, maximum = 100

State C : mean = 1.78, standard deviation = 6.55, minimum = 0, maximum = 94)

- $x_{i\ LC2}$: Percentage of land cover in grid cell that is developed (residential, commercial, and industrial combined)

(State A : mean = 8.58, standard deviation = 13.72, minimum = 0, maximum = 100

State B : mean = 20.3, standard deviation = 22.37, minimum = 0, maximum = 100

State C : mean = 13.65, standard deviation = 18.02, minimum = 0, maximum = 100)

- $x_{i\ LC3}$: Percentage of land cover in grid cell that is barren

(State A : mean = 0.46, standard deviation = 1.54, minimum = 0, maximum = 27

State B : mean = 1.48, standard deviation = 3.1, minimum = 0, maximum = 31

State C : mean = 0.58, standard deviation = 1.57, minimum = 0, maximum = 32)

- $x_{i\ LC4}$: Percentage of land cover in grid cell that is forest

(State A : mean = 51.48, standard deviation = 24.11, minimum = 0, maximum = 100

State B : mean = 25.96, standard deviation = 19.59, minimum = 0, maximum = 84

State C : mean = 46.61, standard deviation = 20.25, minimum = 0, maximum = 100)

- $x_{i\ LC5}$: Percentage of land cover in grid cell that is scrub

(State A : mean = 7.89, standard deviation = 7.45, minimum = 0, maximum = 64

State B : mean = 7.66, standard deviation = 9.42, minimum = 0, maximum = 62

State C : mean = 1.35, standard deviation = 2.24, minimum = 0, maximum = 24)

- $x_{i\ LC7}$: Percentage of land cover in grid cell that is grass

(State A : mean = 3.75, standard deviation = 5.17, minimum = 0, maximum = 84

State B : mean = 3.55, standard deviation = 4.46, minimum = 0, maximum = 36

State C : mean = 7.62, standard deviation = 6.2, minimum = 0, maximum = 52)

- $x_{i\ LC8}$: Percentage of land cover in grid cell that is pasture

(State A : mean = 16.03, standard deviation = 16.98, minimum = 0, maximum = 86

State B : mean = 7.03, standard deviation = 11.61, minimum = 0, maximum = 69

State C : mean = 19.19, standard deviation = 16.44, minimum = 0, maximum = 90)

- $x_{i\ LC9}$: Percentage of land cover in grid cell that is wetland

(State A : mean = 9.45, standard deviation = 13.44, minimum = 0, maximum = 90

State B : mean = 18.92, standard deviation = 16.71, minimum = 0, maximum = 94

State C : mean = 9.22, standard deviation = 12.26, minimum = 0, maximum = 96)

- $y_{i\ DPoles}$: Number of damaged poles in grid cell i

(State A : mean = 108.5, standard deviation = 127, minimum = 0, maximum = 491)

- $y_{i\ DTransformers}$: Number of damaged transformers in grid cell i

(State A : mean = 43.7 standard deviation = 57, minimum = 0, maximum = 292)

- $x_{i\ SPole}$: Total number of poles for the grid cells used in the damage model

(State A : mean = 30,192, standard deviation = 17,110, minimum = 1,066, maximum = 62,587)

- $x_{i\ STransformer}$: Total number of transformers for the grid cells used in the damage model

(State A : mean = 10,256 standard deviation = 6,306, minimum = 196, maximum = 21,993)

In my model, each observation (e.g., each row in the data table) corresponds to a single grid cell during a single hurricane, and all grid cell-hurricane combinations were included. For example, for State A there are five hurricanes and 6,681 grid cells, meaning that my data table for this state has 33,405 rows. In the model, the hurricane indicator variable is treated as any other predictor. For example, the hurricane indicator variable $x_{i,Danny}$ equals one for outages that occurred during hurricane Danny and zero for outages that occurred during the other hurricanes. This essentially acts to shift the statistical model by a constant relative to the other hurricanes. The intercept of the statistical model is the expected value for Hurricane Katrina for a given set of values for the other explanatory variables because all of other hurricane indicator variables equal zero for Hurricane Katrina.

4. POWER OUTAGE PREDICTION MODEL**

For each state, I fit a series of negative binomial GLMs with either hurricane indicator variables or an alternate set of hurricane descriptor variables, discussed in Section 4.3, that are measurable prior to landfall of a hurricane. While I started by fitting a Poisson GLM for each state, there were clear indications of overdispersion in the data set (e.g., the overdispersion parameters in the initial Poisson GLMs were significant), so I focused further model fitting efforts on negative binomial GLMs that explicitly account for this overdispersion. Also, I fit a series of negative binomial GAMs with the same alternate set of hurricane descriptor variables as I fit the negative binomial GLMs for State A. Negative binomial GLMs were used for accounting for non-linearity of the data.

4.1 Handling Correlation in the Explanatory Variables

As discussed above, a GLM is based on the assumption that the explanatory variables are statistically independent of one another. However, there is significant correlation between many of the variables in my data sets. In order to account this high degree of correlation between many of the variables, I used a PCA to transform the input data.

I conducted the PCA using all of the covariates except for the hurricane indicator variables and the alternate hurricane descriptors. While I could have included these variables in the PCA, this would have produced two sets of principal components, one for the model based on hurricane indicator variables and one for the model based on the alternate hurricane descriptors. This would have complicated the comparison of the results from these two models. Instead, I chose to leave these variables out of the PCA. This yields a set of principal components based on the remaining covariates that are identical regardless of whether the hurricane indicator variables or the alternate hurricane descriptors are used. In addition, the other covariates accounted for the correlation problems in the data set. I then fit negative binomial GLMs based on the

** This section is adapted from Han et al. (2008a) and follows the text of that paper closely.

transformed covariates. Rather than using only a portion of the resulting transformed variables, I used all of them in the analysis, resulting in no loss of information in the set of explanatory variables used. In situations with a large number of explanatory variables, PCA can be useful for data reduction as well. PCA guarantees that the variables used in the regression are not collinear, yielding stable regression parameter estimates. However, it does make the interpretation of the model results more challenging relative to a model in which the data was not transformed with a PCA. This will be discussed below.

4.2 Negative Binomial GLMs Using Hurricane Indicator Variables

I first report the fits of negative binomial GLMs based on the hurricane indicator variables and the transformed covariates. The model for each of the three states was fit separately. I fit these models by starting with a model that included all of the transformed covariates and then iteratively removing the transformed covariate with the highest p-value until the p-values for all of the regression parameters were below 0.05. I formally compared the intermediate models using likelihood ratio tests with the null hypothesis that the difference in deviance for the two models was zero and the alternative hypothesis that the difference was different than zero. The full details of all of the model fits are given in the tables in Appendix B. I also used the two pseudo-R^2 described above.

The best fitting model for State A based on hurricane indicator variables had a deviance of 18,891 on 33,380 degrees of freedom, a statistically significant improvement over the intercept-only model at a p-value less than 0.001. The pseudo-R^2 values for the best-fitting model were 0.622 (R^2_{dev}) and 0.843 (R^2_{α}). Together, the deviance values, likelihood ratio test results, and pseudo-R^2 values suggest that the best-fitting model fits the data well and that including the regression parameters helps to both reduce the deviance and explain more of the above-Poisson variability in the data set. In this model, the parameter values for the indicator variables for hurricanes Danny, Dennis, and Georges were -2.1, -0.94, and -1.3, all significant at a p-value less than 0.001. In this model, 21 of the 26 transformed covariates were statistically significant at a p-value less than 0.01. The estimated overdispersion parameter value was 1.22 (significantly different from zero for $p<0.01$), suggesting that there is significant overdispersion in this data set. Recalling that the hurricane indicator variables act as shifts in the model intercept relative to Hurricane Katrina, these results suggest that there were fewer outages on average during hurricanes Danny, Dennis, and Georges than during Hurricane Katrina in this state.

The best fitting model for State B based on hurricane indicator variables had a deviance of 1,898 on 1,789 degrees of freedom, a statistically significant improvement over the intercept-only model at a p-value less than 0.001. The pseudo-R^2 values for the best-fitting model were 0.731 (R^2_{dev}) and 0.817 (R^2_{α}). These fit results suggest that the best fitting model for State B fits the data well and that including covariates helps to reduce the deviance and above-Poisson variability. In this model, the parameter values for the indicator variables for Hurricanes Dennis and Ivan were –0.36 and 2.8, both significant at a p-value less than 0.02. In this model, 14 of the 24 transformed covariates were statistically significant at a p-value less than 0.01. The estimated overdispersion parameter value was 0.81 (significantly different from zero for $p<0.01$), suggesting that there is overdispersion in this data set. These results suggest that there were fewer outages on average during Hurricane Dennis than during Hurricane Katrina in this state but more outages during Hurricane Ivan than during Hurricane Katrina. The larger coefficient for the Ivan indicator variables shows that the effect of Hurricane Ivan on the number of outages was stronger in State B than in State A, both judged relative to the effects of Hurricane Katrina.

The best fitting model for State C based on hurricane indicator variables had a deviance of 10,844 on 58,619 degrees of freedom, a statistically significant improvement over the intercept-only model at a p-value less than 0.001. The pseudo-R^2 values for the best-fitting model were 0.617 (R^2_{dev}) and 0.911 (R^2_{α}). As with States A and B, these fit results suggest that the best fitting model for State C fits the data well and that including covariates helps to reduce the deviance and above-Poisson variability. In this model, the parameter values for the indicator variables for hurricanes Cindy, Frances, Isidore, Ivan, and Jeanne were -0.82, 1.8, -0.90, 0.61, and 0.64, all significant at a p-value less than 0.001. In this model, 15 of the 25 transformed covariates were statistically significant at a p-value less than 0.05. The estimated overdispersion parameter value was 2.45 (significantly different from zero at $p<0.01$), suggesting that there is significant overdispersion in this data set. These results suggest that there were fewer outages on average during Hurricanes Cindy and Isidore than during Hurricane Katrina in this state but more outages during Hurricane Frances, Ivan, and Jeanne than during Hurricane Katrina.

The models discussed in this section have all relied on the use of hurricane indicator variables as Liu et al. (2005) did. However, using these models for prediction would require plugging values into the hurricane indicator variables of the regression model for a certain hurricane. Yet these indicator variables are for past hurricanes, not future hurricanes. While one could assume that the approaching hurricane is like the average of the past hurricanes (e.g., run the model once for each of the past hurricanes and then average the predictions), it would be preferable for a prediction model to be based only on measurable characteristics of hurricanes. This would likely give decision-makers greater confidence in the predictions, and it would allow the model to be used effectively for hurricanes that are not like the average of the previous hurricanes.

4.3 Negative Binomial GLMs with Alternate Hurricane Descriptors

To overcome the difficulties posed by using hurricane indicator variables in predictive models, I replaced the hurricane indicator variables with more directly measurable characteristics of hurricanes. My goal was to replace the indicator variables, which could not be measured for future hurricanes, with variables that could be measured for an approaching hurricane. I tried many different variables, but the ones that gave the best fits and statistical significance were the time between landfall of the hurricane being modeled and the time of the landfall of the previous hurricane (in months), the radius of the maximum winds at landfall (in km), and the central pressure difference at landfall (in mb). Each of these can be reasonably estimated as a hurricane is approaching based on public data provided by the National Hurricane Center web page (www.nhc.noaa.gov/), making them useful covariates in a practical predictive model. I replaced the hurricane indicator variables with these parameters and refit the negative binomial GLMs using the principal components. I refer to the new set of hurricane variables as the alternate hurricane descriptors to distinguish them from the hurricane indicator variables. Tables in Appendix B give the regression parameter estimates and p-values of power outage prediction models fitted by the negative binomial GLM using the principal components and alternate hurricane descriptors.

The best fitting model for State A based on alternate hurricane descriptors had a deviance of 18,884 on 33,379 degrees of freedom, a statistically significant improvement over the intercept-only model at a p-value less than 0.05. The pseudo-R^2 values for the best-fitting model were 0.632 (R^2_{dev}) and 0.842 (R^2_{α}). The results suggest that this model fits the data well and that the inclusion of the explanatory variables reduces the deviance and the above-Poisson variability relative to an intercept-only model. In this model, the parameter values for the alternative hurricane descriptors for Pressure ($x_{Pressure}$) and Time (x_{Time}) were 0.03 and 0.02, both significant at a p-value less than 0.001, meaning that these new variables do improve the fit of the model. In this model, 23 of the 26 transformed covariates were statistically significant at a p-value less than 0.05. The estimated overdispersion parameter value was 1.22 (different from zero at $p<0.01$), showing that there is significant overdispersion in this data set. Furthermore, this overdispersion parameter was the same as the best fitting model for this state based on hurricane indicator variables. This shows that the alternate hurricane descriptors are explaining at least as much of the above-Poisson variability in the data as the hurricane indicator variables, but in a different way, using measurable characteristics of a hurricane. Overall, the results for State A suggest that as central pressure difference increases and the time interval between hurricanes increases, there will be, on average (across grid cells), more outages during a hurricane.

The best fitting model for State B based on alternate hurricane descriptors had a deviance of 1,876 on 1,787 degrees of freedom, a statistically significant improvement over the intercept-only model at a p-value less than 0.05. The pseudo-R^2 values for the best-fitting model were 0.732 (R^2_{dev}) and 0.817 (R^2_{α}). The results suggest that this model fits the data well and that the inclusion of the explanatory variables reduces the deviance and the above-Poisson variability relative to an intercept-only model. In this model, of the three alternate hurricane descriptors, only x_{Time} was significant at a p-value less than 0.001 with a parameter value of 0.03. In this model, 17 of the 24 transformed covariates were statistically significant at a p-value less than 0.05. The estimated overdispersion parameter value was 0.81 (different from zero at $p<0.01$), showing that there is significant overdispersion in this data. The estimated value of the overdispersion parameter was the same as for the best fitting model for State B based on hurricane indicator variables as it was for the State A model. The results for State B suggest that longer time intervals between hurricanes are associated with more outages, on average (across grid cells), during hurricanes.

The best fitting model for State C based on alternative hurricane descriptors had a deviance of 16,642 on 33,378 degrees of freedom, a statistically significant improvement over the intercept-only model at a p-value less than 0.05. The pseudo-R^2 values for the best-fitting model were 0.611 (R^2_{dev}) and 0.904 (R^2_{α}). The results suggest that this model fits the data well and that the inclusion of the explanatory variables reduces the deviance and the above-Poisson variability relative to an intercept-only model. In this model, the parameter values for the alternative hurricane descriptors x_{Time} and x_{RMW} were 0.03 and -0.11, both significant at a p-value less than 0.001. In this model, 19 of the 25 transformed covariates were statistically significant at a p-value less than 0.05. The estimated overdispersion parameter value was 2.62 (different from zero at $p<0.01$), the highest among three states. This state also experienced few strong hurricanes during the period for which I have outage data, perhaps leading to higher variability in the number of outages, even once conditioned on the explanatory variables. Overall, the results for State C suggest that the longer time interval between hurricanes, the higher the number of outages, on average, during hurricanes. This agrees with the results for States A and B. In addition, the results for State C suggest that hurricanes with smaller radii of the maximum wind tend to be associated with more power outages, on average. This may be because a small radius of maximum wind indicates that the hurricane has a "stiff" vortex shape and thus a conspicuous eye. This would tend to concentrate the energy of the vortex more tightly around the center, potentially leading to more damage near the center of the hurricane. However, this still must be treated with caution because this state experienced only weak hurricanes during the period for which I have data. Further analysis based on future storms may help to substantiate or refute this hypothesis.

In order to more directly compare the fits of the models based on the alternate hurricane descriptor variables with those of the models based on the hurricane indicator variables, I focus next on the models that use all available covariates, even if some of them had p-values above 0.05. These are called the saturated models. This is done so that the models are based on the same set of information. The deviances of the saturated models based on alternate hurricane descriptor variables for the three states are 18,883 on 33,375 degrees of freedom, 1,899 on 1,779 degrees of freedom, and 10,782 on 58,611 degrees of freedom respectively. The deviances of the saturated models based on hurricane indicator variables for the three states are 18,880 on 33,374 degrees of freedom, 1,899 on 1,779 degrees of freedom, and 10,843 on 58,607. I see that there is not much of a difference between the deviances for States A and B, indicating that for these two states the two types of models provide very similar fits to the data. The deviance for State C is lower (better) with the alternative hurricane descriptors than with the hurricane indicator variables, indicating that the model based on the alternate hurricane descriptors may provide a better fit to the data for this state. Using the alternative hurricane descriptors, which are relatively easy to obtain, I obtain a more useful model for predicting the number of outages while achieving at least an equivalent goodness of fit to the data. I also examined the residuals (raw, Pearson, and deviance) of the different models and checked for outliers in the predictions. The models based on both the hurricane indicator variables and the alternate hurricane descriptors both had some problems with outliers in the predictions for a few of the grid cells in the most heavily urbanized areas. However, the use of the alternate hurricane descriptors rather than the hurricane indicator variables did not affect the degree to which there were outliers. Assessments of residuals also suggest that the overall fits of the models based on alternate hurricane indicators are at least as good as those based on hurricane indicator variables.

4.4 Examples of Model Predictions and Overall Assessment of Predictive Accuracy

To further examine how well the models fit the data, I provide typical examples of the model fits for Hurricane Katrina. Figures 4.1, 4.2, and 4.3 show both the predicted mean number of outages and the actual number of outages in each grid cell for portions of the three states for Hurricane Katrina. Note that the outage maps shown in these figures are based on interpolating between the grid-based outage numbers using inverse distance weighting in ArcINFO. The geographic pattern of model predictions is generally accurate except for overpredictions in the main urban areas, those areas with the highest number of actual and predicted outages shown on the maps. In these few grid cells there is a much higher amount of overhead line than in the other grid cells. It appears that the relationship between the amount of overhead line and the log of the mean number of outages expected in each grid cell is non-linear. A GLM cannot incorporate this non-linearity. This non-linearity will be discussed further below. The accuracy of the geographic pattern of model predictions is very similar for the models based on alternate hurricane descriptors and the models based on hurricane indicator variables.

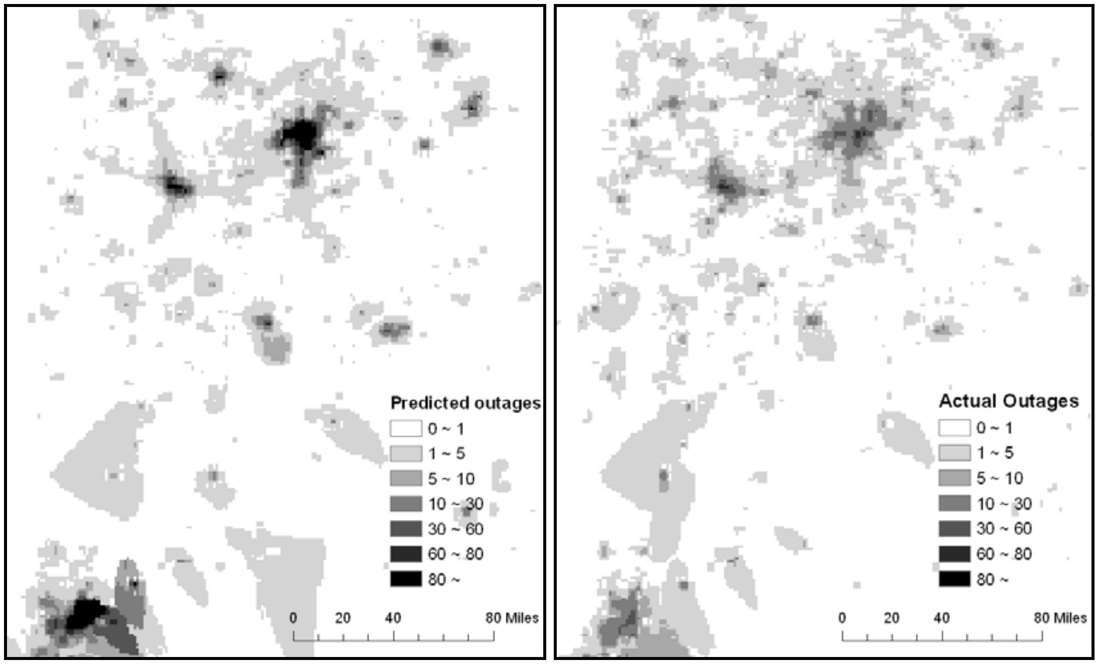

Figure 4.1. Predicted number of outages (left plot) and actual number of outages (right plot) in State A during Hurricane Katrina.

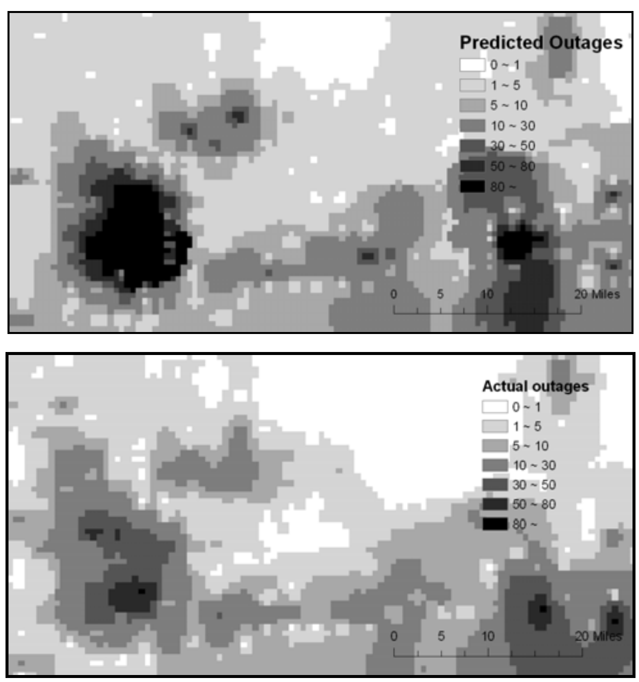

Figure 4.2. Predicted number of outages (above plot) and actual number of outages (below plot) in State B during Hurricane Katrina.

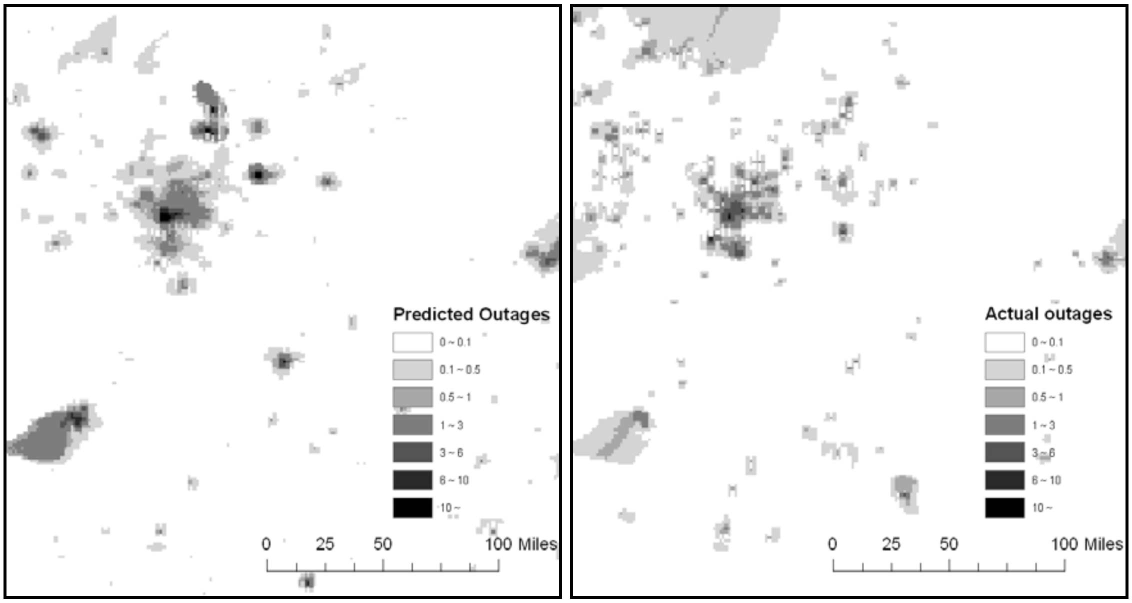

Figure 4.3. Predicted number of outages (left plot) and actual number of outages (right plot) in State C during Hurricane Katrina.

To further test the predictive accuracy of the models, I also conducted hold-out analysis. I removed the data for a single hurricane (e.g., Katrina) from the data set, fit the model to the remaining data, used the fitted model to predict the number of outages in each grid cell during Hurricane Katrina, and then calculated the mean absolute error between the actual number of outages and the predicted number of outages (the MAE). I repeated this process for each of the hurricanes for each state. Dividing the MAE by the mean number of outages yields an estimate of the error in the predictions from the model. Tables 4.1, 4.2, and 4.3 show the results for State A, State B, and State C. Testing the predictive accuracy of the models that utilize the hurricane indicator variables is more challenging because it is not entirely clear how to treat the hurricane indicators when making predictions for a hurricane not in the fitting data set. I used the same hold-out method for the indicator-based models. For a given withheld hurricane, I first re-fit the indicator-based model excluding the indicator variable for the withheld hurricane. I then estimated the number of outages in each grid cell four times (once each per hurricane), with a different indicator variable set equal to one each time. I then averaged across the four predictions to obtain the predictions for the withheld hurricane. These estimates where then used in calculating the MAE values for the indicator-based models.

Table 4.1. Predictive accuracy of the statistical models for hold-out samples in State A.

	Danny (1997)	Georges (1998)	Ivan (2004)	Dennis (2005)	Katrina (2005)
Actual number of Outages	627	1,075	13,568	4,840	10,105
$\mu_{outages}$	0.0938	0.1609	2.0308	0.7244	1.5125
$\frac{MAE_{Hurricane\ indicator\ variables}}{\mu_{outages}}$	56.50	41.40	2.223	21.05	2.605
$\frac{MAE_{Alternative\ hurricane\ descriptors}}{\mu_{outages}}$	8.560	15.35	1.000	13.57	4.707

Table 4.2. Predictive accuracy of the statistical models for hold-out samples in State B.

	Ivan (2004)	Dennis (2005)	Katrina (2005)
Actual number of Outages	14,948	4,683	3,446
$\mu_{outages}$	24.83	7.779	5.724
$\frac{MAE_{Hurricane\ indicator\ variables}}{\mu_{outages}}$	0.8782	18.84	21.16
$\frac{MAE_{Alternative\ hurricane\ descriptors}}{\mu_{outages}}$	0.8240	1.911	1.170

Table 4.3. Predictive accuracy of the statistical models for hold-out samples in State C.

	Hanna (2002)	Isidore (2002)	Francis (2004)	Ivan (2004)	Jeanne (2004)	Cindy (2005)	Dennis (2005)	Katrina (2005)
Actual number of Outages	253	143	2,951	1,843	648	255	1,027	518
$\mu_{outages}$	0.0345	0.0195	0.4026	0.2514	0.0884	0.0348	0.1401	0.0707
$\frac{MAE_{Hurricane\ indicator\ variables}}{\mu_{outages}}$	3.432	9.736	0.9469	1.645	2.006	7.361	2.632	2.949
$\frac{MAE_{Alternative\ hurricane\ descriptors}}{\mu_{outages}}$	7.332	2.676	1.047	2.886	2.455	2.520	1.378	5.377

The earlier discussion of model fit showed that the indicator-based models and the models using the alternate hurricane descriptors yielded very similar fits to the full data set. However, the results in Table 4.1 and Table 4.2 show that the model based on the alternate hurricane descriptors generally provides more accurate predictions for hurricanes not in the fitting data set and that in some cases the difference in predictive accuracy is large. From Table 4.1 I see that the error as a fraction of the average number of outages varies from 2 times up to 57 times for the indicator-based model and from one time up to 14 times for the model based on the alternate hurricane descriptors. For Hurricanes Danny, Georges, and Dennis, the model using the alternate hurricane descriptors has a substantially lower prediction error than the indicator-based model. For Hurricanes Ivan, the prediction error is lower with the alternate hurricane descriptors, but the difference is not as great. Only for Hurricane Katrina is the error of the indicator-based model lower, and in this case the difference is not high. Similar results are seen for State B. Even though the difference in predictive accuracy for State C is not large, the error as a fraction of the average number of outages for the model based on the alternate hurricane descriptors is less than the maximum error for the indicator-based model. Overall, the model based on the alternate hurricane descriptors, physically measurable characteristics of hurricanes, does seem to provide more accurate predictions for hurricanes not in the fitting data set than the model based on hurricane indicator variables together with the ad hoc assumption that a future hurricane is like the average of the past hurricanes. Being able to make outage predictions based on measurable characteristics of hurricanes does increase the accuracy of the predictions for hurricanes not in the fitting data set without a loss of fit to the past data. While these are not perfect predictions, this does provide a strong basis for making resource allocations, especially given the high degree of variability in the spatial distribution of outages during hurricanes.

Overall, the results suggest that the outage model can provide the type and accuracy of information needed to help guide state-wide hurricane preparation. My results show that for a strong hurricane such as Hurricane Katrina, the model is a good predictive model for those areas outside of the urban areas, and the hold-out analysis results suggest that the model accuracy is good for most hurricanes. However, within the main urban areas, the results of the model should be used with caution. The model is more useful in making comparisons between different portions of the state than for comparing precise outage estimates from small grid cells immediately adjacent to one another. This is appropriate given that the model is intended to help guide state-wide resource allocations rather than to provide very precise predictions for small, local areas.

4.5 Relative Importance of Explanatory Variables

In addition to their usefulness for predicting outages in future hurricanes the models can be used to understand the association between the explanatory variables and outages by examining the relative importance of the parameters. To evaluate the relative impacts of the different explanatory variables on the mean number of counts in a GLM, the relative rate of change in $\mu(x)$ with respect to a unit change in x_j can be written as, (Cameron and Trivedi 1998);

$$\delta_j = \left(\frac{1}{\mu(x)}\right)\frac{\partial \mu(x)}{\partial x_j} = \beta_j \tag{4.1}$$

For discrete indicator variables such as the hurricane indicator variable, the interpretation of the derivative is more problematic because the variable can take on only two values, 0 and 1. However, I use the same formula as in Cameron and Trivedi (1998) for consistency. While the original explanatory variables have different units and variability, in the process of conducting the PCA, I standardized the data to have a mean of 0 and a standard deviation of 1, so that the meaning of a unit change is consistent across variables.

The parameters δ_j must be back-calculated from the regression parameters of the models based on principal components. This is done by using the weightings (i.e., the eigenvectors) that result from the PCA to calculate the importance parameter for each of the original covariates as a weighted linear combination of the regression parameter estimates for the principal components, just as the principal components were calculated as a linear combination of the original data. The end result is a set of parameters, δ, that provides an indication of the impacts of changes in each explanatory variable on the expected number of outages and thus is a measure of the relative importance of the different explanatory variables. For comparison purposes, I include the models based on both the hurricane indicator variables and the alternate hurricane descriptors in my analysis in this section. This can yield useful insights into the role of the hurricane indicator variables and which hurricanes were particularly problematic in terms of outages.

Figures 4.4, 4.5, and 4.6 show the relative rate of change of the predicted mean number of outages with respect to changes in the different explanatory variables: transformers, poles, switches, miles of overhead line, miles of underground line, number of customers served, windspeed, duration of strong winds, FSMs, MAP, SPIs, and the land cover variables for each of the states. For example, for State A, if the amount of time that the winds were above 20 m/s in a grid cell increased by 1 minute with all other explanatory variables held constant, I would expect the number of outages to increase by approximately 0.5μ where μ is the number of outages that would have been predicted without the increase in the duration of strong winds. As shown in Figure 4.4, the relative impacts of the land cover variables, MAP, some of the FSMs, the miles of underground line, and the SPIs are lower than the other variables for State A. This indicates that these variables do not have a strong influence on the predicted number of power outages in the first state. Both the wind speed and duration of strong wind covariates have statistically significant and positive effects on the predicted number of outages. In Figure 4.5 I see that some of the FSM and SPI variables have a strong and statistically significant impact on the predicted number of outages for State B. The wind speed variable has a negative impact on the number of outages, the opposite of the case for State A. Figure 4.6 shows that windspeed has a positive impact on outages in State C but that the duration of strong winds does not have a substantial impact. Some of the SPI and FSM variables have a substantial impact while others do not. Looking across all three states, I see that the variables that measure the number of overhead power system components in a grid cell (the transformer, switches, overhead, and poles variables) tend to have a positive impact. While the relative magnitudes of the impacts of these parameters vary across states, the general conclusion is that having more overhead components leads to higher numbers of outages during hurricanes, as would be expected. As with the models based on the hurricane indicator variables, the relative effects of the land cover, MAP, some of the FSMs and SPIs are smaller than the other variables in the models based on the alternate hurricane descriptors. In the models in which the hurricane indicator variables were replaced with the alternative hurricane descriptors, the relative effects of the other explanatory variables tend to increase. The alternate hurricane descriptors are statistically significant, and they do have some impact on the predictions, but this impact is not strong. The overall results are mixed for the wind speed variables. One would initially expect both wind speed and duration of strong winds to have a positive relationship to the number of outages. However, this is not the case. At least one of the wind speed variables has a positive relationship for each of the three states. However, the sign of the impact of the wind variables are not consistent across the states. This is likely due to the fact that the three states have experienced different types of hurricanes in the past. State A has been impacted by large, powerful hurricanes (e.g., Ivan and Katrina). The hurricanes that impacted State B during the period for which I have data have been relatively week. This may mean that outages in this state are caused more flooding or thunderstorms and less by strong wind than in other states. The positive effect of duration and negative effect of wind speed in State B supports such a conclusion. State C is an intermediate case. It was impacted by strong hurricanes during my data collection period, but more on the edges of these storms.

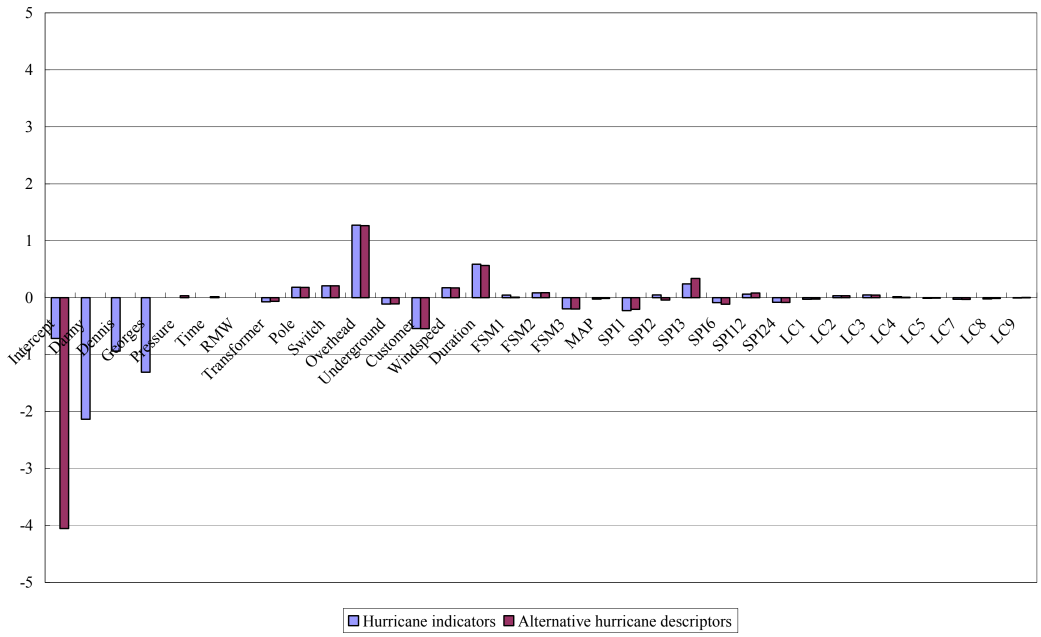

Figure 4.4. Relative effects of fixed effects, hurricane indicators and alternate hurricane descriptors of the final prediction models for State A.

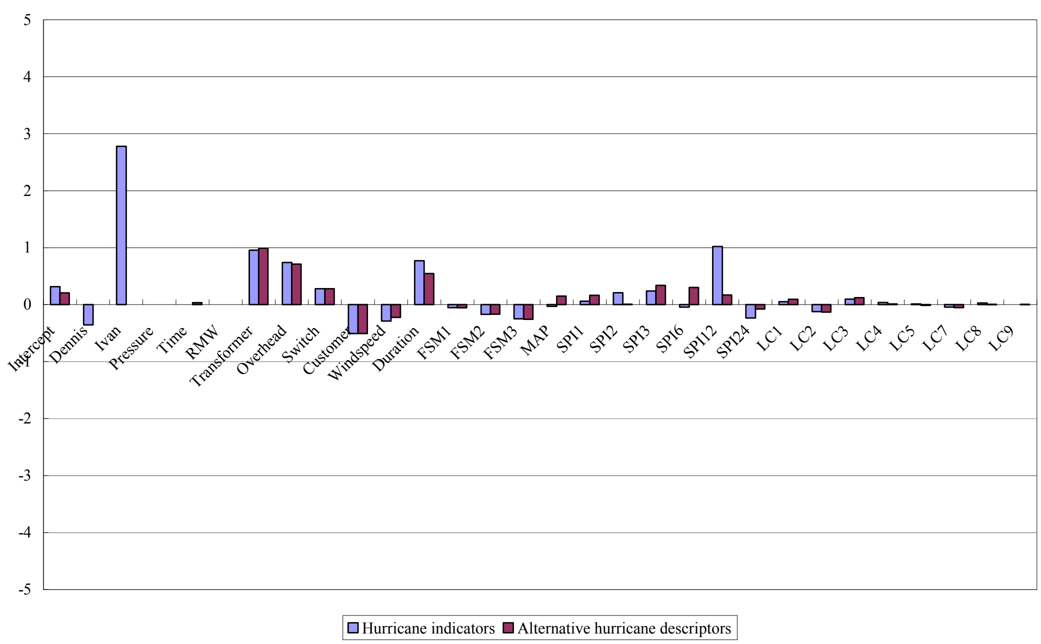

Figure 4.5. Relative effects of fixed effects, hurricane indicators and alternate hurricane descriptors the final prediction models for State B.

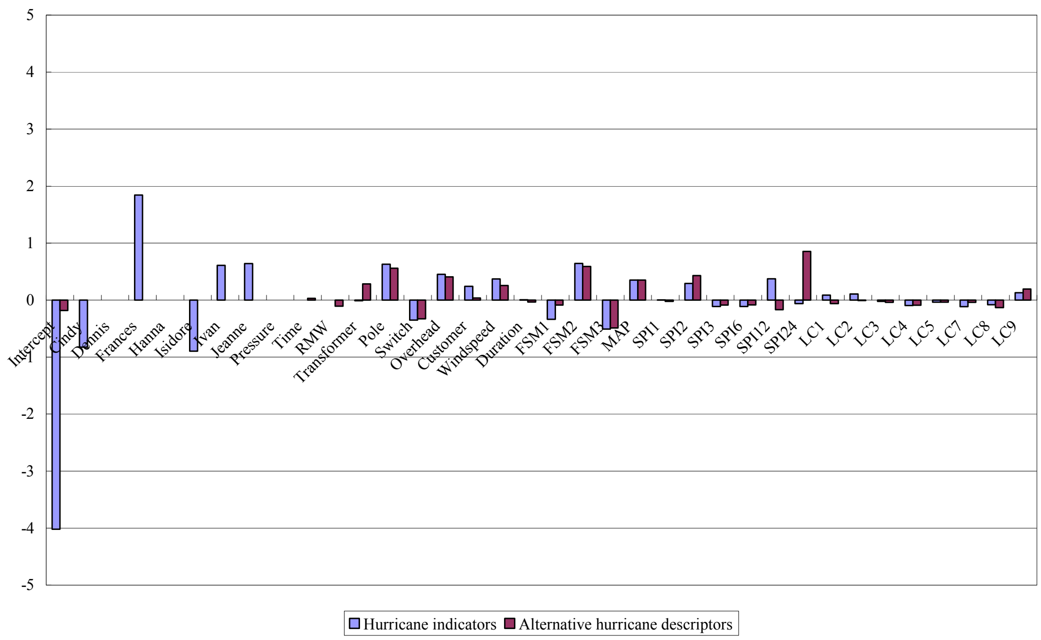

Figure 4.6. Relative effects of fixed effects, hurricane indicators and alternate hurricane descriptors the final prediction models for State C.

4.6 *GAM Fitting Process*

GLMs such as those described in the section above assume that the systematic component of the model uses a linear link function. However, in many cases there can be considerable non-linearity in the relationship between log(μ) and the covariates. No accounting for such non-linearity is one possible cause of the over-predictions in the urban areas with the GLMs. In an effort to capture this non-linearity in the link function and to provide better predictions of power outages during hurricanes, I fit negative binomial GAMs to the data described in Section 3 using the program R. Specifically I used cubic regression splines as smoothing functions (Wood 2006). Figure 4.7 shows the fitted splines for the first four principal components, showing non-linearity in relationship between these principal components and the log of the mean number of power outages. For example, the first subplot indicates considerable nonlinearity in the relationship between the first principal component and the log of the mean number of power outages. In contrast, GLMs such as those developed by Liu et al. (2005) and Han et al. (2008a) assume a linear relationship. I began with one single-term spline per explanatory variable and iteratively removed splines in order of decreasing p-value until I was left with only splines that were statistically significant at a 0.05 level before then testing the predictive accuracy of the models. Because I wanted to keep the models simple and to ease comparisons with Han et al. (2008a) where interactions among covariates were not included, I did not consider higher-order splines. I formally compared all of the models that were fit to the data on the basis of Generalized Cross Validated deviance (GCV) (Hastie and Tibshirani 1990), selecting the model with the lowest GCV as the best fit to the data.

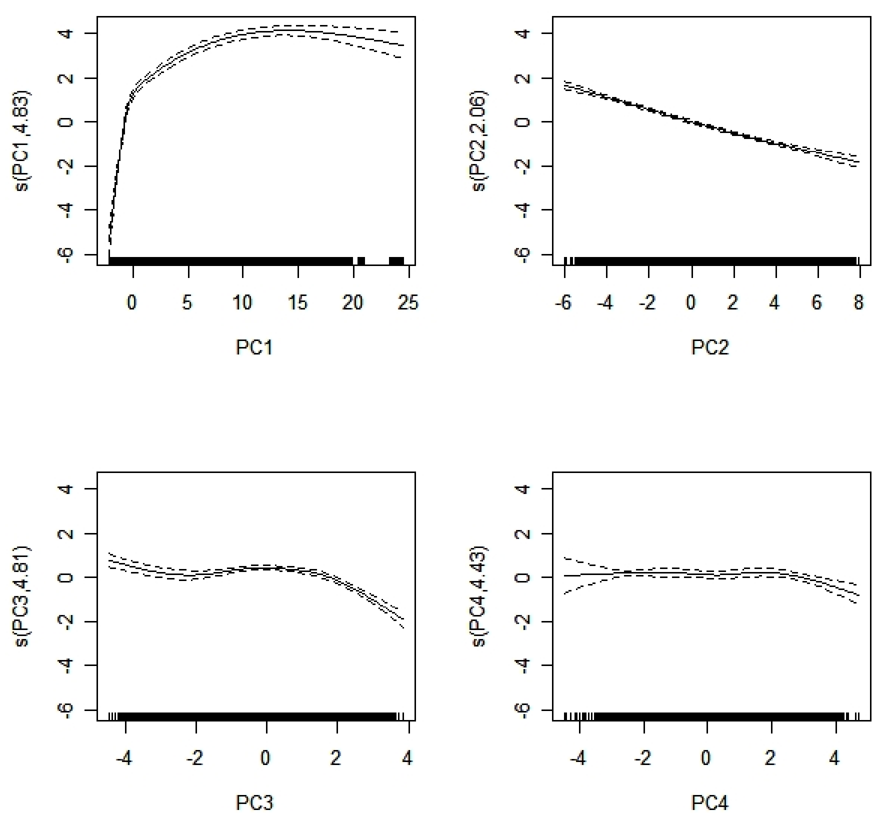

Figure 4.7. Fitted additive splines for 4 principal components.

I also repeated the hold-out sampling of the data that I did with the GLMs to test the predictive accuracy of the best-fitting GAM for hurricanes not included in the fitting data and to compare the predictive accuracy of this model with that of the best-fitting GLM from Han et al. (2008a). I divided the data into a fitting data set from which I removed one of the hurricanes and a validation set consisting of the data from the removed hurricane data. I fit a GAM containing the subset of the variables selected on the basis of the full data and then used these models to predict the number of outages of each grid cell in the validation set. By repeating this process for each hurricane, I was able to estimate the predictive accuracy of the models for data not included in the fitting data set.

4.7 *GAM Results*

Table 4.4 gives the model fit diagnostics for the negative binomial GAMs. For comparison purposes, the best fit negative binomial GLM from Han et al. (2008a) is also included in Table 4.4, and it had a deviance of 18,884 on 33,379 degrees of freedom. In Table 4.4, negative binomial GAM 0 represents the saturated model, the model with single-term splines of all PCA-transformed covariates included. Negative binomial GAM 5 includes only splines of the principal components with p-values below 0.05.

From Table 4.4 I see that the deviance and AIC for the negative binomial GAMs are lower than those for the best-fit negative binomial GLM, suggesting that the GAMs fit the data better than the best-fit GLM. In addition, Table 4.4 shows that for the GAM models, all values of R_α^2, a pseudo-R^2 based on the ovedispersion parameter α, are approximately 1 and are higher than the R_α^2 values for the best-fit negative binomial GLM. This suggests that the GAM models are accounting for more, and in fact nearly all, of the overdispersion. The variability that remains in the predicted counts is primarily due to the Poisson variability about the mean. Another diagnostic for comparison of models is the GCV of the regression model. While lower AIC and deviance values are generally preferable, I selected GCV as my primary criteria in comparing the fits of different negative binomial GAMs because of its advantages in terms of invariance

(Wahba 1990). Based on AIC and GCV, negative binomial GAM 4 gives the best-fit models to the data set.

Table 4.4. Comparison between NB GLM and NB GAMs.

Model	Deviance	Degrees of Freedom	AIC	R_α^2	GCV	Variables Excluded
negative binomial GLM	18,884	33,379	53,154	0.8424	—	RMW, PC 9, 13, 26
negative binomial GAM 0	15,276	33,311	49,395	0.9990	1.0028	None
negative binomial GAM 1	15,276	33,311	49,395	0.9990	1.0028	PC 17
negative binomial GAM 2	15,266	33,314	49,398	0.9990	1.0027	PC 17, 24
negative binomial GAM 3	15,312	33,315	49,382	0.9990	1.0027	RMW, PC 17, 24
negative binomial GAM 4	15,280	33,319	49,395	0.9990	1.0026	RMW, PC 11, 17, 24
negative binomial GAM 5	15,281	33,319	49,396	0.9990	1.0026	RMW, PC 11, 17, 24, 26

Figure 4.8 shows the outage predictions from negative binomial GAM 4 for Hurricane Katrina. Comparing this map of predicted outages with the map of the actual number of outages (Figure 4.1), I see that the GAM predictions match the spatial distribution of outages much more closely than the GLM predictions do. Similar results are seen for the other four hurricanes, though they are not displayed for the sake of brevity.

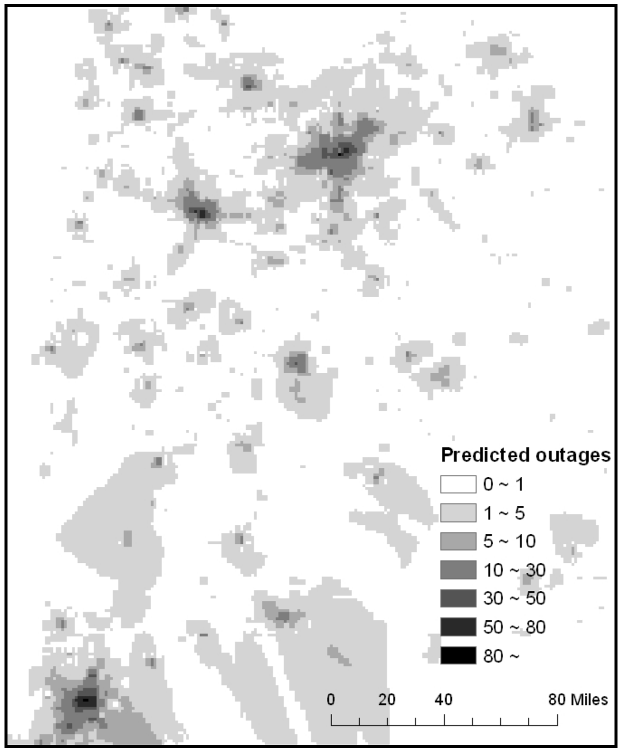

Figure 4.8. Number of outages predicted with the GAM for Hurricane Katrina.

As mentioned above, the negative binomial GLM of Han et al. (2008a) over-estimated the number of outages substantially in some grid cells, and these over-estimates influence the overall MSE for the GLM. In examining the grid cells corresponding to these outliers in detail, it was noticed that the grid cells were predominantly in areas with high amounts overhead line relative to other grid cells and that these seemed to be driving the overprediction for these areas. On the other hand, Figure 4.9 shows that the predicted number of outages grows approximately linearly (with associated variability) with the actual number of outages for the negative binomial GAM, suggesting that the GAM overcomes the over-estimation problem. Again, similar results are seen for the other hurricanes, but these results are not shown here for the sake of brevity.

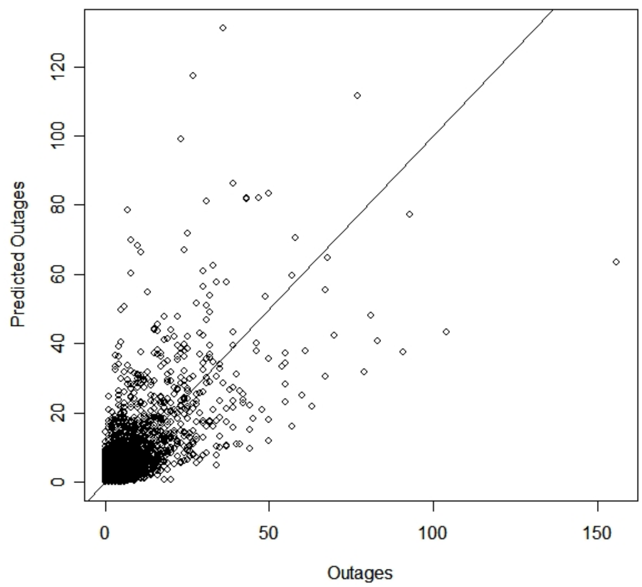

Figure 4.9. Predicted number of outages vs. actual number of outages for the best fit negative binomial GAM for Hurricane Katrina.

In order to check the predictive accuracy of the GAM, hold-out tests were performed for each hurricane and the averages of the absolute values of the difference between the actual number of outages and the predicted number of outages (referred to here as MAE for mean absolute error) were calculated. Table 4.5 shows the MAEs divided by the mean of the actual number of outages ($\mu_{outages}$) for each hurricane for both the GLM and the GAM. Because the MAE gives more weight to large errors, I subdivided the MAE into 4 categories in terms of the actual number of outages in order to get a more complete picture of prediction accuracy for this model. The categorized $MAE/\mu_{outages}$ provides a measure of the relative prediction error for each outage range. For example, for Hurricane Katrina, the GLM outage predictions differ, on average across the grid cells, by 32% of the actual number of outages for grid cells with 0 to 1 outages, 1.3 times for grid cells with 1 to 10 outages, 111 times for grid cells with 10 to 50 outages, and 101 times for grid cells with over 50 outages and the GAM outage predictions differ by 24% of the actual number of outages for grid cells with 0 to 1 outages, 1 time for grid cells with 1 to 10 outages, 7 times for grid cells with 10 to 50 outages, and 14 times for grid cells with over 50 outages. Note that these errors are all defined based on dividing the MAE by the average number of outages per grid cell for the hurricane (μ_{outages} = 10,105/6,681). As discussed above, the GLM over-estimates outages for some grid cells. In addition, the predictive accuracy of the GLM is highly variable across hurricanes. For Hurricane Dennis the MAE of the GAM for 10 to 50 outage range is approximately 1,113 times the actual number of the outage counts while for the GAM it is 10 times. The GAM on the other hand provides consistently low prediction errors than the GLM provides for all hurricanes. Overall, the results suggest that GAMs can provide much more accurate outage predictions than GLMs across a variety of types of hurricanes, including large, powerful hurricanes like Hurricanes Katrina and Ivan and smaller, weaker hurricanes like Hurricane Danny. While there is still error in the predictions, the results provide a much better basis for allocating repair crews among the different geographic portions of the service area.

Table 4.5. Ratio of MAEs to the mean of the actual number of outages for Hold-Out sampling fitted by NB GLM and NB GAM.

		Danny (1997)	Georges (1998)	Ivan (2004)	Dennis (2005)	Katrina (2005)
Actual number of Outages		627	1,075	13,568	4,840	10,105
μ_{outages}		0.0938	0.1609	2.0308	0.7244	1.5125
$\frac{\text{MAE}_{\text{GLM}}}{\mu_{\text{outages}}}$	$0 \sim 1_{\text{ outages}}$	7.111	1.068	0.0003	0.5287	0.3206
	$1 \sim 10_{\text{ outages}}$	36.21	134.2	1.345	5.118	1.299
	$10 \sim 50_{\text{ outages}}$	87.94	344.9	8.576	1,113	111.4
	$50 \sim_{\text{ outages}}$	—	—	35.55	—	101.4
$\frac{\text{MAE}_{\text{GAM}}}{\mu_{\text{outages}}}$	$0 \sim 1_{\text{ outages}}$	1.017	0.6564	0.0003	0.4079	0.2350
	$1 \sim 10_{\text{ outages}}$	17.59	8.628	1.343	2.070	1.072
	$10 \sim 50_{\text{ outages}}$	55.81	95.21	8.575	10.30	7.123
	$50 \sim_{\text{ outages}}$	—	—	35.54	—	14.76

5. CUSTOMERS OUT PREDICTION MODEL[††]

The models developed in the past (Liu et al. 2005) and in the previous sections of this thesis all focus on predicting the number of outages. However, predictions of the number of customers without power would be more closely aligned with the methods currently used for pre-hurricane planning in the utility company that provided the data from as well as in other utility companies. To address this gap, models were developed for predicting the number of customers without power in each grid cell in each of the three service areas. For brevity of terminology, these models are referred to as the customers out models.

In developing the customer models, I used a negative binomial GLM based on the same principal components as used in Han et al (2008a) and in the earlier sections of this thesis. These principal components consist of orthogonal transformations of the input data discussed in Sections 3 and 4. These models can account for both collinearity and overdispersion providing a good starting point for modeling the number of customers without power. The approach accounted for overdispersion and collinearity in order to obtain a better fit and more stable model estimates. The final suggested model is the negative binomial GLM based principal components and alternative hurricane descriptors (pressure difference, time between hurricanes, and radius to maximum winds) rather than the hurricane indicator variables of Liu et al. (2005).

5.1 Fitting Negative Binomial GLMs

For each state, I fit a series of negative binomial GLMs. For each model I further divided the fitting into a model based on the original data and a model based on a transformation of the data through a Principal Components Analysis (PCA).

For all three states I first fit a negative binomial GLM with all covariates. I then iteratively reduced the parameter used in these models until I found a model with all parameter p-values below 0.05. I then used likelihood ratio tests to formally compare the

[††] This section is adapted from Han et al. (2008c) and follows the text of that paper closely.

reduced and full models, in all cases showing that the reduced models provided fits that were either statistically indistinguishable from the full model or, in some cases, provided better fits given the parameters used. The full details of negative binomial GLMs with principal components for the State A, B, and C respectively are given in the tables in Appendix B. Also, I used Deviance and the pseudo-R^2 based on α to provide to the best fit to the data.

5.2 Negative Binomial GLMs Based on Principal Components with Alternate Hurricane Descriptors

Negative binomial GLM customer models were first fitted based on the principal components and the hurricane indicator variables. The principal components with high p-values (larger than 0.05) were iteratively removed and model comparisons done with likelihood ratio tests. The deviances of the saturated models (model 0) with principal components are similar to the deviances of the saturated models with correlated variables, but there are slightly less covariates in the final model with principal components than in the final model with correlated variables for all 3 states (24 covariates with principal components and 25 covariates with correlated variables for State A, 16 covariates with principal components and 18 covariates with correlated variables for State B, and 20 covariates with principal components and 23 covariates with correlated variables for State C). The final negative binomial GLMs using the principal components give more reliable and efficient fits than the final negative binomial GLMs using the original correlated data.

I replaced the hurricane indicator variables with alternative hurricane descriptors and refit the negative binomial GLMs using the principal components. I refer to the new set of hurricane variables as the alternate hurricane descriptors to distinguish them from the hurricane indicator variables. The deviances of the saturated models for State A, State B, and State C are 16,641 on 33,375 degrees of freedom, 1,891 on 1,779 degrees of freedom, and 9,012 on 58,611 degrees of freedom respectively. Comparing the deviances of the saturated models with hurricane indicator variables for State A, State B, and State C (16,641 on 33,374 degrees of freedom, 1,891 on 1,779 degrees of freedom, and 9,006 on 58,607), I see that the deviances are nearly identical. Negative binomial GLMs based on principal components and alternative hurricane descriptors provide the best models for estimating the number of customers without power in each of the grid cells. These models use only variables that are readily measurable for approaching hurricanes, and they provide a good fit to the data.

The best fitting models for State A, State B, and State C had deviances of 16,642 on 33,378 degrees of freedom, 1,891 on 1,787 degrees of freedom, and 9,012 on 58,613 degrees of freedom respectively a statistically significant improvement over the intercept-only model for each state at a p-value less than 0.05. The deviances of the best fitting models based on alternate hurricane descriptors are similar to the deviances of the saturated models, but there is a decrease in the number of principal components in the final model for all 3 states (decreases of 3 covariates with principal components out of 26 PCs for State A, 8 covariates with principal components out of 24 PCs for State B, and 2 covariates with principal components out of 25 PCs for State C). When a likelihood ratio test suggests that the saturated model and the final model are statistically indistinguishable, the preferred final model is the more parsimonious (simple) model.

Besides checking the difference in deviance of the models through likelihood ratio tests, I checked the residual variability of the best fitting models for each state with the pseudo-R^2 based on α. The α values give a sense of how much overdispersion there is in the data that my models do not explain. Higher α values indicate that there will likely be more variability beyond the Poisson variability about the mean value. However, the dispersion parameter α can vary based on the degrees of freedom for each state relatively. The pseudo-R^2 based on α provides a measure of the reduction in above-Poisson variability and thus is preferable. For the sake of comparison, the pseudo-R^2 values of the best-fitting models based on hurricane indicator variables for State A, State B, and State C were 0.391, 0.398, and 0.847 respectively. Similarly, the pseudo-R^2 values of the best-fitting models based on alternate hurricane descriptors for State A, State B, and State C were 0.390, 0.398, and 0.846 respectively. The results suggest that this model fits the data well and that the inclusion of the explanatory variables reduces the deviance and the variability relative to an intercept-only model. Also, there is no change in residual variability when the hurricane indicator variables are replaced with the alternate hurricane descriptors.

5.3 Examples of Model Prediction and Overall Assessment of Predictive Accuracy

Figures 5.1, 5.2, and 5.3 provide examples of the fit of the customers out model for Hurricane Katrina in States A, B, and C. Note that the customer outage maps shown in these figures are based on interpolating between the grid-based customer outage numbers using inverse distance weighting in ArcINFO. The geographic pattern of the customer outage model is accurate outside of the main urban areas but the model overestimates the number of customers without power within the urban areas, just as with the power outage model of Section 4. However, with the customer model, the overprediction is more dramatic. In the urban areas, there is a much higher amount of overhead line than in the other grid cells. It appears that the relationship between the amount of overhead line and the log of the mean number of customers without power expected in each grid cell is non-linear. This non-linearity of the data set causes outliers which lie in the main urban areas. As an effort to remove outlier problem, I adjusted outliers in the predictions for a few of the grid cells in the most heavily urbanized areas, based on the principle that the predicted number of customers without power should be lower than the number of customers in the grid cells. When I tested the predictive accuracy of the models, I conducted hold-out analysis with the adjusted data set.

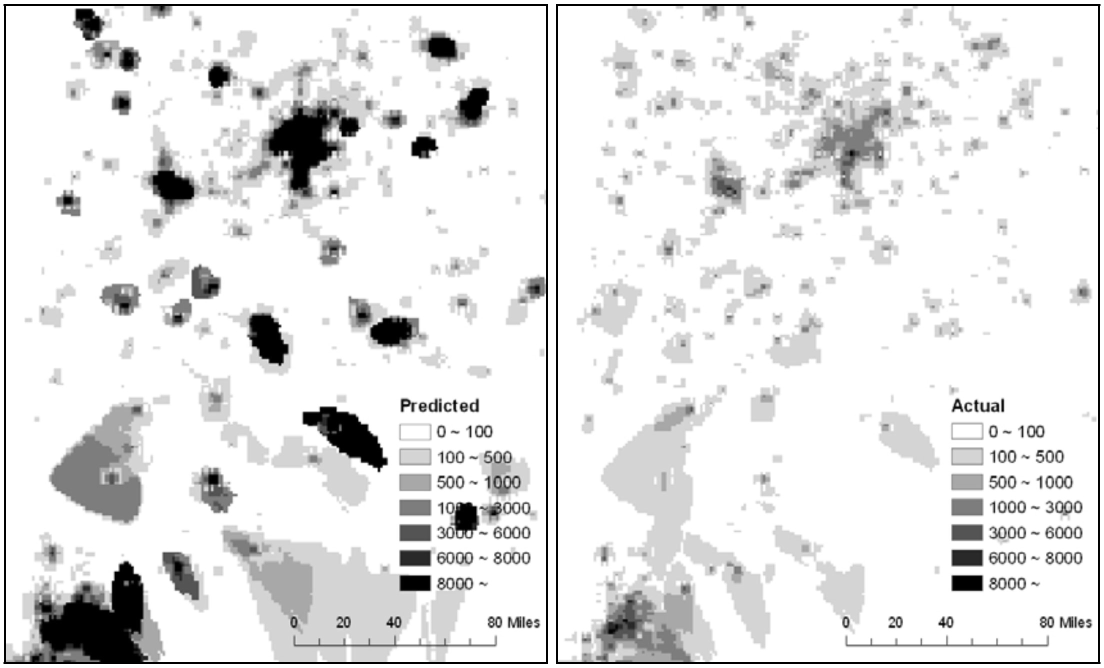

Figure 5.1. Predicted number of customers out (left plot) and actual number of customers out (right plot) in State A during Hurricane Katrina.

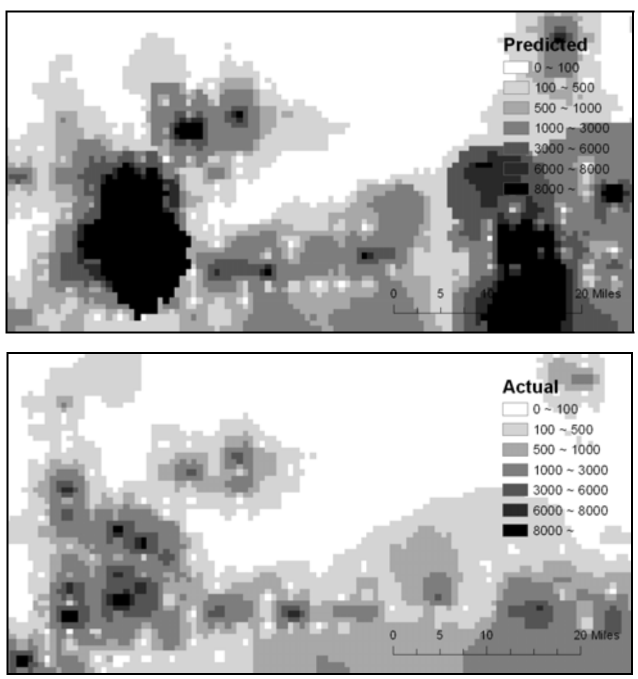

Figure 5.2. Predicted number of customers out (above plot) and actual number of customers out (below plot) in State B during Hurricane Katrina.

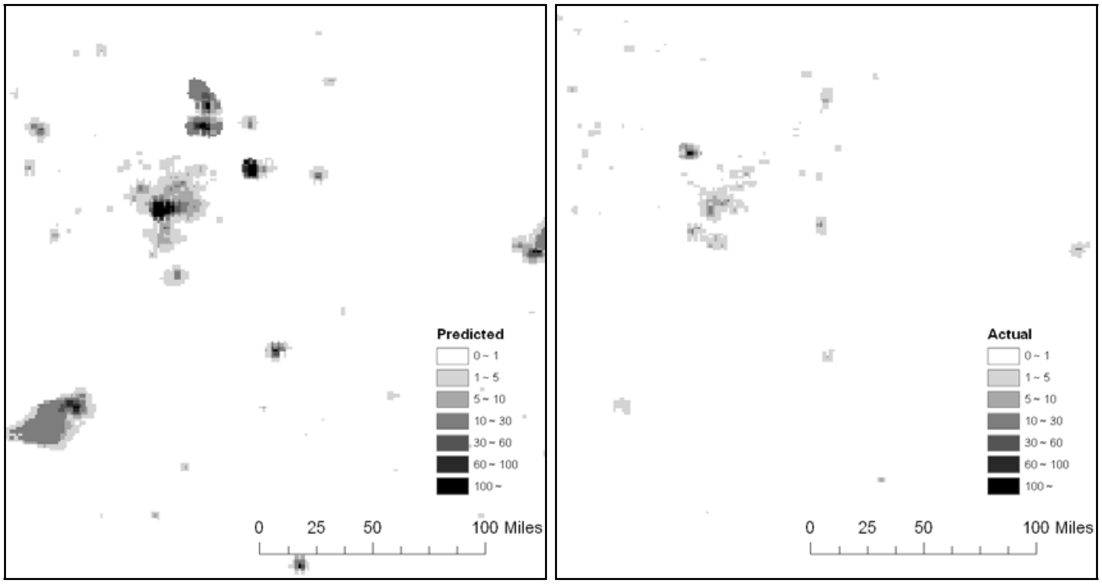

Figure 5.3. Predicted number of customers out (left plot) and actual number of customers out (right plot) in State C during Hurricane Katrina.

Table 5.1 shows the results of the hold-out analysis for State A. For this hold-out analysis, I first removed the data for a single hurricane (e.g., Katrina) from the data set, fit the model to the remaining data, used the fitted model to predict the number of customers without power in each grid cell during Hurricane Katrina, and then calculated the mean value of the absolute error between the actual number of customers without power and the predicted number of without power (the MAE). I repeated this process for each of the hurricanes for each state. Dividing the MAE by the mean number of customers without power yields an estimate of the relative error in the predictions from the model. These are typical of the results for the other states, which are not shown for brevity. Testing the predictive accuracy of the models that utilize the hurricane indicator variables is more challenging. The same hold-out analysis procedure is used in this section as was used in Section 4. That is, the predictions were setting each of the other hurricane indicator variables equal to one, and then these predictions were averaged. From Table 5.1 I see that the error in the estimates of the average number of customers without power varies from one to 14 times the average number of customers without power. The results show that the models generally provide reasonably accurate predictions except for Hurricane Danny, the weakest hurricane in the data set. For Hurricanes Georges and Dennis, the model prediction errors in each grid cell are at most two times to the average number of customers out. For Hurricanes Ivan and Katrina, the prediction errors are around approximately equal to the average number of customers without power. The model based on the alternate hurricane descriptors does seem to provide similar predictions for hurricanes not in the fitting data set as the model based on hurricane indicator variables together with the ad hoc assumption that a future hurricane is like the average of the past hurricanes. Overall, the results suggest that the customers out prediction model can provide the type of information needed to help guide state-wide hurricane preparation. My results show that for a strong hurricane such as Hurricane Katrina and Ivan, the model is a good predictive model for those areas outside of the urban areas, and the hold-out analysis results suggest that the model accuracy is good for most hurricanes. The model is more useful in making comparisons between different portions of the state than for comparing precise customers out estimates from small grid cells immediately adjacent to one another. This is appropriate given that the model is intended to help guide state-wide resource allocations rather than to provide very precise predictions for small, local areas.

Table 5.1. Predictive accuracy of the statistical models for hold-out samples in State A.

	Danny (1997)	Georges (1998)	Ivan (2004)	Dennis (2005)	Katrina (2005)
Actual number of customers out	72,646	326,392	1,244,445	447,966	998,292
$\mu_{\text{Customers Out}}$	10.87	48.85	186.3	67.05	149.4
$\frac{\text{MAE}_{\text{Hurricane indicator variables}}}{\mu_{\text{Customers Out}}}$	9.162	2.237	0.9441	2.042	1.019
$\frac{\text{MAE}_{\text{Alternative hurricane descriptors}}}{\mu_{\text{Customers Out}}}$	13.65	1.860	1.000	2.143	0.9336

5.4 Relative Importance of Explanatory Variables

Figures 5.4, 5.5, and 5.6 show the relative rate of change of the predicted mean number of customers out with respect to changes in the different explanatory variables: transformers, poles, switches, miles of overhead line, miles of underground line, number of customers served, windspeed, duration of strong winds, FSMs, MAP, SPIs, and the land cover variables for each of the states for the models that include the hurricane indicator variables and alternate hurricane descriptors. As shown in Figure 5.4, the relative impacts of the land cover variables, MAP, some of the FSMs, and the miles of underground line are lower than the other variables. This indicates that these variables do not have a strong influence on the predicted number of customers without power in State A. As expected, the wind speed covariate has a statistically significant and positive effect on the predicted number of customers without power. This was the same case for State B and State C as shown in Figures 5.5 and 5.6. In Figures 5.4, 5.5, and 5.6, I see that some of the FSM and SPI variables have a strong and statistically significant impact on the predicted number of customers without power. Looking across all three states, I see that most of the variables that measure the amount of overhead power system components in a grid cell while the LC (land cover) covariate have not a statistically significant impact on the predicted number of customers without power. While the relative magnitudes of the impacts of these parameters vary across states, the general conclusion is that have more overhead components leads to higher numbers of customers out during hurricanes, as would be expected.

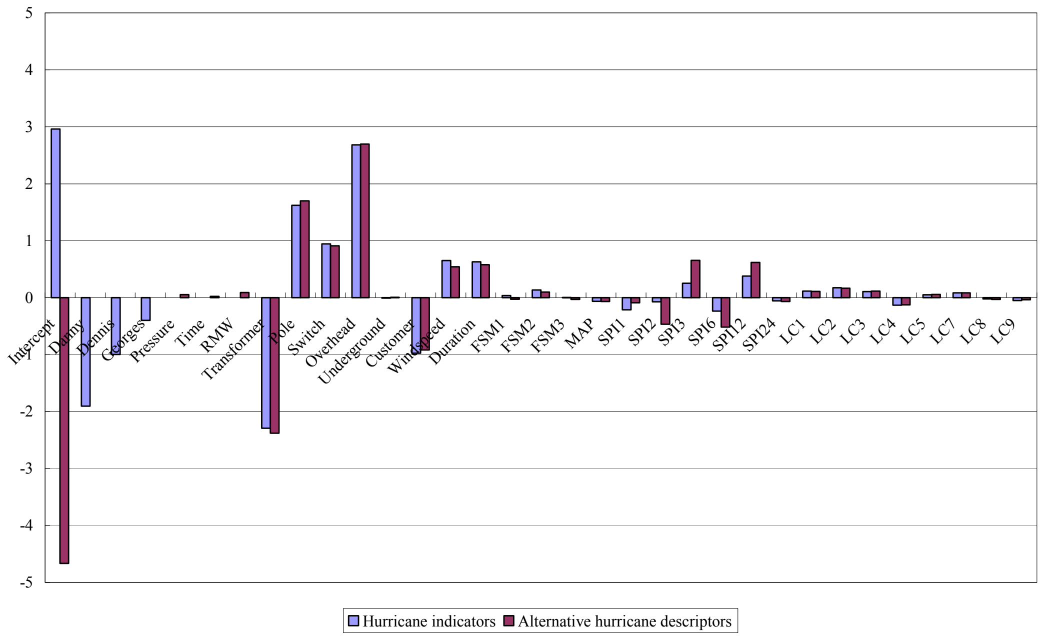

Figure 5.4. Relative effects of fixed effects, hurricane indicators and alternate hurricane descriptors of the final customers out prediction models for State A.

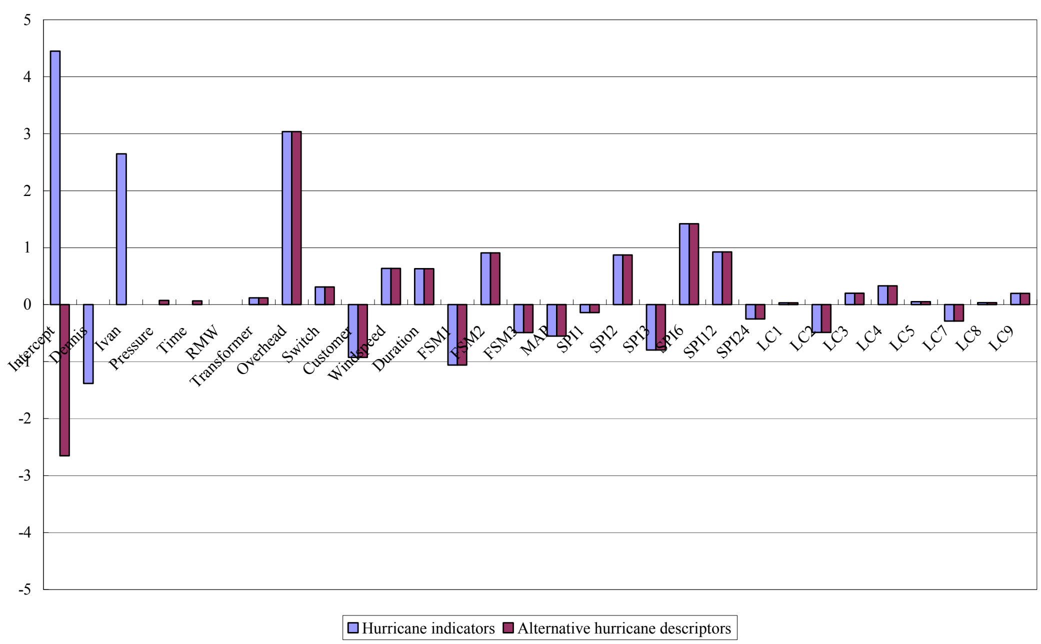

Figure 5.5. Relative effects of fixed effects, hurricane indicators and alternate hurricane descriptors of the final customers out prediction models for State B.

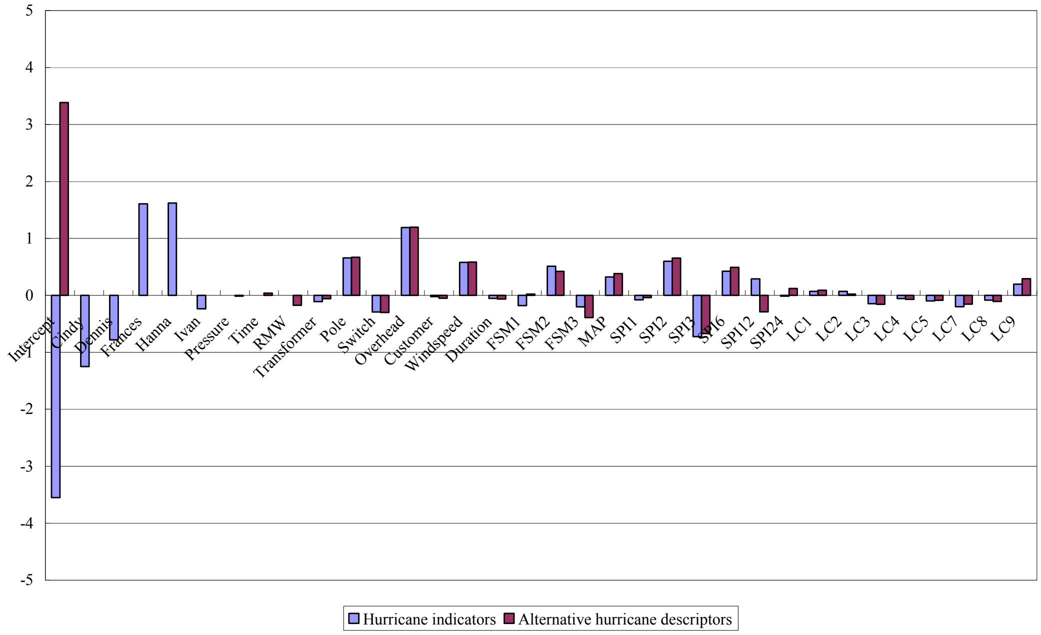

Figure 5.6. Relative effects of fixed effects, hurricane indicators and alternate hurricane descriptors of the final customers out prediction models for State C.

6. STATISTICAL DAMAGE ESTIMATION MODEL

If possible, it would be helpful to have estimates of the amount of actual damage to power distribution systems during hurricanes. For example, estimates of the number of damaged poles and transformers at the grid cell level could enable a utility company to target pre-hurricane resource allocations to those areas most likely to experience high levels of damage. However, developing these damage estimates poses a significant research challenge. No rigorous statistical methods have been reported in the literature for this problem, and there has been only limited detailed damage data available in the past. However, the damage data provided for portions of the State A service area provide a starting point for developing statistical models for estimating damage to poles and transformers at the grid cell level. It should be emphasized that while this data provides a good starting point, it is imperative that more complete damage data be collected for future hurricanes if accurate statistical damage estimation models are to be developed. This section summarizes the models I have developed and discusses their limitations and application.

6.1 Initial Damage Model Fit Results

The data set provided for State A contains the number of poles and transformers damaged in past hurricanes for limited portions of the service area. This damage data is aggregated to much larger areas than the grid cells used in the customer outage model. These damage aggregation areas were irregularly shaped and overlapped a number of smaller grid cells. In some cases the larger data aggregation areas overlapped as many as 224 small grid cells. Due to level of aggregation of the damage data, I assumed that the *rate* of damage of poles and transformers, given as the number of damaged poles or transformers divided by the total number of poles and transformers (treating each separately), was constant throughout the large grid cells. This allowed me to scale down to the smaller grid cells, making use of all of the detailed explanatory variables available at the level of the small grid cells. However, assuming that the damage rate is constant across the aggregated area is a strong assumption. Better damage data is needed for future hurricanes to help overcome this limitation in the current model. The data was scaled down to smaller grid cells by first calculating the total number of poles and transformers in each of the data aggregation areas by summing over the smaller grid cells that were included in each of the larger damage aggregation areas. Then the *rate* of pole and transformer damage was calculated by dividing the total number of damaged poles or transformers by the total number of poles or transformers in the data aggregation areas. Then these two damage rates were assumed to be constant, and the number of damaged poles and transformers in each of the original grid cells was estimated by multiplying the pole or transformer damage rate by the number of poles or transformers in each of the smaller original grid cells. Negative binomial GLMs were then developed for predicting the rate of damaged poles and damaged transformers at the level of the small grid cells in the same way as they were for the customers out models. However, unlike the customers out prediction models, an offset (the number of poles or transformers in each grid cell) is included in the link function to estimate the mean number of poles and transformers damaged based on the estimated damage rates and the total number of poles or transformers in each grid cell. The predictions from the damage models are the number of poles and the number of transformers damaged in each of the small grid cells for State A. This leads to variability in the predictions that is not present in the original data set which included information only at the level of the larger data aggregation areas. This variability will be discussed further below.

In the models for predicting the number of poles and transformers damaged, I found that there is considerable overdispersion in these data sets and that a Poisson GLM is not appropriate for predicting either the number of poles damaged or the number of transformers damaged. A negative binomial GLM is likely a better model. I fit the negative binomial GLM to the damage data. The deviances of the models with all covariates are 2,489 on 2,173 degrees of freedom for the damaged poles model and 2,556 on 2,173 degrees of freedom for the damaged transformers model, suggesting that the models may fit the data. There is a remarkable decrease in deviance relative to the

Poisson GLM. This suggests that the negative binomial model accounts for the extra variability in the damage data better than the Poisson GLM does. However, these models were based on data without conducting a PCA. Some collinearity exists, and a PCA is needed to account for this.

6.2 Negative Binomial Damage Model Fit Results

I conducted a PCA using the covariates from the damage models and refit the negative binomial GLM to the transformed data. As with the customers out prediction models, the principal components with p-values larger than 0.05 were iteratively removed and model comparisons were done by likelihood ratio tests between the different negative binomial GLMs with principal components. The full details of all of the model fits are given in the tables in Appendix B. The deviances of the final models with principal components are approximately the same as the deviances of the final models with correlated variables. Also, comparing the deviances of the final models with hurricane indicator variables to the deviances of the final models with alternative hurricane descriptors, there is little difference between the deviances (2,487 on 2,181 degrees of freedom for the pole damage estimation model and 2,557 on 2,181 degrees of freedom for the transformer estimation model). In addition, the number of covariates of the final model with principal components is approximately the same as the number of covariates of the final model with correlated variables for damaged poles (19 covariates with principal components and 20 covariates with correlated variables for the damaged poles estimation model and 20 covariates with principal components and 20 covariates with correlated variables for the damaged transformers estimation model).

While the damage data provided for State A is the best power system damage data I have seen for hurricanes, there are still limitations due to the aggregation in the data and the limited geographic area covered by this data. Overall, the results from the models that I have fit to the limited damage data available suggest that it may be possible to obtain reasonably accurate estimates of damage to poles and transformers during hurricanes. It is hard to show the spatial distribution of damage prediction due to limited data. Still, it would prove useful to conduct a trial run of this model fit based only on the available data on a limited scope in a future storm to see how well it predicts damage outside of the areas from which data was used to fit the model. One approach to gathering to this data would be to develop a statistically rigorous sampling plan under which a portion of the system elements in some or all of the grid cells were inspected after a hurricane. With proper sampling, the recorded damage data could be used to generalize to develop system-wide damage estimates for future hurricanes.

6.3 Relative Importance of Explanatory Variables

Figures 6.1 and 6.2 show the relative rate of change of the predicted mean number of damaged poles and transformers with respect to changes in the different explanatory variables: transformers, poles, switches, miles of overhead line, miles of underground line, number of customers served, windspeed, duration of strong winds, FSMs, MAP, SPIs, and the land cover variables for each of the states for the models that include the hurricane indicator variables and alternate hurricane descriptors.

In examining the results from the damage models shown in Figures 6.1 and 6.2, the variables that have the strongest impact on the predicted amount of damage are the maximum gust wind speed (positive impact), FSM3 (fractional soil moisture in the deepest layer – negative impact), SPI1 (negative impact), SPI2 (positive impact), SPI3 (negative impact), SPI12 (positive impact), and SPI24 (negative impact). Higher wind speeds tend to increase the amount of damage during hurricanes and higher soil moisture at the deepest layers tends to decrease the amount of damage according to this model. At the same time, the impacts of moisture availability are mixed depending on the time-frame of interest. High values for the longest-term (24 month) moisture availability variable tend to decrease the amount of damage, while values for the shorter moisture availability variables are mixed. The implications of the moisture availability variables for different time frames is not clear, though it is clear that they are having a strong impact on the predicted amount of damage.

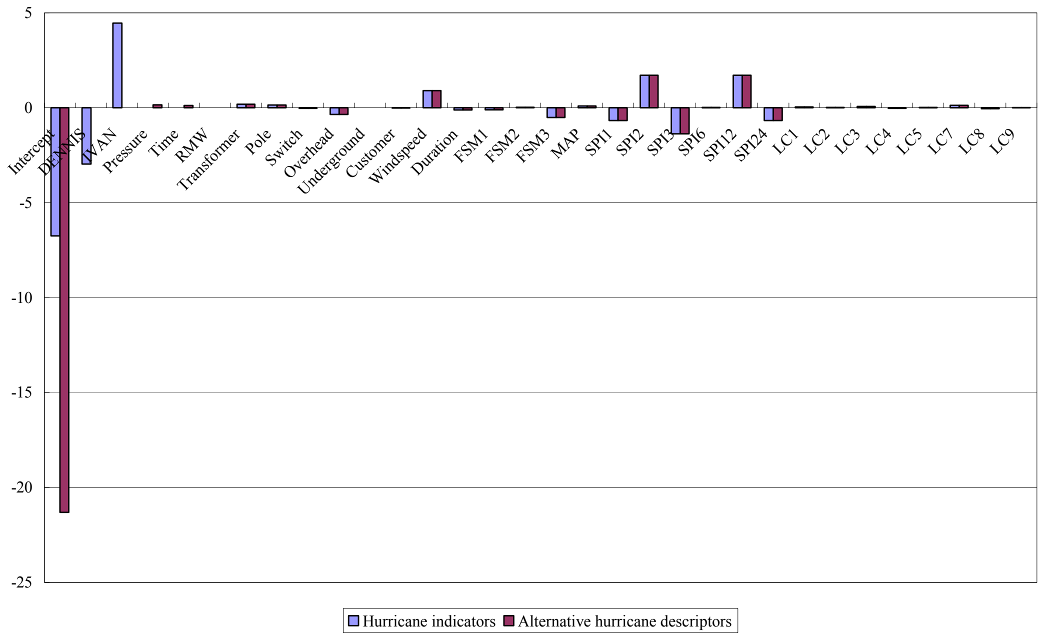

Figure 6.1. Relative effects of fixed effects, hurricane indicators and alternate hurricane descriptors of the final damaged pole prediction models for State A.

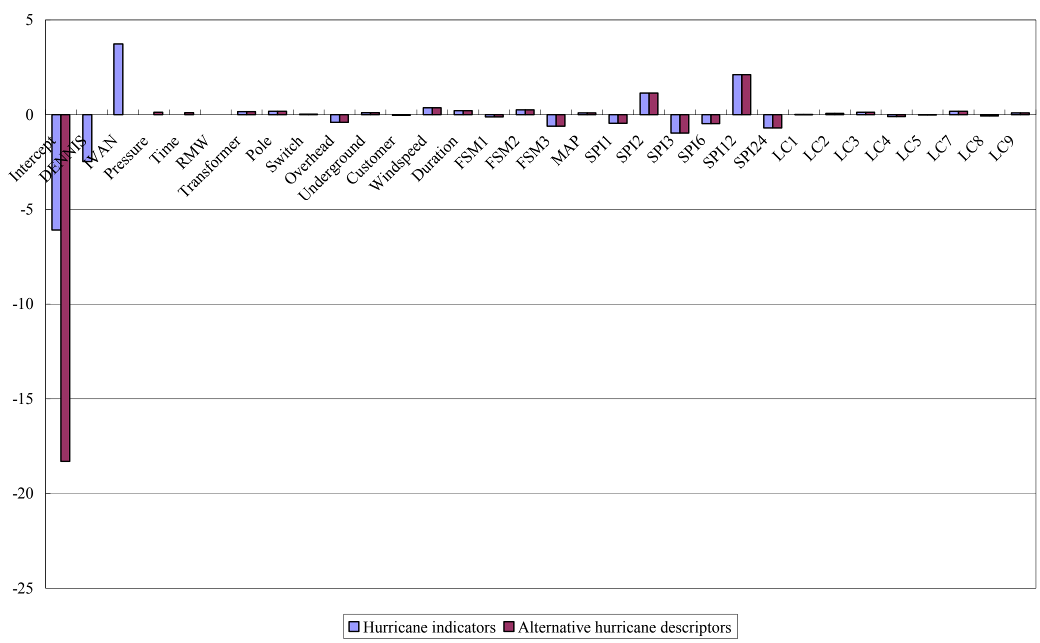

Figure 6.2. Relative effects of fixed effects, hurricane indicators and alternate hurricane descriptors of the final damaged transformer prediction models for State A.

7. PHYSICAL DAMAGE ESTIMATION MODEL

This section focuses on the power distribution system because the vast majority of damage during hurricanes occurs in the distribution system. The models developed in this section predict the damage in the power distribution system. Unlike the statistical damage estimation models discussed in Section 6, physical damage estimation models need geometry, material properties, and loading conditions of the distribution system. Due to the limited amount of detailed data that is available about failures in the distribution system, plausible overhead power line structures which can represent the system were developed for use in this section by following the appropriate codes. These representative systems were then used as the basis for developing damage estimation models. Pole geometry and strength information were derived from the American National Standard Institute (ANSI) O5.1 (ANSI 2002). The National Electrical Safety Code (NESC 2007) establishes overload requirements (Rule 250) for the overhead lines in the power distribution system. Also, ACSE 7-05, "Minimum design loads for buildings and other structures" was considered as a reference standard so as to meet wind load provisions for buildings and other structures. Based on these codes, representative overhead power line structures were developed for the case study used by the damage estimation model. Then the damage on the power line structures was predicted by developing damage estimation models. Based on the damage estimation model, the number of damaged poles could be predicted for future hurricanes.

There has been one previous published study that used structural reliability models to estimate damage to power distribution system poles, the Caribbean Disaster Mitigation Project (1996). The Caribbean Disaster Mitigation Project included hurricane hazard modeling that accounted for the effect of hurricane-related wind together with a structural analysis of the poles in the power distribution system. However, the structural analysis model used by the Caribbean Disaster Mitigation Project (1996) considered only flexural damage to poles under wind loads, not foundation failures. Foundation failure is a significant failure mode during hurricanes because the power distribution system poles can fall by losing the resistance of the foundation due to wet soil conditions. Anecdotal evidence and pictures of hurricane damage suggest that foundation failures do cause at least some pole failures during hurricanes. In this study, fragility curves for utility poles in the power distribution system were developed by using structural reliability methods in combination with Bayesian updating based on limited observed damage information. These damage estimation models were used in conjunction with the hurricane wind field model to estimate pole damage in the distribution system.

Due to the lack of the detailed data for the power distribution system, I can not consider all possible failure mechanisms. In particular, I have not included failures due to trees falling onto lines or poles and damage due to wind-blown debris due to a lack of data. Fortunately, damage data are available for a few hurricanes: Dennis (2004), Ivan (2005), and Katrina (2005). If the actual damage information can be integrated with the information from physical damage estimation models, this would provide better fragility estimates and a better understanding of the uncertainty inherent in physical damage estimation models. This integrated approach should also provide more reliable damage predictions for future events by integrating observed system performance with structural reliability models. Bayesian methods are appropriate for this integration process based on limited data. They produce updated prediction models for future events that account for both the structural reliability model and the observed data. This section develops both a structural reliability model for poles and a Bayesian approach for predicting the number of damaged poles based on both the physical damage estimation model and the observed data. Finally, fragility curves for the poles in a representative distribution system are presented and the number of damaged poles in the case study service area is predicted using these fragility curves.

7.1. *Fragility of the Power Distribution System by Structural Reliability Methods*

7.1.1 Power distribution system failure

In evaluating the reliability or probability of failure of a system, one must account for the fact that the system can often fail due to more than one failure mechanism. In other words, the probability of failure of the power distribution system can be defined by individual failure mechanisms such as trees falling on lines or poles, wind-born debris striking lines or poles, as well as severe wind causing pole failures directly. In this study only two failure mechanisms were considered: (1) flexural failure of poles due to wind and (2) foundation failure of poles due to wind. While the other failure modes (e.g. tree-induced failures) likely play a significant role in terms of overall system reliability, the focus of this section is on only direct wind-induced failures. I did not address other wind-induced failures such as trees and debris falling on or being blown into poles and lines due to high winds. I have also not addressed failures due to other hazards such as inland flooding or storm surge along the coast. The model developed in this section is an important first step in developing a model for estimating damage in power distribution systems during hurricanes, but future work is needed to develop a complete model. Future work can build from this starting point to include additional wind-induced failure modes and failure modes induced by other hurricane-related hazards such as flooding and storm surge. With $P(E_i)$ representing the probability of failure of the i^{th} failure mechanism, the probabilities of the individual failure modes can be defined by

$P(E_1) = P(\textit{flexural failure} \mid V)$: conditional probability of a pole breaking due to a bending moment induced by wind speed V

$P(E_2) = P(\textit{foundation failure} \mid V)$: conditional probability that the soil that the pole is planted in loses strength given wind speed V

Assuming that the two failure events are statistically independent, the probability of failure of the power distribution system is

$$p_{fs} = 1 - \prod_{i=1}^{n} \left[1 - P(E_i)\right] \tag{7.1}$$

and the cumulative density function is

$$F_{system}(fail \mid V) = \prod_{i=1}^{n} \left[1 - P(E_i)\right] \tag{7.2}$$

While the assumption of independence is not strictly speaking correct given that the formulation of the limit state functions involves common random variables, it facilitates the analysis for the prior which can be simply obtained. Furthermore, assumption of independence is really one of conditional independence here: the two failure modes are assumed to be conditionally independent given wind speed. While there may be still be sources of dependence (e.g., span length appearing in both failure mode equations, inducing a dependency), the assumption of conditional independence is a reasonable first approximation. For evaluating the probability of failure for the individual failure modes, first-order reliability methods (FORM) were used because the limit state function of each failure mechanism is linear and the random variables (e.g. modulus of rupture of poles, moment of resistance of soil, and span length of the distribution system) are uncorrelated. Specifically, the advanced first-order second-moment (AFOSM) method was used in order to include non-Normal random variables. The limit state function is described in detail below. The AFOSM requires the determination of the design point (e.g. the point of minimum distance to the limit state function). Because some algorithms may fail to converge to find the design point, the improved HL-RF algorithm (Zhang and Der Kiureghian 1994) was used in this study. Using this AFOSM approach, the probability of failure of a single pole as a function of wind speed was estimated. The fragility of a power distribution system pole is defined in this section as the conditional probability of the pole failing given a specified 3-sec gust wind speed. Using the fragility developed for individual poles, the number of damaged poles affected by hurricane-related wind can be predicted by mapping pole locations which can facilitate simulation of the event for model evaluation. Then the predicted number of damaged poles is directly compared with the data provided by the utility company.

The limit state function for each of the individual failure modes for the reliability analysis is defined as

$$g(x) = g(X_1, X_2, \cdots\cdots, X_n) = R - W \quad (7.3)$$

where $X_1, X_2, \cdots\cdots, X_n$ are random variables, R is the resistance capacity for the individual failure mode and W is the wind load. The resistance capacity and geometry information of the power distribution system is obtained from the ANSI standard O5.1 classification of pole structures. The wind load was calculated using wind pressure provision in NESC 2007 and the 3-sec gust wind speed obtained from the hurricane wind field model discussed in Section 3.

Because of the limited amount of detailed data available about power distribution systems, plausible overhead power line structures were used to represent power distribution systems. From ANSI O5.1, three types of utility poles were considered for this study: Southern pine, Douglas-fir, and Western red cedar. A 34.5 kV transmission line and a 12.47 kV distribution line were used for each of pole types because the power distribution system is typically composed of 2 types of lines. Span length for the two types of lines and height for a utility pole were obtained from the Caribbean Disaster Mitigation Project (1996) where they used a mean span length of 144 ft and a variance of 36.7 ft^2 for 12.47 kV line, a mean of 341 ft and a variance of 85 ft^2 for 34.5 kV lines, and height of 45 ft. Based on the results of Keshavarzian and Priebe (2002) which says the effect of pole height variations between 45 ft and 60 ft, the range of heights generally used in power distribution systems, is negligible on the wind loading calculation, a 45 ft tall utility pole was used as a baseline structure for the reliability analysis. The utility pole is planted 6.5 ft deep in the ground following ANSI O5.1. The loading condition and dimension are shown Figure 7.1. Note that in developing the example system, design standards are being used to represent as-built conditions. This is a strong assumption. As-build conditions often differ substantially from design specifications, and there is often considerable variation in actual pole conditions throughout a large power distribution system. However, detailed information about as-built conditions is not available. In this section, I use design standards to represent the actual system as an approximation. This is in line with the goal of the structural reliability model. This model is intended to provide a sound basis for the prior for the Bayesian analysis, not to provide a highly accurate, system-wide reliability model on its own. While creating an accurate, system-wide reliability model based on structural reliability analysis methods is desirable, the data needed to do this is currently not available. However, enough information is available to support the development of priors for a Bayesian analysis of pole reliability.

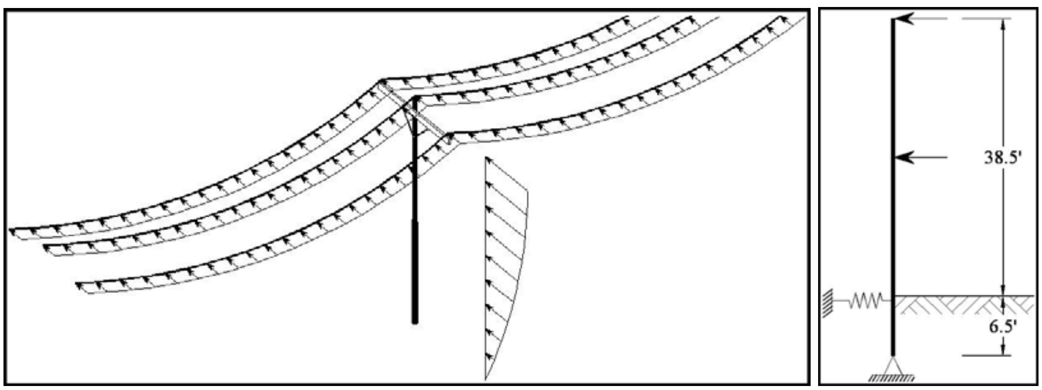

Figure 7.1. Loading condition and dimension of a baseline structure.

7.1.2. Flexural failure

When a pole structure is subjected to a wind load, the wind pressure acts on the conductors and the pole, causing a base bending moment at the ground line of the pole. The inner fibers of the pole are compressed and the outer fibers are extended due to the base bending moment. If the tensile stress of the extended outer fiber exceeds the maximum rupture stress, then the pole will fail. The limit state function of the flexural failure mode is

$$g(x) = R - W = \sigma_r - \sigma_{groundline} = \sigma_r - \frac{M_{groundline}}{Z} = \sigma_r - \frac{32\,M_{groundline}}{\pi D^3_{pole}} \qquad (7.4)$$

where R=resistance capacity, W=wind load, σ_r=mean modulus of rupture (MOR) of the pole, $\sigma_{groundline}$ =tensile stress of the pole at groundline, M=bending moment at groundline, Z=modulus of section, and D=diameter of the pole at groundline. Table 7.1

shows the mean modulus of rupture (MOR) and the coefficient of variation (COV) for 3 types of poles (ANSI O5.1 2002).

Table 7.1. Groundline strength for less than 50 feet long poles, used in unguyed, single-pole structures only.

Species	MOR (<50ft)	
	Mean (psi)	COV
Southern pine	10190	0.169
Douglas-fir: Coastal	9620	0.135
Western red cedar	6310	0.204

For the calculation of force due to extreme wind loading, the NESC suggests the following equation (NESC 2007).

$$P = 0.00256(V_{mi/h})^2 k_z G_{RF} I C_f A \tag{7.5}$$

where P=wind load in pounds, V=3-s gust speed in m/s at 10m above ground, k_z=velocity pressure exposure coefficient (Rule 250-2), G_{RF}=gust response factor (Rule 250C2), I=importance factor (Rule 252B), C_f=shape factor (Rule 252B), and A=projected wind area in ft^2. Equation 7.5 is assumed herein to provide the actual (deterministic) wind force. This assumption, while not strictly speaking correct, facilitates the simple FORM analysis. All coefficients except wind speed V are assumed to be deterministic, but this might not be a problem if the uncertainty in the random variables that are considered in the limit state functions dominates in the wind load calculation. Moreover, because the purpose of the structural reliability model is to provide a solid prior for the final integrated model, not the final model itself, the assumption that equation (7.5) provides the wind force is an acceptable approximation. The values for the deterministic coefficients and factors for the pole and conductors are found in Table 7.2 (NESC 2007).

Finally, the bending moment due to the wind load P for the baseline structure as shown in the right figure of Figure 7.1 is

$$M_{groundline} = 0.00256(V_{mi/h})^2(0.97 \times 38.5 \times D_{pole} \times 20.25 \\ + 3 \times 1.1 \times 0.88 \times L_{span} \times 7/16/12 \times 39.5) \quad (7.6)$$

In the final limit state function, σ_r and L_{span} were treated as random variables shown in Table 7.1 and Section 7.1.1.

Table 7.2. Parameter values for an extreme wind calculation (NESC Rule 250C).

Extreme wind pressure on the pole	Extreme wind pressure on the wires
K_z structure = 1.0 (35ft<Height<50ft) G_{RF} structure = 0.97 (35ft<Height<50ft) I = 1.0 for utility structures C_f = 1.0 for cylindrical structures	K_z wire = 1.1 (35ft<Height<50ft) G_{RF} wire = 0.88 (35ft<Height<50ft) I = 1.0 for utility structures C_f = 1.0 for cylindrical shapes

7.1.3 Foundation failure

Though a flexural failure is a primary failure mechanism for power distribution system poles, a foundation failure is also a critical failure mechanism for power distribution system poles. In order to find out how well the power distribution system can withstand extreme winds, we must consider the resistance of the foundation that the pole is planted in. It is known that a pole pivots about a point below ground level (Wareing 2005). A foundation failure is defined in this section to occur if the moment at a pivot point $h/\sqrt{2}$ below under the ground is greater than the moment of the resistance of the foundation where h is the depth to which the pole is planted. The pivot point has zero stress assuming that the stress distribution is linear. The limit state function of the foundation failure mode is

$$g(x) = R - W = M_g - M_{@pivot} = \frac{kDh^3}{10} - M_{@pivot} \quad (7.7)$$

where $M_g = \frac{kDh^3}{10}$ = moment of resistance of the soil/foundation (Wareing 2005), $M_{@pivot}$=moment at pivot point, k=maximum rupturing intensity in lb/ft^2/ft, D=average diameter of pole below ground level in ft, and h=depth of planting in ft. The relation for

M_g given by Wareing (2005) assumes that a pole will pivot about some point below ground level and the stress distribution of the soil under the ground is linear with depth. The maximum rupturing intensity k was set at 2000 lb/ft^2/ft for average soil (Wareing 2005). Phoon et al. (1995) suggested that 0.32 is an appropriate value for the coefficient of variation (COV) of k for clay soil as an example of soil. The bending moment due to the wind load P for the baseline structure is

$$\begin{aligned} M = 0.00256(V_{mi/h})^2 (0.97 \times 38.5 \times D_{pole} \times (19.25 + h/\sqrt{2}) \\ + 3 \times 1.1 \times 0.88 \times L_{span} \times 7/16/12 \times (38.5 + h/\sqrt{2})) \\ - P_{horrizontal\ reaction} \times (h/\sqrt{2} - 1) \end{aligned} \quad (7.8)$$

where $P_{\text{horrizontal reaction}}$ = reaction force of the soil, which can be calculated analytically.

7.2. *Fragility of the Power Distribution System Using Bayesian Approach*

There are many approaches for estimating the probability of failure of various structures. One approach estimates the probability of failure through statistical methods like those used in Section 6 based on past failure data. However, estimating the probability of failure using the past data is quite difficult, especially for infrastructure systems for which little failure data is available. Another approach is based on structural reliability analyses such as first-order reliability methods (FORM) discussed above and second-order reliability methods (SORM). If the behavior of the structures being modeled is complex, it may not be possible to express the limit state functions in closed-form in such cases. Another approach for estimating the probability of failure is to use Monte Carlo simulation. It could be fairly simple to estimate the probability of failure for a certain structure but it would be computationally burdensome to repeat the simulation for a large number of grid cells such as the 6,000 grid cells in State A used in this study. Because it is necessary to estimate the probability of failure for state-wide areas, more reliable data-based and analytical approaches should be considered for estimating the probability of failure. The Bayesian approach is suitable for this kind of problem in which analytical results and limited data can be integrated.

The Bayesian approach is based on Bayes' theorem (e.g., Gelman et al. 2003). Bayesian probability theory starts from a prior probability distribution representing the initial information about a parameter. This is multiplied by the likelihood function based on the observed data and then normalized by the total probability of the data. The resulting posterior distribution is the conditional probability distribution for the uncertain quantity given the observed data. This is shown in equation (7.9).

$$posterior = \frac{likelihood}{total\ probability\ of\ the\ data}\ prior \tag{7.9}$$

The Bayesian approach can be used for predicting the number of damaged poles by updating the probability of pole failure estimated with the structural reliability model with the observed damage data for poles in the power distribution system. The formulation is defined as

$$f(p|f,t,V) = \frac{f(f,t|p,V)}{\int_x f(f,t|x,V) f(x|V) dx} f(p|V) \tag{7.10}$$

The parameter p is the probability of failure of a power distribution system pole given observed data consisting of f failed poles out of t poles under a wind speed of V. Considering p as the frequency of pole failure, $f(p|V)$ represents the prior probability density function of the probability of failure for poles under wind loads. The structural reliability model provides this prior. In estimating the posterior probability mass function (PMF) for the number of damaged poles in a given grid cell, a beta distribution is an appropriate prior as long as pole failures are assumed to be conditionally independent given wind speed. Pole failures are discrete, non-negative counts, and, barring additional information that could be used as additional conditioning variables, the probability of failure can reasonably be assumed to be constant across all poles. I will use the binomial PMF as my likelihood function. A convenient prior for the parameter p conditioned on wind speed is the beta distribution. This distribution is constrained to the (0, 1) interval, as are probabilities. The beta distribution is also highly flexible in its ability to model differing degrees of information and accuracy in the prior.

Finally, using the beta-binomial pair is also mathematically convenient because they form a conjugate pair, allowing the Bayesian updating to be done analytically.

Raiffa and Schlaifer (2000) show the general form of the Bayesian updating for conjugate prior-likelihood pairs. Conjugate priors allow Bayesian updating to be done analytically in a simple manner. A beta prior with a binomial likelihood is a conjugate prior-likelihood pair that is attractive for this problem. With a beta prior with parameters *f* and *t* and binomial likelihood for the observed data consisting of f' failures in t' trials, the posterior distribution is also a beta distribution with parameters f', t', *f*, and *t*, which are f failed poles out of t poles obtained from the structural reliability model. As shown by Raiffa and Schlaifer (2000), the posterior is given by

$$f(p|f',t') = \frac{\Gamma(t+t')}{\Gamma(f+f')\Gamma(t-f+t'-f')} p^{(f+f'-1)}(1-a)^{(t-f+t'-f'-1)} \quad (7.11)$$

The mean and variance of the posterior beta distribution are

$$mean = \frac{f+f'}{t+t'}, \qquad variance = \frac{(f+f')(t+t'-f-f')}{(t+t')^2(t+t'+1)} \quad (7.12)$$

This process produces the mean frequency of pole failure given a wind speed as well as the full PDF for the future frequency of pole failure given a wind speed. This updated fragility curve for the power distribution system given wind speeds can be used to predict the number of damaged poles under wind speeds for future hurricanes.

A key challenge in using Bayesian updating in this situation is selecting a prior based on the structural reliability model results. The structural reliability model provides an estimate of the mean probability of failure, not an estimate of the uncertainty (e.g., the variance). In order to compose a prior for a specific grid cell based on the results of the structural reliability model, an assumption must be made about the variance of the beta distribution used as the prior. I assumed that the prior mean is known from the structural reliability model. I also assumed that this mean is equal to the unknown number of failed poles, *f*, divided by the known total number of poles, *t*, in the grid cell. I estimated *f* by multiplying the mean from the structural reliability model for the estimated wind speed at the grid cell by the total number of poles in the grid cell. I then assumed that the prior is a beta distribution with $\alpha = f$ and $\beta = (t\text{–}f)$. The mean and variance of this beta prior are given by equations (7.13) and (7.14).

$$mean = \frac{f}{t}, \tag{7.13}$$

$$variance = \frac{f(t-f)}{t^2(t+1)} \tag{7.14}$$

Note that there are strong assumptions about the variance of the prior that are implicit in this approach. Specifically, using the approach outlined above implicitly assumes that the variance of the prior as a function of the two known parameters, the mean failure probability from the structural reliability model (μ) and the total number of poles (t) in the grid cell, is given by equation (7.15)

$$variance = \frac{t^2\mu - t^2\mu^2}{t^3 + t^2} \tag{7.15}$$

Equation (7.15) shows that I have implicitly assumed that as the prior mean increases, the prior variance increases and that the prior variance is lower for a given wind speed (and thus for a given prior mean) for grid cells with more poles. While these are strong assumptions, I will show below that with the data used in this dissertation, there is not a strong relationship between the prior variance and the prior mean due to relatively small variations in the mean.

In order to check the adequacy of the assumption made about the variance of prior distributions, Figures 7.2 and 7.3 give plots of the mean and variance of the priors and the posteriors respectively. In Figure 7.2, there is no clear pattern between the mean and the variance of the prior for the 3 hurricanes. This suggests that the implicit assumptions made about the variance of priors, e.g., that the variance of the prior increases with the mean of the prior and decreases with increases in the number of poles in the grid cell, do not significantly influence the posteriors in unintended ways. If there had been a clear trend between mean and variance in Figure 7.2, this would have suggested that the implicit assumptions about the variance might have unintended influence on the posteriors.

Figure 7.3 suggests that the posterior variance increases with the posterior mean for Hurricane Ivan. Figure 7.3 also shows that there is considerable uncertainty in the posteriors for the higher failure rates. The points in Figure 7.3 with posterior mean values above 0.1 are from the observed data for one of the sampling areas in which damage data was collected for Hurricane Ivan, and these data points are investigated further in Figures 7.8 and 7.9. Based on the results of the mean and variance plot, I could conclude that the prior distribution, a beta distribution with the assumptions discussed above, is acceptable because the variance of the prior does not increase with the mean of the prior for my data set. For other data sets, one must carefully investigate the relationship between the prior mean and the prior variance to make sure that the assumptions implicit in this type of prior are not influencing the prior in unintended ways.

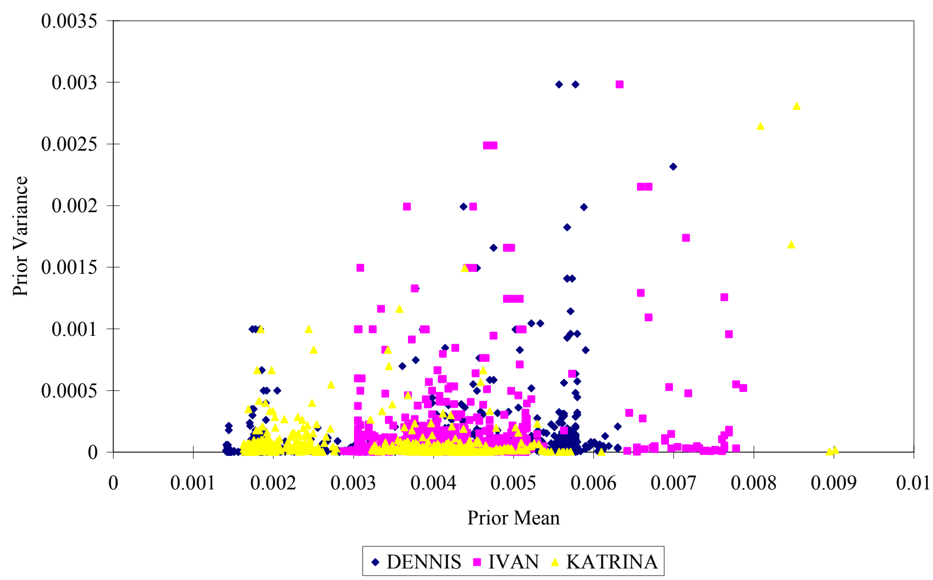

Figure 7.2. Mean and variance of priors for 3 hurricanes.

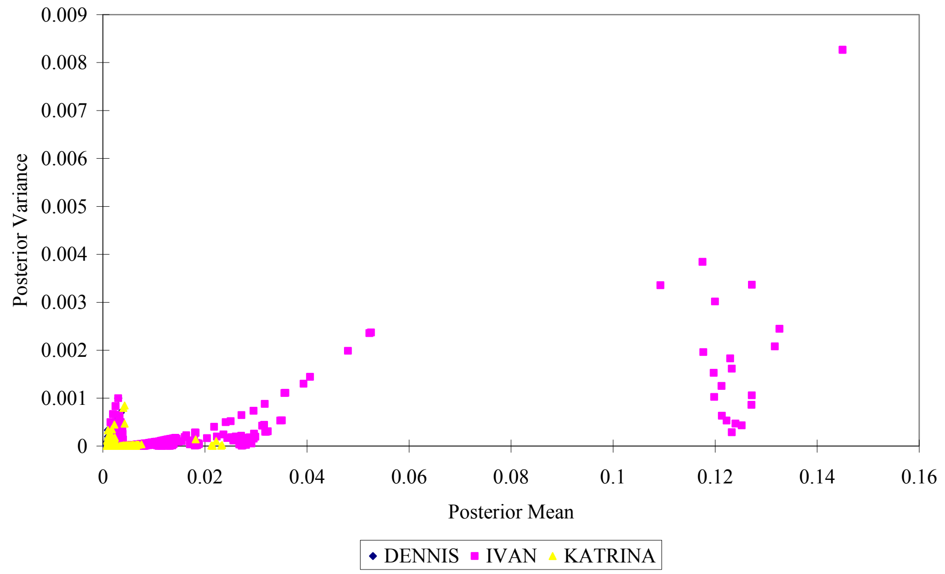

Figure 7.3. Mean and variance of posteriors for 3 hurricanes.

Based on the Bayesian model with the assumptions about the prior discussed above, the overall Bayesian updating process I used for a given type of pole and span length (to be discussed further below) was:

1. I used the structural reliability model to estimate the probability of pole failure in each grid cell based on the estimated wind speed in that grid cell.
2. I used the mean failure probability from the structural reliability model to estimate the mean number of pole failures in each grid cell by multiplying the mean probability by the number of poles in the grid cell.
3. I estimated the parameters of the beta prior distribution (α and β) using the approach discussed above.
4. I updated the prior found in the third step with the observed failure data based on Equation 7.11. This gave the posterior for the number of failed poles in a gird cell for the estimated wind speed in that grid cell.
5. I used Equation 7.12 to estimate the posterior mean fraction of poles failed and the variance of the posterior in each grid cell with the parameters of the beta prior distribution (t and f) and the observed failure data (t' and f').

6. I fit the posterior fragility curve to the posterior mean fraction of poles failed with a normal CDF estimating the mean and the variance of the posterior fragility curve.
7. Based on the posterior mean fraction failed and the variance of the posterior, the lower and upper bounds were also calculated with 95 % asymptotic confidence using the following equation:

$$\text{Confidence interval} = (\text{posterior mean}) \pm 1.96(\text{posterior variance}). \qquad (7.16)$$

7.3. *Physical Damage Estimation Model Results*

Fragilities as a function of wind speeds were evaluated using the AFOSM reliability method. Figure 7.4 shows fragility curves for 3 types of poles for both types of line systems (e.g. both span lengths). The dotted lines represent 12.47 kV distribution systems (short span length) and the solid lines represent 34.5 kV transmission systems (long span length) for the 3 types of poles. Overall, the fragility curves for 12.47 kV line are located to the right in the figure, which means that the probability of failure for 12.47 kV line is lower than the probability of failure for 34.5 kV line. In other words, the probability of failure of the power distribution system is governed by the span length of the system, not the types of poles. This makes sense. Longer span lengths for a given number of poles allow a longer length of line per pole, increasing the effective loading per pole. However, there are still differences, even given a span length.

With the fragility curves, the number of damaged poles can be estimated for each of the distribution systems (e.g. 3 types of poles for both types of line systems). The number of damaged poles is calculated by multiplying the probability of failure of a power distribution system pole for the wind speed experienced in a grid cell (from the structural reliability model) and the total number of poles in the grid cell together. Figures 7.5, 7.6, and 7.7 show the number of damaged poles from structural reliability analysis and observed number of damaged poles for the 5 areas in State A in which damage data was collected for Hurricane Dennis, Ivan, and Katrina respectively. Note that this is a limited data set, and the areas for which damage information was collected are not necessarily representative of the entire service area. However, this is the only damage data available from the utility that provided the data for this analysis. The thicker line represents the number of observed damaged poles during each of the hurricanes and the thinner lines represent the number of damaged poles predicted if the power distribution system were to be entirely composed of on one type of (identical) pole and span length. These figures show that the number of damaged poles from the observed data differs substantially for the three hurricanes even though all of the hurricanes were Category 3 hurricanes when they made a landfall. For Hurricane Dennis and Hurricane Ivan, the observed number of damaged poles is not in between the expected number of damaged poles estimated by fragility curves for the pole types and line systems used while it is for Hurricane Katrina. Overall, the observed number of damaged poles is not within the range predicted by the structural reliability model for the 6 pole-span length combinations. These results suggest that the physical damage estimation model alone does not provide enough accuracy for predicting the number of damaged poles. Integrating the fragility curves based on the structural reliability analysis with the observed data is needed for improving the accuracy of the physical damage estimation models for future hurricanes.

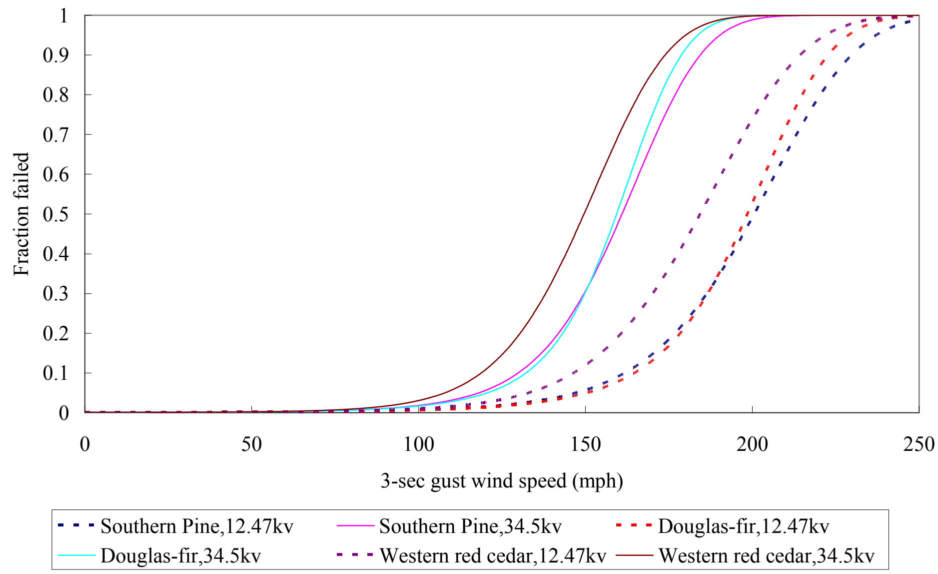

Figure 7.4. Fragility curves given wind speeds for various pole types by structural reliability analysis.

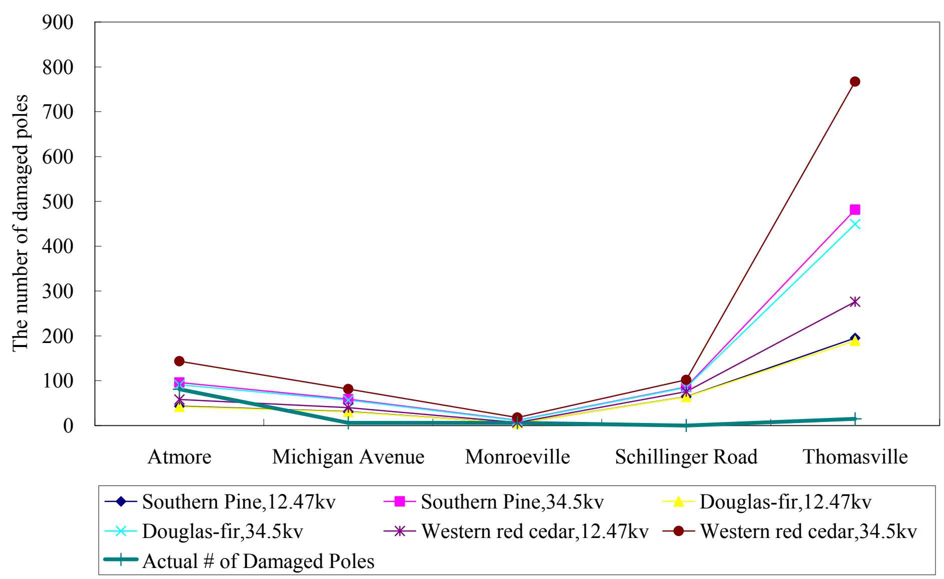

Figure 7.5. The number of damaged poles from structural reliability analysis and observed data for Hurricane Dennis.

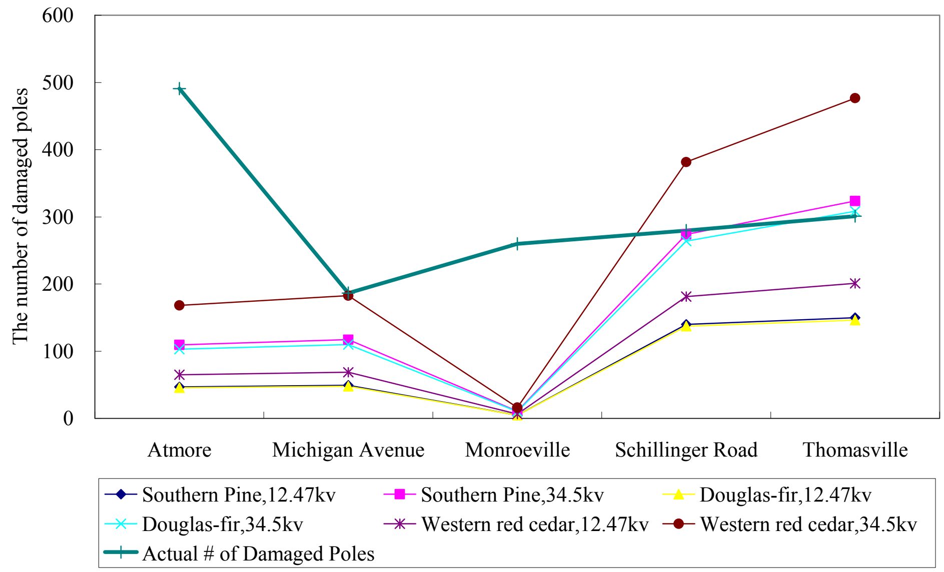

Figure 7.6. The number of damaged poles from structural reliability analysis and observed data for Hurricane Ivan.

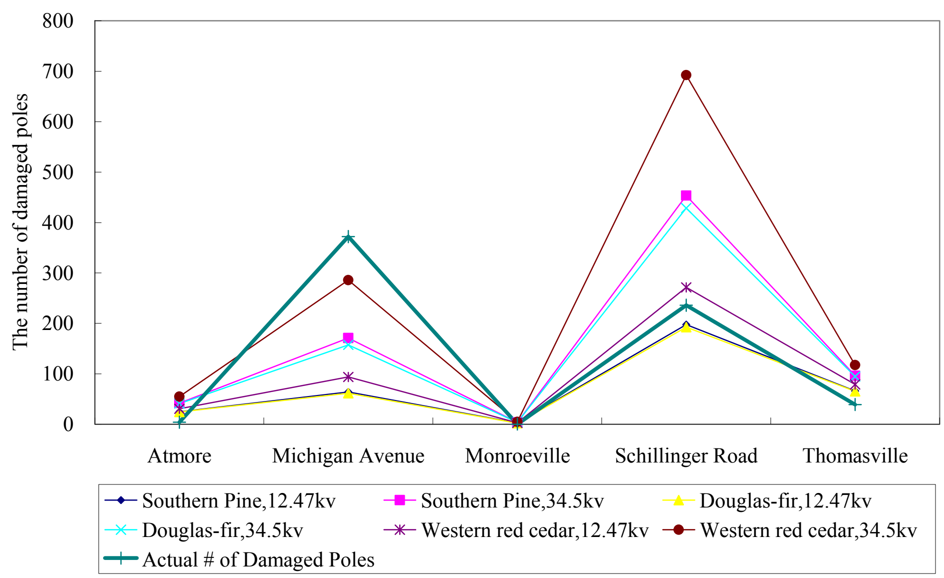

Figure 7.7. The number of damaged poles from structural reliability analysis and observed data for Hurricane Katrina.

As mentioned in Section 7.3, Bayesian updating provides a way to integrate the results of the structural reliability model with the observed data. Using the Bayesian approach with conjugate pairs, the priors based on the fragility curves from the structural reliability analysis were updated with the observed data. For priors for Southern Pine poles, the main type of utility pole used in the service area of the utility company that provided the data for this thesis, the results for 12.47 kV line and 34.5 kV line are shown in Figures 7.8 and 7.9 respectively. In Figures 7.8 and 7.9, points are the posterior mean probability of failure for 3 hurricanes (one point per grid cell), the dotted lines are the priors for Southern Pine poles (12.47 kV in Figure 7.8 and 34.5 kV in Figure 7.9) based on the structural reliability model, , and the solid line is a smoothed fit (a normal CDF fit) of the mean probability of failure from the posterior distribution. Based on the mean and variance of posterior distributions, the lower and upper bounds were also calculated with 95 % asymptotic confidence coefficients (e.g., a student's t distribution was used to estimate the 95% confidence interval using the posterior mean and variance) and presented in Figures 7.8 and 7.9. The "fraction failed" measure used on y axis represents the percentage of failed poles in a grid cell that the model estimates will fail. The number of pole failures for each type of pole was calculated by assuming all of the poles are identical. If additional information about poles (e.g., the fraction of poles and failed poles with different geometries, sizes, transformer attachments, etc.) were available, this information could be used to refine these estimates. However, this data is not available. After updating with the observed data, the updated number of pole failures is valid subject to the assumption that all poles are identical. Again, if more information on the distribution system becomes available, particularly information about the fraction of the total poles that are of the different pole types, the fragility for each pole type could be developed and used to more accurately estimate the total number of damaged poles. Overall, the posterior mean probability of failure for Hurricane Ivan, the pink points, is higher than the mean probability of failure for Hurricane Dennis and Hurricane Katrina because the observed damage for Hurricane Ivan is much more severe than for the other hurricanes. The one clump of dots for Ivan above the rest is from one of areas in which damage data was collected for Hurricane Ivan. Even though the area did not experience high wind speeds, the damage in the area is higher than the other areas. It is not clear why this is the case, but this warrants further investigation in the future. The observed data are available for only relatively low wind speeds (i.e., winds no greater than 110 mph in my data set) so that the lower and upper bounds are available only for the limited wind range. If I assume that the probability of failure has the same pattern for wind speeds stronger than those experienced in Hurricanes Dennis, Ivan, and Katrina, I could extend the probability of failure to the higher wind speeds. However, as with any probably model, one must be cautious about using the results beyond the range of conditions for which the model was developed. At the same time, extending the results to higher wind speeds may prove useful, and if more damage data for higher wind speeds are collected during future hurricanes, uncertainty in the model predictions for the higher wind speeds can be reduced by simply updating the model with the additional damage data. Figure 7.10 shows the updated and extended fragility curve for Southern Pine and the 12.47 kV line together with the original fragility curve used as the prior. Again note that the posterior fragility curve given in Figure 7.10 should be used with caution above wind speeds of approximately 110 mph.

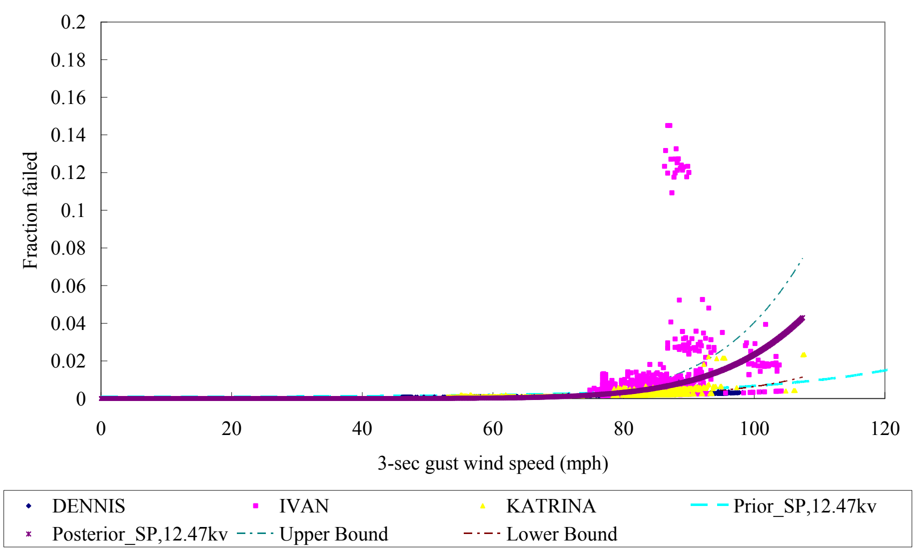

Figure 7.8. Mean fraction failed of poles for 3 Hurricanes, prior fragility curve and posterior fragility curve for Southern Pine, 12.47 kV distribution line.

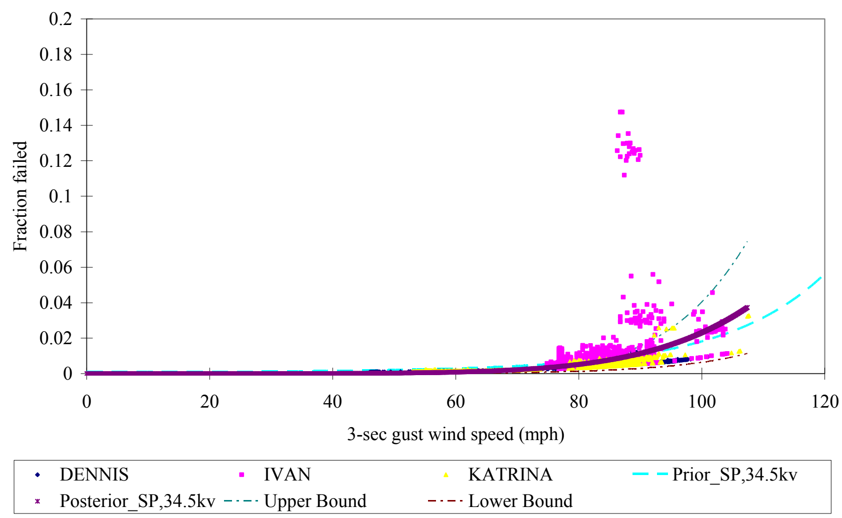

Figure 7.9. Mean fraction failed of poles for 3 Hurricanes, prior fragility curve and posterior fragility curve for Southern Pine, 34.5 kV distribution line.

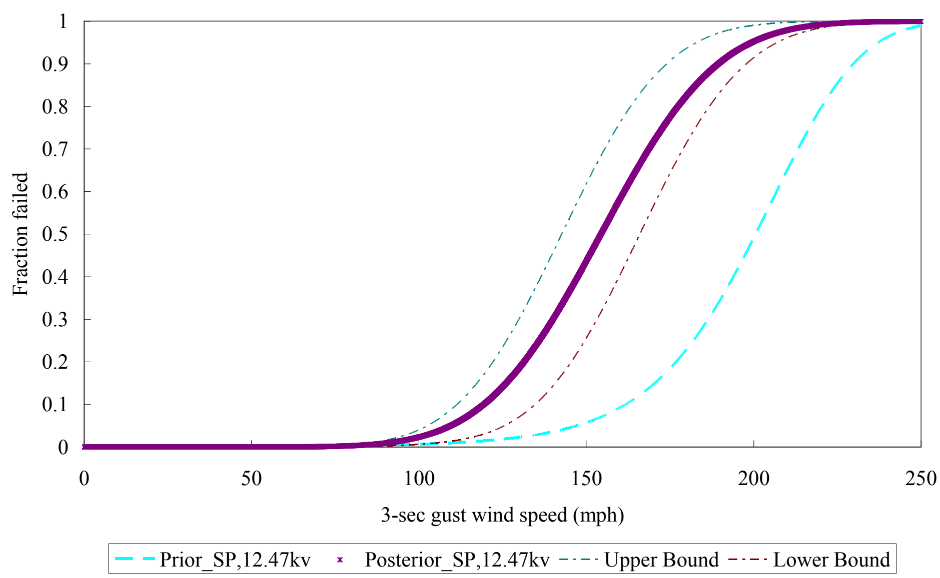

Figure 7.10. Prior fragility curve, posterior fragility curve, and its confidence intervals for Southern Pine, 12.47 kV distribution line.

In order to determine whether or not the informative prior based on structural reliability models adds value to the analysis, I used three priors, a beta(0.1, 0.1), a beta (1, 1), and a beta(10, 10), and updated them with the observed damage data. These three priors range from non-informative for the beta(0.1, 0.1) and beta(1, 1) distributions to mildly informative with a mean of 0.5 for the beta(10, 10) distribution. If the prior based on the structural reliability model adds value to the analysis, the posterior based on this prior should be substantially different from the posteriors based on the other priors. If the posteriors were very similar, it would be a clear indication that the model-based prior is not adding value to the analysis.

Figure 7.11 shows the prior fragility curve using the structural reliability model for a 12.47 kV line with Southern Pine poles, its posterior fragility curve, and the posterior fragility curves with the three other priors. There is large difference between the posterior fragility curve with the structural reliability model and the posterior fragility curves with three priors. Figure 7.11 also shows that the posterior obtained by updating from the model-based prior exhibits substantially more spread than the posteriors found by updating from the other priors. That is, there is considerably more uncertainty in the posterior using the model-based prior than in the other posteriors using the other priors. Given that the posteriors were all based on the same data and the same likelihood, this means that the mode-based prior is having less of an influence on the posterior than the other priors. Essentially, it is a ‘weaker’ prior in the sense that it contains less information (e.g., has higher entropy) than the other priors. This suggests that the model-based prior is likely a more accurate reflection of the degree of uncertainty than the other priors. The prior based on the structural reliability model is adding value to the analysis by more accurately reflecting the prior uncertainty.

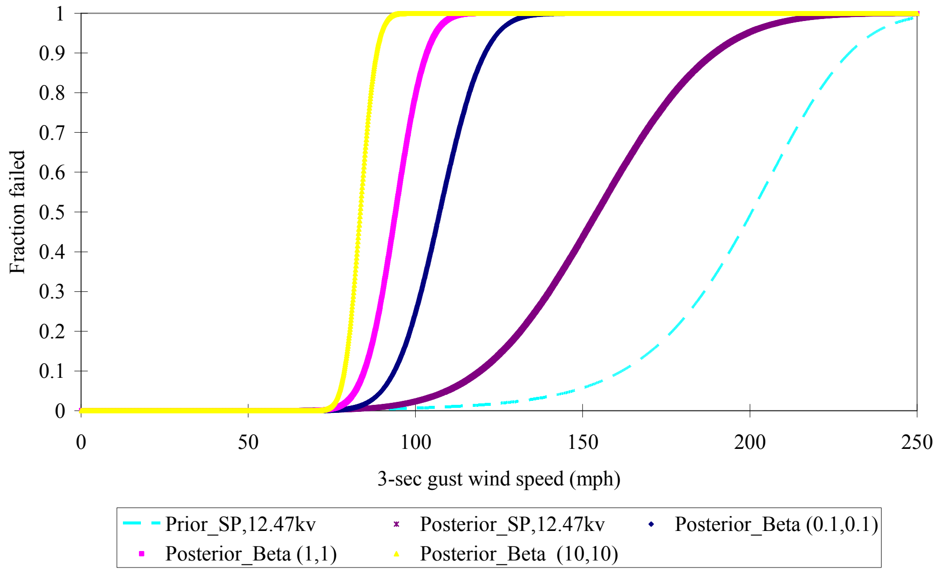

Figure 7.11. Posterior fragility curves with structural reliability prior for Southern Pine, 12.47 kV distribution line and three priors, beta(0.1, 0.1), beta(1, 1), and beta(10, 10).

Figure 7.12 shows the updated and extended posterior fragility curves for 12.47 kV and 34.5 kV distribution lines with Southern Pine poles together with the original fragility curves used as priors. These fragility curves are meant to represent the types of lines and poles used in the service area of the case study area. Figure 7.12 shows that the two updated fragility curves are closer than the priors. This reinforces the discussion of Figure 7.11 above. The data is driving the posteriors to a large degree, indicating that the model-based priors contain considerable uncertainty.

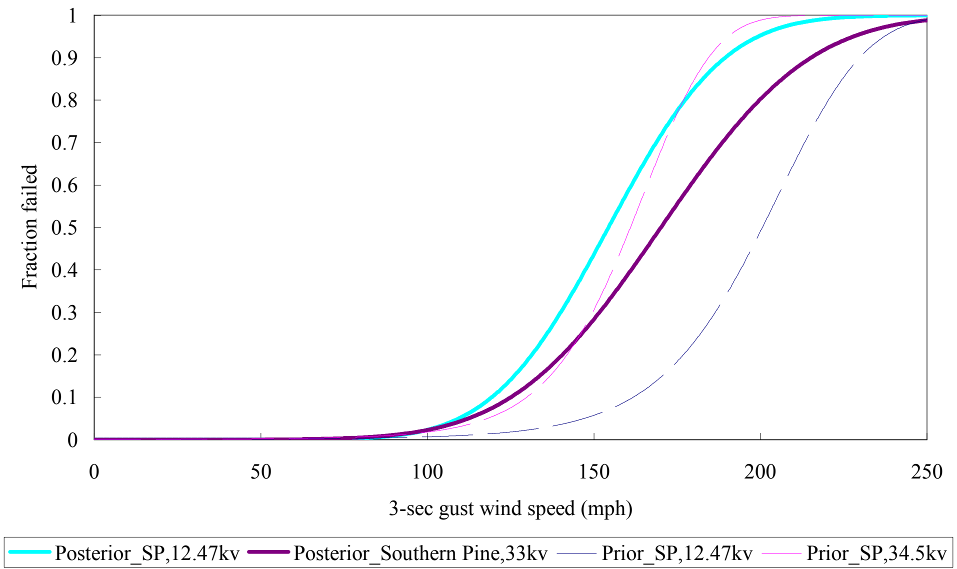

Figure 7.12. Prior fragility curve and posterior fragility curves for Southern Pine, two distribution lines.

With the updated fragility curves shown in Figure 7.12, I can estimate the number of damaged poles for a given wind speed. In this study, the number of damaged poles can be calculated by multiplying the probability of failure for a given wind speed by the total number of poles in each grid cell. The model developed in this section is an innovative approach for integrating Bayesian updating with structural reliability analysis for estimating the reliably of power utility poles during hurricanes and accurately predicting damage to the power distribution system during hurricanes. Finally, the updated fragility curves shown in Figure 7.12 can provide the basis for a data-based approach for predicting the number of damaged poles during hurricanes.

The damage estimation models developed in this dissertation are not perfectly accurate for predicting the damage in the power distribution system. However, the models can provide a good starting point even though the models do not consider all possible failure modes (e.g. tree-induced failure) and detailed information of a power distribution system (e.g. the age of power distribution system poles and the proportions of failures caused by different failure modes in the actual system). A useful extension of the damage estimation model is to consider the detailed information of the power distribution system if the data is available and various other priors as well as the prior developed based on the structural reliability model. Also, damage for other power distribution system structures such as a concrete pole and a transmission tower could be considered in future.

8. SUMMARY AND CONCLUSIONS

8.1 *Summary*

The goals of this dissertation are to develop models which are useful for managing power outage risks and to enable proper preparation for pre-storm planning. The models developed in this study provide a basis for managing the effects of hurricanes before they make landfall, and for restoring electric power after a hurricane. The power outage prediction models estimate the number of outages expected to be caused in the Gulf Coast region by an incoming hurricane. The customers out prediction models estimate the number of customers without power. The statistical damage estimation models are used for predicting the number of damaged poles and damaged transformers based on the past data for hurricanes at the limited area. The physical damage estimation model can estimate the probability of failure of a power distribution system pole given a wind speed. By adopting Bayesian approach, it is possible to more reliably estimate damage to the power distribution system than based on either a structural reliability model or observed data alone, integrating the fragility curves based on the structural reliability model with the observed data.

8.2 *Conclusions*

For accurately estimating the spatial distribution of power outages, customers without power, and damage in the power distribution system during an approaching hurricane based only on measurable characteristics of hurricanes, statistical and physical models were developed. These models can directly help utility companies improve their post-hurricane response through improved pre-hurricane planning. The statistical models developed are based on negative binomial GLMs and negative binomial GAMs in combination with principal components analysis to account for both collinearity and overdispersion in the data sets used. Previous work for predicting power outages used binary variables representing particular hurricanes in order to achieve a good fit to the past data. To use these models for predicting power outage risk and damage during future hurricanes, one must implicitly assume that an approaching hurricane is similar to the average of the past hurricanes. The model developed in this study replaces these binary variables with physically measurable variables, enabling future predictions to be based on only well-understood characteristics of hurricanes.

Through the use of GAMs, this study has improved the accuracy of models for estimating the spatial distribution of power outages during an approaching hurricane. This will in turn help utility companies improve their post-hurricane response through improved pre-hurricane allocation of repair crews to different portions of the service area. Furthermore, it has shown that semi-parametric GAMs can provide substantially improved accuracy in power outage estimates relative to GLMs.

This study also involved using the Bayesian approach for predicting the number of damaged poles with the physical damage estimation model. This Bayesian approach was used in updating the probability that poles fail based on structural reliability analysis together with actual damage data for poles in the power distribution system. This integrated model presents an innovative approach for predicting damage to the power distribution system poles during hurricanes. Finally, fragility curves for representative distribution system poles are presented and the number of damaged poles is predicted using the probability of failure from the updated damage estimation models.

The major research contributions of this thesis are:

- Using the alternative hurricane descriptors, which are relatively easy to obtain, I obtained more useful prediction models.
- By developing the customers out prediction model, I could provide risk measures more closely aligned with the methods currently used for pre-hurricane planning in utility companies.
- By developing the statistical damage estimation model, I could make it enable a utility company to estimate the amount of actual damage in their service areas during hurricanes.
- Fitting negative binomial GAMs to the data provided better predictions of power outages during hurricanes, capturing non-linearity of the data set.

- The physical damage estimation model produced updated prediction models for future events by integrating Bayesian updating with structural reliability analysis to reliably predict damage to the power distribution system during hurricanes, providing a data-based tool for predicting the number of damaged poles in certain wind speeds before hurricane landfall.

These statistical models and physical models can provide a basis for improving pre-hurricane planning for post-hurricane response, and it can provide a basis for future research to further improve hurricane risk estimation models for hurricane-prone areas. The models developed both (a) provide grid-cell level estimates of power outages, customers without power, damaged poles and transformers for future hurricanes and (b) provide insight into which parameters most strongly affect the predictions from the models. These models can provide valuable information for pre-hurricane planning within the particular large investor-owned utility company in the Gulf Coast region, and they also yield more general insights into factors that most influence hurricane risks in the Gulf Coast region of the U.S. during hurricanes. By quantifying where the impacts of the hurricane are likely to be the worst, the results of the models can help managers decide how many crews and how much extra material to have on hand before a hurricane makes landfall, where to position crews and material to enable the fastest possible response after the hurricane, and how the distribution line should be installed based on the expected hurricane seasonal losses of poles. The damaged estimation models can be used to evaluate insurance needs. The models can also be used to examine a number of potentially ‘worst case’ scenarios by running the model with a particularly strong hurricane (past or hypothetical) and an assumed track. This would provide an estimate of how bad things might be in a future hurricane, providing a case against which current response plans could be tested.

REFERENCES

Abdulla, F., Lettenmaier, D., Wood, E., and Smith, J. (1996). "Application of a macroscale hydrologic model to estimate the water balance of the Arkansas-Red River Basin." *J. Geophys. Res.*, 101(D3), 7449–7459.

Agresti, A. (2002), *Categorical Data Analysis*, 2nd Ed. Hoboken, NJ: Wiley-Interscience.

American National Standards Institute (ANSI), (2002). "Specifications and dimensions for wood poles." O5.1, New York.

Andreadis, K. M., Clark, E. A., Wood, A. W., Hamlet, A. F., and Lettenmaier, D. P. (2005). "Twentieth-century drought in the conterminous United States." *J Hydrometeorology*, 6(6), 985–1001.

ASCE (American Society of Civil Engineers), (2005). "Minimum design loads for buildings and other structures." ASCE 7-05, Reston, VA.

Cameron, A. C., and Trivedi., P. K. (1998). *Regression Analysis of Count Data*, Econometric Society Monographs No. 30, Cambridge, UK: Cambridge University Press.

Caribbean Disaster Mitigation Project (1996). "Hurricane vulnerability and risk analysis of the VINLEC transmission and distribution system." Organization of American States.

Cherkauer, K. A., and Lettenmaier, D. P. (1999). "Hydrologic effects of frozen soils in the upper Mississippi River basin." *J. Geophys. Res.*, 104(19), 599–519,610.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003). *Bayesian Data Analysis*, 2nd Ed. Chapman & Hall/CRC, Portland, OR.

Georgiou, P. N. (1985). "Design wind speeds in tropical cyclone-prone regions." Ph.D. Dissertation, University of Western Ontario.

Guttman, N. B. (1998). "Comparing the palmer drought index and the standardized precipitation index." *J the American Water Resources Association*, 34(1), 113–121.

Guttman, N. B. (1999). "Accepting the standardized precipitation index: A calculation algorithm." *J the American Water Resources Association*, 35(2), 311–322.

Han, S. R., Guikema, S. Quiring, S., Lee, K. H., Davidson, R., and Rosowsky, D. (2008a). "Estimating the spatial distribution of power outages during hurricanes in the Gulf Coast region." *Reliability Engineering and System Safety*, in press.

Han, S. R., Guikema, S. D., and Quiring, S. M. (2008b). "Improving the accuracy of hurricane power outage predictions using generalized additive models." submitted to *Technometrics* [under revision after first review].

Han, S. R., Guikema, S. Quiring, S., Lee, K. H., Davidson, R., and Rosowsky, D. (2008c). "Estimating risk of customer outages and damage to power distribution systems during hurricanes." working paper, Zachry Department of Civil Engineering, College Station, TX. To be submitted to *Reliability Engineering and System Safety*.

Hansen, M. C., Defries, R. S., Townshend, J. R. G., and Sohlberg, R. (2000). "Global land cover classification at 1 km spatial resolution using a classification tree approach." *Int J Remote Sensing*, 21(6), 1331–1364.

Hastie, T. J., and Tibshirani, R. J. (1990). *Generalized Additive Models*, Chapman & Hall/CRC, New York.

Heim, R. R., Jr. (2002). "A review of twentieth century drought indices used in the United States." *Bulletin of the American Meteorological Society*, 83(8), 1149–1165.

Huang, Z., Rosowsky, D. V., and Sparks, P. R. (2001). "Hurricane simulation techniques for the evaluation of wind speeds and expected insurance losses." *Journal of Wind Engineering and Industrial Aerodynamics*, 89, 605–617.

Keshavarzian, M., and Priebe, C. H. (2002). "Wind performance of short utility pole structures." *Practice Periodical on Structural Design and Construction*, 7(4), 141–146.

Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S. J. (1994). "A simple hydrologically based model of land surface water and energy fluxes for GCMs." *Journal of Geophysical Research*, 99(14), 415–414,428.

Liang, X., Lettenmaier, D. P., and Wood, E. F. (1996a). "One-dimensional statistical dynamical representation of subgrid spatial variability of precipitation in the two-layer variable infiltration capacity model." *Journal of Geophysical Research*, 101(21), 403–421,422.

Liang, X., Wood, E. F., and Lettenmaier, D. P. (1996b). "Surface soil moisture parameterization of the VIC-2L model: Evaluation and modifications." *Global and Planetary Change*, 13, 195–206.

Liu, H., Davidson, R. A., Rosowsky, D. V., and Stedinger, J. R. (2005). "Negative binomial regression of electric power outages in hurricanes." *J Infrastructure Systems*, 11(4), 258–267.

Liu, H., Davidson, R. A., and Apanasovich, T. V. (2008). "Spatial generalized linear mixed models of electric power outages due to hurricanes and ice storms." *Reliab Engng and Syst Safety*, 93(6), 897–912.

Maurer, E. P., Wood, A. W., Adam, J. C., Lettenmaier, D. P., and Nijssen, B. (2002). "A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States." *J. Clim.*, 15(22), 3237–3251.

McKee, T. B., Doesken, N. J., and Kleist, J. (1993). "The relationship of drought frequency and duration to time scales." *International 8th Conference on Applied Climatology*, Anaheim, CA, 179–184.

McKee, T. B., Doesken, N. J., and Kleist, J. (1995). “Drought monitoring with multiple time scales.” *International 9th Conference on Applied Climatology*, Dallas, TX, 233–236.

Meng, L., and Quiring, S. M. (2007). “A comparison of soil moisture models using Soil Climate Analysis Network (SCAN) observations.” *Journal of Hydrometeorology*, in press.

National Electrical Safety Code (NESC 2007). *IEEE Std*., Piscataway, NJ.

Nijssen, B., Lettenmaier, D., Liang, X., Wetzel, S., and Wood, E. (1997). “Streamflow Simulation for Continental-Scale River Basins.” *Water Resources Research*, 33(4), 711–724.

Nijssen, B, Schnur, R., and Lettenmaier, D. P. (2001). “Global retrospective estimation of soil moisture using the variable infiltration capacity land surface model, 1980–93.” *J. Clim.*, 14(8), 1790–1808.

Phoon, K. K., Kulhawy, F. H., and Grigoriu, M. D. (1995). “Reliability-based design of foundations for transmission line structures.” Rep. No. TR-105000, Electric Power Research Institute, Palo Alto, CA.

Raiffa, H., and Schaifer, R. (2000). *Applied Statistical Decision Theory*, Wiley Classics Library, New York.

Robock, A., Luo, L., Wood, E. F., Wen, F., Mitchell, K. E., et al. (2003). “Evaluation of North American land data assimilation system over the southern great plains during the warm season.” *Journal of Geophysical Research*, 108(D22), 8846, doi:8810.1029/2002JD003245.

Rosowsky, D., Sparks, P., and Huang, Z. (1999). “Wind field modeling and hurricane hazard analysis.” Report to the South Carolina Grant Consortium and Civil Engineering Department, Clemson University, Civil Engineering Department.

Vickery, P. J., and Twisdale, L. A. (1995). “Wind-field and filling models for hurricane wind- speed prediction.” *Journal of Structural Engineering*, 121(11), 1700–1709.

Wahba, G. (1990). *Spline models for observational data*, SIAM, Philadelphia.

Wareing, B. (2005). *Wood pole overhead lines*, The Institution of Electrical Engineers, London, UK.

Wood, E. F., Lettenmaier, D., Liang, X., Nijssen, B., and Wetzel, S. W. (1997). “Hydrological modeling of continental-scale basins.” *Annual Review of Earth and Planetary Sciences*, 25, 279–300.

Wood, S. N. (2006). “Generalized additive models: An introduction with R.” Chapman & Hall/CRC, New York.

Wu, H., Svoboda, M. D., Hayes, M. J., Wilhite, D. A., and Wen, F. (2007). “Appropriate application of the standardized precipitation index in arid locations and dry seasons.” *Int J of Climatology*, 27(1), 65–79.

Zhang, Y., and Der Kiureghian, A. (1994). "Two improved algorithms for reliability analysis." *Proc. of the 6th IFIP WG 7.5 Working Conference on Reliability and Optimization of Structural Systems*, Assisi, Italy, 297–304.

APPENDIX A

— The commands history in the program R for the PCA

```
data<-read.table('regressiondata.txt',header=TRUE)
summary(data)
attach(data)
library(MASS)
PCA1<-
prcomp(~Transformer+Pole+Switch+Overhead+Underground+Customer+Windspeed+Duration+FSM1+F
SM2+FSM3+MAP+SPI1+SPI2+SPI3+SPI6+SPI12+SPI24+LC1+LC2+LC3+LC4+LC5+LC7+LC8+LC9,
scale=TRUE)
summary(PCA1)
write.table(PCA1$x,file="PC.txt",sep=",",row.names=FALSE,col.names=TRUE)
```

Table A.1. Variable loadings of principal components for the State A

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24	PC25	PC26
Transformer	0.400	-0.032	0.013	-0.017	0.006	-0.020	-0.009	0.007	0.019	-0.024	0.011	0.118	-0.013	-0.008	-0.017	0.202	-0.029	-0.139	-0.014	-0.022	-0.076	0.200	0.031	0.151	-0.831	0.000
Pole	0.386	-0.032	0.014	0.003	0.008	-0.027	-0.009	0.014	0.033	-0.053	-0.015	0.313	0.017	-0.010	-0.019	0.174	-0.032	-0.143	-0.009	-0.015	-0.075	0.176	0.012	0.659	0.472	0.000
Switch	0.392	-0.034	0.017	-0.032	0.008	-0.006	-0.029	0.035	0.019	0.007	-0.003	0.017	0.000	0.010	-0.020	0.131	-0.076	0.373	0.039	0.070	0.278	-0.760	-0.102	0.065	-0.036	0.000
Overhead	0.389	-0.027	0.012	-0.028	-0.001	-0.033	0.012	-0.043	0.007	-0.038	0.050	0.001	-0.030	0.002	-0.048	0.356	0.005	-0.531	0.064	0.047	-0.002	-0.053	-0.006	-0.600	0.243	0.000
Underground	0.242	-0.027	0.009	-0.069	-0.008	0.063	-0.024	0.041	-0.042	0.162	0.107	-0.898	-0.108	0.021	-0.001	-0.026	0.028	-0.086	-0.014	-0.005	-0.029	0.067	-0.003	0.233	0.056	0.000
Customer	0.386	-0.038	0.021	-0.023	0.015	-0.014	-0.045	0.059	0.024	0.027	-0.020	0.004	-0.007	0.005	0.004	0.009	-0.125	0.672	-0.108	-0.071	-0.166	0.435	0.057	-0.339	0.148	0.000
Windspeed	-0.007	-0.249	-0.399	0.106	0.074	0.207	-0.283	-0.049	-0.068	-0.235	0.035	-0.003	-0.070	0.283	-0.073	-0.025	-0.009	0.019	-0.099	0.658	0.045	0.105	-0.168	0.005	-0.008	0.000
Duration	0.008	-0.251	-0.462	0.063	0.025	0.152	-0.187	-0.026	-0.070	-0.152	0.037	0.009	-0.052	0.220	-0.046	0.044	0.244	-0.008	-0.149	-0.693	0.100	-0.053	0.034	-0.014	0.009	0.000
FSM1	-0.015	-0.311	-0.226	-0.086	0.140	0.028	0.216	-0.166	0.319	0.205	-0.056	0.004	0.020	-0.209	0.396	0.160	0.480	0.119	0.306	0.144	-0.085	0.011	0.100	0.008	0.003	0.000
FSM2	-0.024	-0.376	-0.189	-0.049	-0.098	-0.216	0.171	0.065	-0.033	0.073	-0.027	-0.012	0.007	-0.253	0.303	-0.059	-0.471	-0.097	-0.174	-0.026	0.524	0.161	-0.028	-0.008	0.006	0.000
FSM3	-0.031	-0.300	0.063	0.018	-0.275	-0.279	-0.059	0.181	-0.404	-0.147	0.008	-0.014	0.008	-0.417	-0.010	0.058	0.299	0.031	-0.275	0.063	-0.346	-0.119	-0.218	-0.002	-0.002	0.000
MAP	0.001	-0.016	-0.065	0.099	0.606	0.157	0.276	0.173	0.132	-0.376	0.035	-0.056	-0.212	-0.432	-0.248	-0.075	-0.101	-0.013	-0.080	-0.029	-0.073	-0.050	0.034	-0.004	0.000	0.000
SPI1	-0.040	-0.362	-0.056	-0.070	0.039	-0.059	0.215	-0.071	0.293	0.325	-0.053	0.029	0.102	0.218	-0.277	-0.103	-0.290	-0.067	-0.069	-0.093	-0.425	-0.127	-0.404	-0.006	-0.011	0.000
SPI2	-0.049	-0.405	0.197	-0.023	-0.002	-0.092	0.084	0.069	-0.003	0.088	-0.026	0.004	0.069	0.177	-0.347	0.001	0.073	-0.006	-0.185	0.135	0.047	-0.077	0.737	0.009	0.003	0.000
SPI3	-0.042	-0.332	0.382	0.029	0.053	0.029	-0.050	0.033	-0.072	-0.053	0.007	-0.007	0.038	-0.009	-0.370	0.033	0.201	0.070	0.451	-0.076	0.398	0.239	-0.342	0.001	-0.003	0.000
SPI6	-0.032	-0.337	0.253	0.094	0.037	0.128	-0.178	0.005	-0.162	-0.279	0.045	-0.021	-0.127	0.118	0.408	-0.062	-0.358	-0.024	0.410	-0.110	-0.327	-0.126	0.160	0.005	-0.002	0.000
SPI12	-0.022	-0.110	0.520	0.079	0.195	0.218	0.021	-0.095	0.168	-0.039	-0.025	0.010	-0.070	0.200	0.359	0.074	0.182	-0.036	-0.566	-0.015	0.113	0.008	-0.197	-0.012	0.003	0.000
SPI24	-0.001	0.081	-0.078	0.158	0.159	-0.377	0.408	0.540	-0.136	-0.073	0.008	-0.017	-0.057	0.487	0.173	0.058	0.135	0.020	0.102	0.031	0.010	0.028	-0.085	0.000	-0.003	0.000
LC1	0.002	-0.019	-0.026	0.079	-0.072	0.510	0.326	0.058	-0.432	0.413	0.356	0.183	-0.232	-0.031	-0.015	0.050	-0.019	0.028	0.001	0.029	-0.004	0.003	0.000	-0.008	0.004	0.204
LC2	0.371	-0.035	0.021	0.021	-0.014	-0.013	-0.010	0.037	0.029	0.002	-0.053	0.108	0.051	-0.009	0.076	-0.783	0.215	-0.172	0.033	0.017	0.039	-0.035	0.005	-0.078	-0.002	0.369
LC3	0.000	-0.013	-0.005	-0.150	-0.350	0.373	0.276	0.239	0.210	-0.383	0.220	-0.080	0.576	0.006	0.030	0.038	-0.012	0.015	-0.004	0.007	-0.010	0.011	-0.004	-0.003	-0.001	0.041
LC4	-0.135	0.015	0.000	-0.602	0.292	-0.090	-0.141	0.028	-0.145	-0.036	-0.012	-0.008	0.104	0.061	0.026	0.202	-0.052	0.028	-0.028	-0.007	-0.018	0.012	-0.024	0.024	0.002	0.649
LC5	-0.088	-0.019	0.027	0.233	-0.011	-0.209	-0.360	0.254	0.410	0.137	0.672	0.055	-0.069	-0.093	0.011	0.053	-0.007	0.005	-0.010	-0.004	0.004	-0.007	0.015	-0.001	0.000	0.201
LC7	-0.061	-0.011	0.021	-0.257	-0.458	0.091	0.076	0.190	0.333	-0.168	-0.208	0.032	-0.676	0.037	-0.077	0.029	0.025	0.014	0.021	0.008	0.015	0.009	-0.007	0.002	0.002	0.139
LC8	-0.013	0.025	-0.002	0.411	-0.168	-0.229	0.316	-0.553	0.008	-0.270	0.079	-0.122	-0.038	0.060	-0.075	0.125	-0.038	0.106	-0.002	-0.011	0.003	0.014	0.009	0.039	-0.002	0.457
LC9	-0.049	0.005	-0.026	0.480	-0.034	0.216	-0.182	0.336	0.085	0.203	-0.541	-0.078	0.174	-0.123	-0.001	0.206	-0.071	-0.034	0.017	-0.010	-0.016	-0.008	0.021	-0.008	-0.001	0.362

Table A.2. Variable loadings of principal components for the State B

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24
Transformer	0.337	0.167	-0.222	-0.061	-0.010	0.003	0.007	-0.013	0.040	-0.017	0.003	0.030	-0.039	-0.015	-0.137	-0.027	0.001	0.385	-0.099	0.002	0.044	0.788	-0.011	0.000
Overhead	0.333	0.157	-0.228	-0.090	-0.020	0.015	-0.001	-0.021	0.038	-0.032	0.000	-0.016	-0.045	0.023	-0.095	-0.094	-0.448	0.505	-0.177	0.016	0.011	-0.539	0.038	0.000
Switch	0.330	0.161	-0.219	-0.062	-0.007	0.032	0.002	-0.020	0.028	-0.022	0.018	0.013	-0.049	0.013	-0.321	-0.060	-0.387	-0.740	0.010	0.013	0.019	0.057	-0.002	0.000
Customer	0.323	0.173	-0.219	-0.010	0.008	0.017	0.028	-0.018	0.029	-0.049	-0.013	0.112	-0.062	-0.079	-0.358	0.151	0.716	0.002	0.204	-0.053	-0.059	-0.279	-0.014	0.000
Windspeed	0.126	0.210	0.409	-0.231	0.010	-0.028	-0.022	-0.161	0.019	-0.066	-0.119	-0.032	-0.119	0.158	0.038	0.336	0.037	-0.065	-0.408	-0.578	0.064	-0.012	-0.090	0.000
Duration	0.063	0.294	0.337	-0.044	-0.126	-0.028	-0.049	-0.288	-0.015	-0.034	-0.098	-0.012	0.168	-0.244	-0.026	0.512	-0.146	0.044	0.190	0.509	-0.030	0.019	0.099	0.000
FSM1	-0.177	0.348	-0.004	0.022	-0.246	0.002	-0.064	0.066	0.021	-0.022	-0.049	0.113	0.289	-0.089	-0.038	-0.249	0.083	-0.048	-0.370	0.048	-0.668	0.005	-0.115	0.000
FSM2	-0.163	0.283	-0.066	0.139	-0.360	0.035	-0.119	0.261	0.058	-0.068	-0.031	0.032	0.433	-0.263	-0.049	-0.025	-0.060	0.027	0.166	-0.351	0.481	-0.014	0.027	0.000
FSM3	0.154	-0.248	0.046	-0.082	0.301	0.090	-0.028	0.486	0.102	-0.140	-0.130	0.205	0.510	0.316	-0.088	0.300	-0.043	0.011	-0.067	0.084	-0.082	0.007	0.060	0.000
MAP	0.229	0.002	0.337	-0.201	-0.199	-0.101	0.069	0.124	-0.017	-0.015	0.269	-0.581	0.214	0.226	-0.117	-0.304	0.179	-0.014	-0.051	0.220	0.153	-0.024	-0.059	0.000
SPI1	-0.140	0.251	0.016	-0.174	0.210	0.089	-0.048	0.582	0.129	-0.083	-0.016	-0.160	-0.419	-0.352	0.068	0.074	0.065	-0.054	-0.241	0.222	0.108	-0.015	0.088	0.000
SPI2	0.124	-0.052	0.275	-0.397	0.393	0.023	0.045	-0.159	-0.030	0.002	-0.198	0.249	0.203	-0.430	0.000	-0.465	0.024	0.003	0.094	-0.044	0.089	-0.010	-0.037	0.000
SPI3	0.211	-0.322	0.069	-0.016	-0.077	0.017	-0.069	0.169	0.052	-0.031	0.109	-0.358	-0.020	-0.428	-0.007	0.165	-0.156	0.056	0.298	-0.350	-0.458	0.040	-0.040	0.000
SPI6	0.159	-0.359	-0.058	0.166	-0.176	0.011	-0.022	-0.027	-0.003	-0.015	-0.069	0.073	0.033	-0.290	0.014	0.120	0.050	-0.048	-0.372	0.198	0.212	-0.043	-0.666	0.000
SPI12	0.187	-0.341	0.009	0.117	-0.251	-0.007	-0.036	-0.096	-0.011	-0.009	-0.066	0.048	0.042	-0.222	0.020	-0.036	0.130	-0.086	-0.412	0.028	0.066	-0.006	0.709	0.000
SPI24	0.139	-0.111	0.224	-0.147	-0.490	0.020	-0.200	0.191	0.162	-0.181	-0.411	0.266	-0.335	0.216	0.067	-0.189	-0.021	0.018	0.267	0.082	-0.041	0.012	-0.044	0.000
LC1	0.079	0.061	0.337	0.478	0.129	0.155	0.002	0.082	-0.285	-0.254	0.093	0.097	-0.123	-0.010	-0.229	-0.124	-0.060	0.062	-0.002	-0.020	-0.011	0.013	0.002	-0.588
LC2	0.319	0.160	-0.220	-0.057	-0.006	0.014	0.036	0.002	0.016	0.018	-0.031	-0.058	0.129	0.022	0.729	0.017	0.103	-0.135	0.064	0.015	-0.015	-0.027	-0.011	-0.481
LC3	0.084	0.052	0.212	0.293	0.039	0.251	0.265	-0.013	0.690	0.485	-0.051	0.004	-0.007	0.008	-0.045	-0.062	-0.010	0.014	0.005	-0.011	-0.001	-0.008	0.000	-0.067
LC4	-0.155	-0.112	-0.037	-0.405	-0.267	-0.110	0.365	0.173	-0.235	0.458	0.040	0.177	-0.035	-0.035	-0.217	0.147	-0.054	0.039	-0.009	-0.013	-0.005	0.005	0.010	-0.421
LC5	-0.235	-0.075	-0.219	-0.076	0.013	0.287	0.238	-0.178	0.032	-0.242	-0.609	-0.455	0.069	0.031	-0.170	0.001	0.016	0.016	-0.024	0.003	0.003	0.019	0.004	-0.203
LC7	-0.153	-0.100	-0.041	-0.290	-0.150	0.386	0.196	-0.173	0.318	-0.475	0.513	0.196	-0.004	-0.021	0.018	0.044	-0.005	0.004	-0.009	0.013	0.010	0.009	0.001	-0.096
LC8	-0.095	-0.073	-0.068	-0.196	0.044	0.366	-0.767	-0.135	0.037	0.308	0.080	-0.087	0.029	0.061	-0.130	0.011	0.046	0.019	-0.039	0.030	0.034	-0.002	-0.008	-0.250
LC9	-0.152	-0.074	-0.073	-0.028	0.104	-0.710	-0.182	-0.097	0.466	-0.187	0.004	-0.025	0.013	-0.028	-0.157	-0.001	-0.014	0.007	-0.030	0.005	0.017	0.000	0.003	-0.359

Table A.3. Variable loadings of principal components for the State C

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24	PC25
Transformer	0.387	-0.142	0.017	-0.016	0.008	-0.028	0.063	-0.026	0.001	-0.004	0.049	-0.026	0.095	-0.144	0.006	-0.069	0.035	-0.004	-0.306	0.144	0.047	0.021	-0.039	-0.816	0.000
Pole	0.361	-0.137	0.029	-0.026	0.023	-0.007	0.098	-0.036	-0.014	-0.014	0.077	-0.036	0.170	-0.165	0.018	0.573	-0.379	0.024	0.165	-0.481	-0.201	-0.031	0.003	0.036	0.000
Switch	0.377	-0.139	0.016	-0.017	0.007	-0.030	0.057	-0.021	0.010	0.001	0.039	-0.005	0.092	-0.157	0.004	-0.167	0.093	-0.052	0.776	0.357	0.165	0.007	0.003	0.050	0.000
Overhead	0.383	-0.139	0.016	-0.011	0.012	-0.013	0.066	0.002	-0.006	0.002	0.028	-0.008	0.088	-0.092	0.005	0.156	-0.126	-0.041	-0.497	0.463	0.233	0.026	0.020	0.504	0.000
Customer	0.368	-0.138	0.017	-0.039	0.014	-0.041	0.063	-0.034	0.024	0.008	0.039	-0.039	0.051	-0.154	-0.003	-0.538	0.380	0.070	-0.164	-0.457	-0.239	-0.032	0.018	0.270	0.000
Windspeed	0.099	0.282	-0.312	-0.101	0.212	-0.082	0.003	-0.162	-0.140	0.281	-0.013	0.036	0.199	0.170	0.121	-0.419	-0.578	-0.036	-0.002	-0.094	0.146	-0.013	-0.018	-0.002	0.000
Duration	0.050	0.231	-0.262	-0.110	0.348	-0.135	0.013	-0.198	-0.145	0.389	0.023	-0.047	0.142	0.051	-0.438	0.304	0.445	-0.028	0.001	0.038	-0.032	0.013	-0.027	0.000	0.000
FSM1	0.064	0.075	-0.144	-0.281	-0.519	0.217	-0.026	-0.066	-0.163	0.352	-0.025	0.066	-0.328	-0.229	0.085	0.074	0.079	0.000	0.002	-0.176	0.332	0.300	-0.062	0.000	0.000
FSM2	0.066	0.167	-0.277	-0.353	-0.455	0.157	0.030	-0.040	-0.030	-0.072	0.018	-0.068	0.141	0.109	0.008	0.011	-0.010	0.107	0.001	0.279	-0.514	-0.354	0.100	0.001	0.000
FSM3	0.084	0.277	-0.143	-0.189	-0.155	0.017	0.004	0.035	0.173	-0.578	-0.036	-0.010	0.369	0.183	-0.149	0.032	0.133	-0.216	-0.007	-0.201	0.369	0.164	-0.095	0.001	0.000
MAP	0.103	0.087	0.008	0.417	-0.111	0.246	-0.197	-0.160	-0.253	-0.063	-0.370	0.616	0.205	-0.083	-0.134	-0.006	0.016	0.037	-0.006	0.018	-0.108	0.039	0.059	0.000	0.000
SPI1	0.136	0.250	-0.366	0.150	0.187	0.025	-0.034	0.119	0.088	-0.172	-0.039	0.088	-0.402	-0.274	0.170	0.086	0.085	-0.120	-0.012	-0.080	0.199	-0.561	0.091	-0.016	0.000
SPI2	0.161	0.366	-0.023	0.100	0.053	-0.024	0.013	0.152	0.117	-0.092	0.050	-0.051	-0.327	-0.166	-0.191	-0.071	-0.162	-0.396	0.000	0.124	-0.428	0.466	0.014	0.012	0.000
SPI3	0.134	0.365	0.227	0.074	-0.052	-0.015	0.011	0.115	0.075	-0.050	0.039	-0.069	-0.129	-0.078	-0.350	-0.058	-0.137	0.647	0.018	0.021	0.087	-0.094	-0.398	0.016	0.000
SPI6	0.087	0.342	0.404	-0.061	-0.051	-0.088	0.027	0.016	0.010	0.074	0.047	-0.019	0.032	0.019	-0.033	0.006	0.012	0.113	0.000	-0.045	0.145	-0.027	0.805	-0.041	0.000
SPI12	0.054	0.217	0.540	-0.027	-0.160	-0.046	0.018	-0.038	-0.075	0.226	0.014	0.041	0.068	0.061	0.026	-0.008	0.030	-0.529	-0.019	-0.036	0.009	-0.393	-0.350	0.010	0.000
SPI24	0.102	0.356	0.097	-0.036	0.165	-0.107	0.027	-0.083	-0.034	-0.046	-0.032	0.076	0.079	0.050	0.735	0.163	0.251	0.212	0.008	0.084	-0.136	0.235	-0.177	0.015	0.000
LC1	-0.003	-0.019	-0.022	-0.063	-0.053	-0.311	-0.620	0.178	-0.417	-0.141	0.482	0.084	0.042	-0.098	0.001	-0.004	-0.002	0.003	-0.007	0.002	-0.010	-0.001	-0.011	0.012	-0.185
LC2	0.339	-0.116	0.012	-0.020	0.018	0.020	-0.050	0.041	0.002	0.004	-0.110	0.102	-0.364	0.664	-0.029	0.057	0.028	0.016	0.037	-0.048	-0.002	0.003	0.010	-0.020	-0.510
LC3	0.011	-0.009	-0.103	0.036	-0.136	-0.379	0.078	0.667	-0.169	0.159	-0.522	-0.121	0.166	-0.040	0.029	0.021	0.024	0.001	0.008	-0.022	-0.008	0.009	0.001	-0.009	-0.044
LC4	-0.106	0.096	-0.154	0.478	-0.309	-0.170	0.102	-0.195	0.273	0.151	0.159	-0.174	0.209	-0.152	0.062	0.007	0.029	-0.009	-0.014	0.004	0.028	0.008	0.010	0.011	-0.574
LC5	-0.042	0.042	0.013	0.187	-0.023	0.035	0.446	-0.169	-0.721	-0.299	0.018	-0.316	-0.123	0.018	-0.019	-0.031	0.010	-0.023	0.005	-0.003	0.035	-0.001	0.014	0.003	-0.064
LC7	-0.159	-0.036	-0.014	-0.237	0.016	-0.259	0.553	0.130	0.006	-0.045	0.258	0.640	-0.029	-0.095	-0.070	-0.024	-0.020	0.025	-0.011	0.018	-0.020	-0.002	-0.020	-0.001	-0.176
LC8	-0.092	0.063	0.094	-0.183	0.306	0.639	0.014	0.339	-0.078	0.059	0.042	-0.074	0.218	-0.211	0.028	-0.040	0.013	-0.015	-0.007	0.011	0.003	0.007	-0.004	-0.001	-0.466
LC9	-0.111	-0.049	0.140	-0.397	0.115	-0.264	-0.152	-0.404	0.023	-0.199	-0.478	-0.058	-0.108	-0.338	-0.062	-0.025	-0.100	0.002	-0.013	0.041	-0.038	-0.025	-0.012	0.006	-0.350

Table A.4. Variable loadings of principal components for damage estimation models provided by the State A

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24	PC25	PC26
Transformer	0.374	-0.088	0.062	-0.043	0.028	-0.006	0.023	-0.024	-0.002	0.057	0.061	-0.084	-0.045	0.027	0.029	0.202	-0.035	-0.036	-0.167	0.011	0.065	0.295	0.143	0.366	-0.714	-0.001
Pole	0.361	-0.086	0.053	-0.023	0.023	-0.004	0.035	-0.034	-0.012	0.132	0.150	-0.275	-0.158	0.043	0.015	0.153	-0.043	-0.025	-0.154	0.010	-0.005	0.103	0.242	0.404	0.657	0.000
Switch	0.376	-0.081	0.037	-0.037	0.030	0.008	0.010	-0.018	-0.013	0.001	0.033	-0.056	-0.026	0.015	0.007	0.046	-0.201	-0.096	0.171	0.013	-0.062	-0.858	0.060	-0.005	-0.135	0.001
Overhead	0.372	-0.077	0.058	-0.046	-0.006	0.004	0.004	-0.042	-0.006	0.022	0.051	0.012	-0.009	0.009	0.042	0.361	0.070	0.022	-0.456	0.028	-0.072	0.063	-0.339	-0.608	0.065	0.001
Underground	0.268	-0.052	0.020	-0.065	0.014	-0.005	-0.035	0.043	0.016	-0.320	-0.284	0.682	0.402	-0.061	0.039	0.024	0.047	0.035	-0.120	0.011	-0.002	0.001	0.123	0.208	0.157	0.000
Customer	0.367	-0.086	0.033	-0.045	0.045	0.014	0.007	0.007	-0.021	-0.065	-0.017	0.012	0.041	0.023	0.014	-0.095	-0.426	-0.173	0.628	-0.043	0.052	0.385	-0.154	-0.219	0.077	0.000
Windspeed	-0.054	-0.277	-0.191	0.095	0.139	-0.016	0.286	-0.088	-0.586	-0.004	0.015	-0.010	0.031	-0.395	0.022	0.221	0.016	0.423	0.164	0.003	-0.072	0.007	0.057	-0.013	-0.008	0.000
Duration	-0.006	-0.253	-0.406	0.008	-0.033	-0.060	-0.046	0.017	-0.143	-0.141	0.102	-0.088	0.058	-0.345	-0.028	-0.240	-0.014	-0.572	-0.230	0.243	0.291	-0.007	-0.028	-0.016	0.009	0.000
FSM1	0.018	0.014	-0.434	0.044	0.303	0.126	0.089	-0.143	0.179	0.158	-0.084	0.070	-0.065	0.056	0.343	-0.211	-0.341	0.185	-0.239	-0.399	0.118	-0.033	-0.209	0.099	0.006	0.000
FSM2	0.031	-0.026	-0.486	-0.034	0.156	0.109	-0.127	-0.023	0.264	-0.001	0.091	0.032	-0.014	0.101	-0.016	0.014	-0.036	0.096	0.057	0.424	-0.533	0.087	0.331	-0.147	-0.039	0.000
FSM3	0.100	0.216	-0.355	-0.090	-0.150	-0.029	-0.142	0.183	0.164	-0.190	0.277	-0.043	0.076	-0.018	-0.417	0.190	-0.035	0.354	0.067	-0.136	0.464	-0.033	0.081	-0.059	0.009	0.000
MAP	-0.025	-0.166	0.127	0.204	0.447	0.191	-0.072	-0.342	0.198	0.311	-0.163	0.084	0.066	-0.109	-0.572	0.066	0.061	-0.072	0.031	0.040	0.162	-0.009	-0.012	-0.011	0.016	0.000
SPI1	0.096	0.249	-0.335	-0.025	0.096	-0.106	0.302	-0.002	-0.161	0.161	-0.198	0.043	-0.004	0.274	-0.007	0.190	0.409	-0.374	0.166	-0.307	0.039	0.007	0.228	-0.131	0.002	0.000
SPI2	0.086	0.396	-0.105	0.050	0.056	-0.044	0.234	-0.004	-0.304	0.025	-0.139	0.022	-0.023	0.320	-0.218	-0.089	-0.151	0.078	-0.086	0.519	0.057	-0.021	-0.378	0.197	0.016	0.000
SPI3	0.084	0.418	0.092	0.101	0.021	0.009	0.044	0.015	-0.194	-0.101	0.095	-0.042	0.055	-0.285	-0.356	-0.224	-0.274	-0.126	-0.247	-0.344	-0.412	0.057	0.166	-0.067	-0.034	0.000
SPI6	0.087	0.365	-0.045	0.142	0.235	0.195	-0.177	-0.032	0.094	-0.136	0.289	-0.004	0.039	-0.328	0.219	0.200	0.329	-0.124	0.223	0.010	-0.126	-0.021	-0.395	0.241	0.019	0.000
SPI12	0.033	0.349	0.226	0.198	0.233	0.177	0.008	-0.135	-0.085	-0.084	0.072	0.017	-0.038	-0.049	0.353	-0.063	-0.073	0.054	-0.051	0.245	0.400	0.009	0.468	-0.283	-0.006	0.000
SPI24	-0.093	-0.251	0.000	0.201	0.129	0.320	-0.196	-0.108	-0.328	-0.414	0.306	0.061	-0.069	0.533	-0.093	-0.001	0.026	-0.085	-0.047	-0.182	-0.035	-0.003	-0.018	0.025	-0.007	0.000
LC1	-0.006	-0.012	-0.034	0.463	0.057	-0.463	-0.201	-0.074	0.019	-0.009	-0.053	-0.319	0.565	0.156	0.110	0.081	-0.067	0.034	-0.013	-0.021	-0.039	-0.005	0.000	0.003	0.001	-0.214
LC2	0.365	-0.081	0.019	0.004	0.020	0.007	0.019	-0.007	-0.011	0.000	0.036	-0.079	-0.061	0.001	-0.052	-0.604	0.461	0.256	0.032	-0.020	0.002	0.018	-0.031	-0.071	-0.029	-0.440
LC3	0.012	-0.026	0.006	0.325	0.088	-0.607	-0.016	-0.045	0.091	-0.070	0.189	0.395	-0.551	-0.032	-0.038	0.023	-0.031	-0.016	0.028	0.003	-0.003	0.004	-0.003	0.001	0.000	-0.034
LC4	-0.150	0.060	0.034	-0.414	0.392	-0.098	-0.334	0.207	-0.173	-0.069	-0.203	-0.047	-0.158	-0.027	-0.001	0.192	-0.132	-0.062	-0.036	-0.013	0.018	-0.007	0.014	0.011	0.013	-0.566
LC5	-0.128	0.029	0.080	-0.403	0.029	-0.164	0.192	-0.338	-0.028	0.250	0.602	0.236	0.316	0.088	0.028	-0.036	-0.060	-0.063	0.006	0.024	0.005	-0.010	0.000	0.007	0.005	-0.192
LC7	-0.094	-0.059	0.087	-0.117	0.199	-0.080	0.566	-0.162	0.359	-0.596	-0.065	-0.257	-0.010	-0.013	-0.059	0.068	0.001	-0.019	-0.020	0.013	-0.020	-0.002	-0.023	0.010	0.001	-0.096
LC8	-0.028	0.096	-0.126	0.147	-0.541	0.218	-0.025	-0.505	0.026	-0.040	-0.165	0.068	-0.151	-0.074	0.012	0.192	-0.113	-0.061	0.040	0.000	-0.006	0.003	0.012	0.041	0.006	-0.487
LC9	-0.071	-0.104	0.044	0.351	-0.016	0.268	0.367	0.581	0.136	0.182	0.190	0.158	0.072	0.008	-0.010	0.115	-0.116	-0.103	-0.026	0.037	0.004	-0.008	0.004	0.003	0.005	-0.395

APPENDIX B

— The commands history in the program R for the saturated model fitting of negative binomial GLMs

```
data<-read.table('regressiondata.txt',header=TRUE)
summary(data)
attach(data)
library(MASS)
NBPCA<-
glm.nb(Outage~Pressure+Time+RMW+PC1+PC2+PC3+PC4+PC5+PC6+PC7+PC8+PC9+PC10+PC11+
PC12+PC13+PC14+PC15+PC16+PC17+PC18+PC19+PC20+PC21+PC22+PC23+PC24+PC25+PC26,dat
a=data)
summary(NBPCA)
```

Table B.1. Regression parameter estimates and p-values (second line of each cell) of power outage prediction models fitted by the negative binomial GLM for State A.

Model	Intercept	Transformer	Pole	Switch	Overhead	Underground	Customer	Hurricane Danny	Hurricane Dennis	Hurricane Georges	Hurricane Ivan	Windspeed	Duration	FSM1	FSM2	FSM3	MAP
0	18.6467 0.1328	-0.0004 0.2786	0.0005 0.0002	0.0073 <.0001	0.1425 <.0001	-0.0302 <.0001	-0.0010 <.0001	-2.0729 <.0001	-0.8812 <.0001	-1.2453 <.0001	0.1346 0.3723	0.0153 <.0001	0.0654 <.0001	0.8842 0.2742	1.4866 0.0321	-3.6997 <.0001	-0.0003 0.2148
1	-0.1805 0.8112	-0.0004 0.2897	0.0005 0.0002	0.0073 <.0001	0.1424 <.0001	-0.0302 <.0001	-0.0010 <.0001	-2.1530 <.0001	-0.8166 <.0001	-1.2543 <.0001	NA	0.0142 <.0001	0.0637 <.0001	NA	1.9264 0.0002	-3.7292 <.0001	NA
2	-0.1632 0.8288	NA	0.0004 0.0001	0.0073 <.0001	0.1396 <.0001	-0.0309 <.0001	-0.0010 <.0001	-2.1534 <.0001	-0.8168 <.0001	-1.2532 <.0001	NA	0.0143 <.0001	0.0636 <.0001	NA	1.9062 0.0002	-3.7128 <.0001	NA

MAP	SPI1	SPI2	SPI3	SPI6	SPI12	SPI24	LC1	LC2	LC3	LC4	LC5	LC7	LC8	LC9	Dispersion Parameter	D.F	Deviance
-0.0003 0.2148	-0.2705 <.0001	0.0509 0.5176	0.4025 <.0001	-0.1603 0.0152	0.1341 0.0743	-0.2974 0.0002	-0.2144 0.0838	-0.2087 0.0925	-0.1852 0.1358	-0.2109 0.0891	-0.2126 0.0865	-0.2168 0.0807	-0.2120 0.0875	-0.2111 0.0888	1.2239 0.0292	33374	18880.06
NA	-0.2467 <.0001	NA	0.4491 <.0001	-0.1057 0.0380	NA	-0.2397 <.0001	-0.0293 0.0002	-0.0237 0.0021	NA	-0.0259 0.0006	-0.0274 0.0004	-0.0317 0.0002	-0.0269 0.0004	-0.0257 0.0007	1.2247 0.0292	33380	18883.70
NA	-0.2449 <.0001	NA	0.4485 <.0001	-0.1049 0.0393	NA	-0.2379 <.0001	-0.0295 0.0002	-0.0239 0.0019	NA	-0.0260 0.0006	-0.0275 0.0004	-0.0318 0.0002	-0.0270 0.0003	-0.0258 0.0007	1.2226 0.0291	33381	18896.95

Table B.2. Model comparisons by likelihood ratio tests for the negative binomial GLM

Model comparison	Likelihood Ratio Test Statistic	Degrees of Freedom	Likelihood Ratio Test p-value	Conclusion
0 to 1	3.14	5	0.6784	Models 0 and 1 are statistically indistinguishable
0 to 2	16.83	6	0.0099	Model 0 outperforms Model 2
Null to 1	32468	25	0	Model 1 outperforms Null Model

Table B.3. Regression parameter estimates and p-values (second line of each cell) of power outage prediction models fitted by the negative binomial GLM with principal components for State A.

Model	Intercept	Hurricane Danny	Hurricane Dennis	Hurricane Georges	Hurricane Ivan	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12
0	-0.7811 <.0001	-2.0729 <.0001	-0.8812 <.0001	-1.2453 <.0001	0.1346 0.3722	0.3982 <.0001	-0.2007 <.0001	-0.2154 <.0001	0.0193 0.0368	0.0702 <.0001	0.1800 <.0001	-0.1811 <.0001	-0.1658 <.0001	-0.0031 0.8730	-0.2495 <.0001	0.0823 <.0001	0.1488 <.0001
1	-0.7159 <.0001	-2.1368 <.0001	-0.9362 <.0001	-1.3093 <.0001	NA	0.3984 <.0001	-0.2003 <.0001	-0.2372 <.0001	NA	0.0671 <.0001	0.1656 <.0001	-0.1696 <.0001	-0.1568 <.0001	NA	-0.2410 <.0001	0.0819 <.0001	0.1502 <.0001

PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24	PC25	PC26	Dispersion Parameter	D.F	Deviance
-0.0075 0.6142	0.1560 <.0001	-0.1240 <.0001	0.5176 <.0001	0.2762 <.0001	-0.9559 <.0001	0.1319 0.0047	-0.1616 <.0001	0.5595 <.0001	-0.3273 <.0001	-0.0179 0.8133	-0.4901 <.0001	0.3582 <.0001	-7.8450 0.0887	1.2239 0.0292	33374	18880.07
NA	0.1604 <.0001	-0.1191 <.0001	0.5173 <.0001	0.2730 <.0001	-0.9579 <.0001	0.1358 0.0023	-0.1736 <.0001	0.5469 <.0001	-0.3361 <.0001	NA	-0.4897 <.0001	0.3705 <.0001	NA	1.2234 0.0292	33380	18891.46

Table B.4. Model comparisons by likelihood ratio tests for the negative binomial GLM with principal components

Model comparison	Likelihood Ratio Test Statistic	Degrees of Freedom	Likelihood Ratio Test p-value	Conclusion
0 to 1	11.39	6	0.0770	Models 0 and 1 are statistically indistinguishable
Null to 1	32485	26	0	Model 1 outperforms Null Model

Table B.5. Regression parameter estimates and p-values (second line of each cell) of power outage prediction models fitted by the negative binomial GLM with principal components and alternative hurricane descriptors for State A.

Model	Intercept	Pressure	Time	RMW	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
0	-3.6630 <.0001	0.0330 <.0001	0.0158 <.0001	-0.0076 0.6775	0.3993 <.0001	-0.1810 <.0001	-0.1997 <.0001	0.0191 0.0380	0.0717 <.0001	0.1791 <.0001	-0.1738 <.0001	-0.1614 <.0001	0.0030 0.8761	-0.2385 <.0001	0.0817 <.0001	0.1492 <.0001	-0.0036 0.8094
1	-4.0528 <.0001	0.0349 <.0001	0.0155 <.0001	NA	0.3994 <.0001	-0.1803 <.0001	-0.1995 <.0001	0.0196 0.0312	0.0712 <.0001	0.1761 <.0001	-0.1727 <.0001	-0.1579 <.0001	NA	-0.2398 <.0001	0.0818 <.0001	0.1485 <.0001	NA
2	-4.1003 <.0001	0.0357 <.0001	0.0154 <.0001	NA	0.3995 <.0001	-0.1777 <.0001	-0.1999 <.0001	0.0195 0.0322	0.0702 <.0001	0.1694 <.0001	-0.1653 <.0001	-0.1504 <.0001	NA	-0.2383 <.0001	0.0815 <.0001	0.1480 <.0001	NA

Model	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24	PC25	PC26	Dispersion Parameter	D.F	Deviance
0	0.1431 <.0001	-0.1458 <.0001	0.5159 <.0001	0.2728 <.0001	-0.9548 <.0001	0.1612 0.0003	-0.1827 <.0001	0.5821 <.0001	-0.3049 <.0001	-0.1139 0.0706	-0.4860 <.0001	0.3599 <.0001	-7.8402 0.0888	1.2244 0.0292	33375	18882.60
1	0.1416 <.0001	-0.1486 <.0001	0.5148 <.0001	0.2680 <.0001	-0.9545 <.0001	0.1655 0.0001	-0.1828 <.0001	0.5866 <.0001	-0.3032 <.0001	-0.1178 0.0555	-0.4860 <.0001	0.3592 <.0001	NA	1.2247 0.0292	33379	18884.26
2	0.1440 <.0001	-0.1453 <.0001	0.5190 <.0001	0.2753 <.0001	-0.9553 <.0001	0.1748 <.0001	-0.1898 <.0001	0.5922 <.0001	-0.3065 <.0001	NA	-0.4849 <.0001	0.3599 <.0001	NA	1.2236 0.0292	33380	18894.18

Table B.6. Model comparisons by likelihood ratio tests for the negative binomial GLM with principal components and alternative hurricane descriptors

Model comparison	Likelihood Ratio Test Statistic	Degrees of Freedom	Likelihood Ratio Test p-value	Conclusion
0 to 1	1.66	4	0.7980	Models 0 and 1 are statistically indistinguishable
0 to 2	11.58	5	0.0410	Model 0 outperforms Model 2
Null to 1	32468	25	0	Model 1 outperforms Null Model

Table B.7. Regression parameter estimates and p-values (second line of each cell) of power outage prediction models fitted by the negative binomial GLM for State B.

Model	Intercept	Transformer	Switch	Overhead	Customer	Hurricane Dennis	Hurricane Ivan	Windspeed	Duration	FSM1	FSM2	FSM3	MAP	SPI1	SPI2	SPI3
0	-4.4153 0.9411	0.0035 <.0001	0.0042 0.0021	0.0507 <.0001	-0.0005 <.0001	0.1001 0.7583	2.2195 0.0217	-0.0323 0.0077	0.0858 <.0001	-0.2472 0.9726	-8.3381 0.0879	-7.9990 0.0703	0.0008 0.5439	0.6525 0.1552	0.1154 0.8482	0.9927 0.0280
1	-2.3418 <.0001	0.0035 <.0001	0.0043 0.0017	0.0514 <.0001	-0.0005 <.0001	NA	1.5050 <.0001	-0.0328 0.0040	0.0928 <.0001	NA	-9.6441 <.0001	-6.0298 0.0017	NA	0.6968 0.0048	NA	1.0037 <.0001

Model	SPI6	SPI12	SPI24	LC1	LC2	LC3	LC4	LC5	LC7	LC8	LC9	Dispersion Parameter	D.F	Deviance
0	-0.4846 0.5261	0.9391 0.0460	-0.7270 0.4858	0.0133 0.9822	0.0039 0.9948	0.0454 0.9395	0.0104 0.9862	0.0070 0.9907	-0.0045 0.9940	0.0114 0.9848	0.0111 0.9851	0.8020 0.0438	1779	1899.13
1	-1.0213 0.0312	0.8755 0.0032	NA	0.0093 <.0001	NA	0.0403 <.0001	0.0066 0.0184	NA	NA	0.0072 0.0345	0.0073 0.0059	0.8056 0.0438	1787	1897.93

Table B.8. Model comparisons by likelihood ratio tests for the negative binomial GLM

Model comparison	Likelihood Ratio Test Statistic	Degrees of Freedom	Likelihood Ratio Test p-value	Conclusion
0 to 1	1.2	8	0.9966	Models 0 and 1 are statistically indistinguishable
Null to 1	5194	18	0	Model 1 outperforms Null Model

Table B.9. Regression parameter estimates and p-values (second line of each cell) of power outage prediction models fitted by the negative binomial GLM with principal components for State B.

Model	Intercept	Hurricane Dennis	Hurricane Ivan	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14
0	0.3454 0.2185	0.1000 0.7587	2.2200 0.0217	0.6373 <.0001	0.1242 0.4632	-0.0899 0.0420	-0.0129 0.8880	-0.1504 0.0009	-0.0184 0.5336	-0.0353 0.3939	-0.2305 0.2616	0.0571 0.3795	0.0825 0.1291	0.0694 0.1797	-0.2552 0.0084	-0.1517 0.1843	-0.4965 0.0418
1	0.3178 <.0001	-0.3557 0.0138	2.7776 <.0001	0.6890 <.0001	NA	-0.0899 <.0001	NA	-0.1504 <.0001	NA	NA	-0.4604 <.0001	NA	0.1166 0.0009	NA	-0.1372 0.0208	NA	-0.6997 <.0001

Model	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24	Dispersion Parameter	D.F	Deviance
0	-0.2255 <.0001	-0.0431 0.6550	-0.8625 <.0001	0.5579 <.0001	-0.2895 0.1829	0.5033 0.0022	-0.1668 0.5538	0.5460 0.0059	0.8229 0.0495	-0.4636 0.9867	0.8020 0.0438	1779	1899.13
1	-0.2174 <.0001	NA	-0.8234 <.0001	0.5255 <.0001	-0.4526 0.0009	0.5493 0.0004	NA	0.5333 0.0069	0.8675 0.0352	NA	0.8104 0.0440	1789	1898.33

Table B.10. Model comparisons by likelihood ratio tests for the negative binomial GLM with principal components

Model comparison	Likelihood Ratio Test Statistic	Degrees of Freedom	Likelihood Ratio Test p-value	Conclusion
0 to 1	0.8	10	0.9999	Models 0 and 1 are statistically indistinguishable
Null to 1	5163	16	0	Model 1 outperforms Null Model

Table B.11. Regression parameter estimates and p-values (second line of each cell) of power outage prediction models fitted by the negative binomial GLM with principal components and alternative hurricane descriptors for State B.

Model	Intercept	Pressure	Time	RMW	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
0	-0.4320 0.8129	0.0080 0.6542	0.0342 0.0713	0.0000 0.0000	0.6373 <.0001	0.1242 0.4632	-0.0899 0.0420	-0.0129 0.8880	-0.1504 0.0009	-0.0184 0.5336	-0.0353 0.3939	-0.2305 0.2616	0.0571 0.3795	0.0825 0.1291	0.0694 0.1797	-0.2552 0.0084	-0.1517 0.1843
1	0.2053 0.0098	NA	0.0331 <.0001	NA	0.6395 <.0001	0.1141 0.0001	-0.1046 <.0001	NA	-0.1585 <.0001	NA	NA	-0.1680 <.0001	0.0655 0.0348	0.0718 0.0425	0.0887 0.0268	-0.3059 <.0001	-0.1846 0.0014

Model	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24	Dispersion Parameter	D.F	Deviance
0	-0.4965 0.0418	-0.2255 <.0001	-0.0431 0.6550	-0.8625 <.0001	0.5579 <.0001	-0.2895 0.1829	0.5033 0.0022	-0.1668 0.5538	0.5460 0.0059	0.8229 0.0495	-0.4636 0.9867	0.8020 0.0438	1779	1899.13
1	-0.4731 <.0001	-0.2315 <.0001	NA	-0.8565 <.0001	0.5664 <.0001	-0.2771 0.0421	0.4902 0.0017	NA	0.5624 0.0043	NA	NA	0.8097 0.0440	1787	1895.71

Table B.12. Model comparisons by likelihood ratio tests for the negative binomial GLM with principal components and alternative hurricane descriptors

Model comparison	Likelihood Ratio Test Statistic	Degrees of Freedom	Likelihood Ratio Test p-value	Conclusion
0 to 1	3.42	8	0. 9053	Models 0 and 1 are statistically indistinguishable
Null to 1	5170	18	0	Model 1 outperforms Null Model

Table B.13. Regression parameter estimates and p-values (second line of each cell) of power outage prediction models fitted by the negative binomial GLM for State C.

Model	Intercept	Transformer	Pole	Switch	Overhead	Customer	Hurricane Cindy	Hurricane Dennis	Hurricane Frances	Hurricane Hanna	Hurricane Isidore	Hurricane Ivan	Hurricane Jeanne	Windspeed	Duration	FSM1	FSM2
0	10.2560 0.6895	0.0003 0.3762	0.0008 <.0001	-0.0010 0.1608	0.0286 <.0001	-0.0002 <.0001	-1.0152 <.0001	-0.0497 0.6818	1.8215 <.0001	-0.5056 0.0374	-0.9249 <.0001	0.8020 0.0014	0.6847 0.0065	0.0843 <.0001	-0.0044 0.3933	4.7225 0.0006	-6.0042 <.0001
1	-7.2597 <.0001	NA	0.0007 <.0001	NA	0.0291 <.0001	-0.0002 <.0001	-0.9896 <.0001	NA	1.8580 <.0001	-0.5309 0.0180	-0.9126 <.0001	0.8454 <.0001	0.6875 0.0019	0.0806 <.0001	NA	4.3918 0.0009	-5.9223 <.0001
2	-7.2739 <.0001	NA	0.0008 <.0001	NA	0.0289 <.0001	-0.0002 <.0001	-0.9916 <.0001	NA	1.8894 <.0001	-0.4944 0.0267	-0.8726 <.0001	-0.8726 <.0001	0.7196 0.0011	0.0806 <.0001	NA	4.4035 0.0009	-5.9560 <.0001

FSM3	FSM3	MAP	SPI1	SPI2	SPI3	SPI6	SPI12	SPI24	LC1	LC2	LC3	LC4	LC5	LC7	LC8	LC9	Dispersion Parameter	D.F	Deviance
0.9858 0.1461	0.9858 0.1461	0.0018 <.0001	0.4746 <.0001	-0.5462 <.0001	0.1286 0.3660	0.0269 0.8511	0.4755 <.0001	-0.5558 <.0001	-0.1708 0.5058	-0.1770 0.4904	-0.2001 0.4363	-0.1845 0.4721	-0.1718 0.5034	-0.2031 0.4291	-0.1881 0.4636	-0.1759 0.4932	2.4338 0.0975	58607	10843.42
1.0369 0.1023	1.0369 0.1023	0.0018 <.0001	0.4692 <.0001	-0.5773 <.0001	0.1588 0.0973	NA	0.4785 <.0001	-0.5500 <.0001	NA	NA	-0.0219 0.1384	-0.0085 <.0001	NA	-0.0278 <.0001	-0.0124 <.0001	NA	2.4279 0.0972	58616	10858.99
1.0475 0.0988	1.0475 0.0988	0.0018 <.0001	0.4580 <.0001	-0.5727 <.0001	0.1523 0.1113	NA	0.4947 <.0001	-0.5476 <.0001	NA	NA	NA	-0.0083 <.0001	NA	-0.0285 <.0001	-0.0121 <.0001	NA	2.4252 0.0972	58617	10864.94

Table B.14. Model comparisons by likelihood ratio tests for the negative binomial GLM

Model comparison	Likelihood Ratio Test Statistic	Degrees of Freedom	Likelihood Ratio Test p-value	Conclusion
0 to 1	15.57	9	0.0764	Models 0 and 1 are statistically indistinguishable
0 to 2	21.52	10	0.0177	Model 0 outperforms Model 2
Null to 1	17507	23	0	Model 1 outperforms Null Model

Table B.15. Regression parameter estimates and p-values (second line of each cell) of power outage prediction models fitted by the negative binomial GLM with principal components for State C.

Model	Intercept	Hurricane Cindy	Hurricane Dennis	Hurricane Frances	Hurricane Hanna	Hurricane Isidore	Hurricane Ivan	Hurricane Jeanne	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
0	-3.9582 <.0001	-1.0152 <.0001	-0.0498 0.6817	1.8215 <.0001	-0.5056 0.0374	-0.9249 <.0001	0.8020 0.0014	0.6847 0.0065	0.4834 <.0001	-0.0530 0.0454	0.0292 0.5529	0.0562 0.0411	0.0298 0.4468	-0.0310 0.2218	-0.1222 <.0001	-0.1834 <.0001	-0.2172 <.0001
1	-4.0207 <.0001	-0.8241 <.0001	NA	1.8441 <.0001	NA	-0.8972 <.0001	0.6093 <.0001	0.6406 <.0001	0.4984 <.0001	NA	NA	NA	NA	NA	-0.1243 <.0001	-0.2039 <.0001	-0.2265 <.0001

Model	PC10	PC11	PC12	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24	PC25	Dispersion Parameter	D.F	Deviance
0	0.2318 <.0001	-0.0951 0.0002	0.1199 0.0003	0.0899 0.0233	-0.0336 0.3473	-0.1152 0.0683	0.1495 <.0001	-0.6133 <.0001	-0.2730 0.0126	-0.4367 <.0001	0.0655 0.1886	-0.8205 <.0001	-0.4400 <.0001	-0.1216 0.3693	0.3457 <.0001	6.4397 0.4774	2.4338 0.0975	58607	10843.42
1	0.2209 <.0001	-0.0986 <.0001	0.1146 0.0003	0.1372 <.0001	NA	NA	0.1537 <.0001	-0.5934 <.0001	-0.1989 0.0107	-0.4350 <.0001	NA	-0.8990 <.0001	-0.4385 <.0001	NA	0.3154 0.0002	NA	2.4466 0.0973	58619	10844.04

Table B.16. Model comparisons by likelihood ratio tests for the negative binomial GLM with principal components

Model comparison	Likelihood Ratio Test Statistic	Degrees of Freedom	Likelihood Ratio Test p-value	Conclusion
0 to 1	0.62	12	1	Models 0 and 1 are statistically indistinguishable
Null to 1	17454	20	0	Model 1 outperforms Null Model

Table B.17. Regression parameter estimates and p-values (second line of each cell) of power outage prediction models fitted by the negative binomial GLM with principal components and alternative hurricane descriptors for State C.

Model	Intercept	Pressure	Time	RMW	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
0	-0.7634 0.4704	0.0036 0.4569	0.0344 <.0001	-0.0962 <.0001	0.5446 <.0001	0.1424 <.0001	-0.1535 <.0001	-0.0513 0.0466	0.1274 <.0001	-0.1132 <.0001	-0.1115 <.0001	-0.2638 <.0001	-0.2474 <.0001	0.2129 <.0001	-0.1120 <.0001	0.1893 <.0001	0.0046 0.9014
1	-0.1817 0.4609	NA	0.0306 <.0001	-0.1056 <.0001	0.5434 <.0001	0.1413 <.0001	-0.1425 <.0001	-0.0508 0.0143	0.1303 <.0001	-0.1159 <.0001	-0.1107 <.0001	-0.2648 <.0001	-0.2492 <.0001	0.2270 <.0001	-0.1102 <.0001	0.1905 <.0001	NA

Model	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24	PC25	Dispersion Parameter	D.F	Deviance
0	-0.1171 0.0005	0.4255 <.0001	0.2060 <.0001	-0.6180 <.0001	-0.0965 0.2454	-0.4438 <.0001	0.1329 0.0078	-0.7151 <.0001	-0.1050 0.2789	0.5223 <.0001	0.3307 0.0002	6.2503 0.4936	2.6130 0.1018	58611	10782.40
1	-0.1216 <.0001	0.4286 <.0001	0.2025 <.0001	-0.6235 <.0001	NA	-0.4387 <.0001	0.1231 0.0111	-0.7430 <.0001	NA	0.5018 <.0001	0.3328 0.0002	NA	2.6223 0.1019	58616	10773.90

Table B.18. Model comparisons by likelihood ratio tests for the negative binomial GLM with principal components and alternative hurricane descriptors

Model comparison	Likelihood Ratio Test Statistic	Degrees of Freedom	Likelihood Ratio Test p-value	Conclusion
0 to 1	8.50	5	0.1307	Models 0 and 1 are statistically indistinguishable
Null to 1	16906	23	0	Model 1 outperforms Null Model

Table B.19. Regression parameter estimates and p-values (second line of each cell) of customers out prediction models fitted by the negative binomial GLM with principal components for State A

Model	Intercept	Hurricane Danny	Hurricane Dennis	Hurricane Georges	Hurricane Ivan	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12
0	2.8561 <.0001	-1.7949 <.0001	-0.8918 <.0001	-0.4089 0.0106	0.3154 0.2938	0.8181 <.0001	-0.3689 <.0001	-0.2672 <.0001	0.1039 <.0001	-0.0387 0.1288	0.3443 <.0001	-0.1983 <.0001	-0.2052 <.0001	0.0351 0.3231	-0.3838 <.0001	0.2269 <.0001	0.3144 0.0006
1	2.9617 <.0001	-1.9056 <.0001	-0.9951 <.0001	-0.3993 0.0032	NA	0.8146 <.0001	-0.3861 <.0001	-0.3238 <.0001	0.1098 <.0001	-0.0527 0.0186	0.3099 <.0001	-0.1956 <.0001	-0.1793 <.0001	NA	-0.3945 <.0001	0.2309 <.0001	0.3115 <.0001

Model	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24	PC25	PC26	Dispersion Parameter	D.F	Deviance
0	-0.1241 0.0001	0.2723 <.0001	-0.1217 0.0126	0.8244 <.0001	0.5283 <.0001	-1.6372 <.0001	0.0607 0.4476	0.2995 0.0008	0.9909 <.0001	-1.2541 <.0001	-0.5616 0.0001	-0.5261 0.0020	3.1802 <.0001	-13.3615 0.1744	17.2218 0.2179	33374	16641.46
1	-0.1289 <.0001	0.2682 <.0001	-0.1195 0.0109	0.8134 <.0001	0.4870 <.0001	-1.6416 <.0001	NA	0.3027 0.0007	0.9534 <.0001	-1.2803 <.0001	-0.4681 0.0005	-0.5273 0.0017	3.1508 <.0001	NA	17.2275 0.2180	33378	16642.08

Table B.20. Model comparisons by likelihood ratio tests for the negative binomial GLM with principal components

Model comparison	Likelihood Ratio Test Statistic	Degrees of Freedom	Likelihood Ratio Test p-value	Conclusion
0 to 1	0.62	4	0.9608	Models 0 and 1 are statistically indistinguishable
Null to 1	8541	26	0	Model 1 outperforms Null Model

Table B.21. Regression parameter estimates and p-values (second line of each cell) of customers out prediction models fitted by the negative binomial GLM with principal components and alternative hurricane descriptors for State A

Model	Intercept	Pressure	Time	RMW	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
0	-5.6783 0.0001	0.0570 <.0001	0.0183 <.0001	0.1110 0.0002	0.8280 <.0001	-0.2579 <.0001	-0.1570 0.0020	0.0995 <.0001	-0.0207 0.4131	0.3191 <.0001	-0.1392 <.0001	-0.1765 <.0001	0.0799 0.0207	-0.3119 <.0001	0.2137 <.0001	0.3084 <.0001	-0.0845 0.0080
1	-4.6673 0.0009	0.0516 <.0001	0.0227 <.0001	0.0898 0.0014	0.8262 <.0001	-0.2553 <.0001	-0.1106 0.0124	0.1045 <.0001	NA	0.3587 <.0001	-0.1602 <.0001	-0.2089 <.0001	0.1092 0.0007	-0.3170 <.0001	0.2118 <.0001	0.3148 <.0001	-0.0905 0.0041

Model	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24	PC25	PC26	Dispersion Parameter	D.F	Deviance
0	0.2217 <.0001	-0.2424 <.0001	0.8398 <.0001	0.5607 <.0001	-1.6194 <.0001	0.1563 0.0433	0.1818 0.0326	1.1242 <.0001	-1.1152 <.0001	-1.0035 <.0001	-0.5440 0.0017	3.2671 <.0001	-13.2646 0.1792	17.2569 0.2183	33375	16641.40
1	0.2255 <.0001	-0.2233 <.0001	0.8460 <.0001	0.6025 <.0001	-1.6132 <.0001	NA	0.1856 0.0292	1.1305 <.0001	-1.0950 <.0001	-1.0102 <.0001	-0.5211 0.0023	3.2732 <.0001	NA	17.2638 0.2184	33378	16642.40

Table B.22. Model comparisons by likelihood ratio tests for the negative binomial GLM with principal components and alternative hurricane descriptors

Model comparison	Likelihood Ratio Test Statistic	Degrees of Freedom	Likelihood Ratio Test p-value	Conclusion
0 to 1	1	3	0.8013	Models 0 and 1 are statistically indistinguishable
Null to 1	8511	26	0	Model 1 outperforms Null Model

Table B.23. Regression parameter estimates and p-values (second line of each cell) of customers out prediction models fitted by the negative binomial GLM with principal components for State B

Model	Intercept	Hurricane Dennis	Hurricane Ivan	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14
0	4.9687 <.0001	-1.2182 0.1110	0.8776 0.7117	0.8430 <.0001	0.2972 0.4703	-0.1201 0.2747	-0.2798 0.2026	-0.1839 0.0935	-0.1541 0.0351	0.1029 0.2989	-0.6821 0.1578	-0.0125 0.9338	0.1834 0.1645	-0.6943 <.0001	0.4313 0.0627	-0.1098 0.7001	-0.5604 0.3562
1	4.4521 <.0001	-1.3833 0.0002	2.6464 <.0001	0.9476 <.0001	NA	-0.1918 0.0002	-0.1034 0.0313	-0.2373 <.0001	-0.1675 0.0135	NA	-0.9359 <.0001	NA	0.2980 0.0002	-0.6471 <.0001	0.5653 <.0001	NA	-1.0038 <.0001

Model	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24	Dispersion Parameter	D.F	Deviance
0	-0.5134 0.0001	0.2013 0.4303	-2.2484 <.0001	1.3420 <.0001	-1.1597 0.0395	0.5732 0.1400	1.7136 0.0023	-1.2936 0.0219	0.1192 0.8867	43.7555 0.5115	6.0183 0.2064	1779	1890.62
1	-0.4350 0.0011	NA	-2.1735 <.0001	1.3090 <.0001	-1.4093 <.0001	NA	2.0308 <.0001	-1.3711 0.0137	NA	NA	6.0567 0.2076	1787	1890.80

Table B.24. Model comparison by likelihood ratio for the negative binomial GLM with principal components

Model comparison	Likelihood Ratio Test Statistic	Degrees of Freedom	Likelihood Ratio Test p-value	Conclusion
0 to 1	0.18	8	1	Models 0 and 1 are statistically indistinguishable
Null to 1	1220	18	0	Model 1 outperforms Null Model

Table B.25. Regression parameter estimates and p-values (second line of each cell) of customers out prediction models fitted by the negative binomial GLM with principal components and alternative hurricane descriptors for State B

Model	Intercept	Pressure	Time	RMW	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
0	-0.5107 0.9066	0.0586 0.1657	0.0338 0.4630	0.0000 0.0000	0.8430 <.0001	0.2972 0.4703	-0.1201 0.2748	-0.2798 0.2027	-0.1839 0.0935	-0.1541 0.0351	0.1029 0.2989	-0.6821 0.1578	-0.0125 0.9338	0.1834 0.1645	-0.6943 <.0001	0.4313 0.0627	-0.1098 0.7001
1	-2.6530 0.0421	0.0757 <.0001	0.0650 <.0001	NA	0.9476 <.0001	NA	-0.1918 0.0002	-0.1034 0.0313	-0.2373 <.0001	-0.1675 0.0135	NA	-0.9359 <.0001	NA	0.2980 0.0002	-0.6471 <.0001	0.5653 <.0001	NA

Model	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24	Dispersion Parameter	D.F	Deviance
0	-0.5604 0.3562	-0.5134 0.0001	0.2013 0.4303	-2.2484 <.0001	1.3420 <.0001	-1.1597 0.0395	0.5732 0.1400	1.7136 0.0023	-1.2936 0.0219	0.1192 0.8867	43.7555 0.5115	6.0183 0.2064	1779	1890.54
1	-1.0038 <.0001	-0.4350 0.0011	NA	-2.1735 <.0001	1.3090 <.0001	-1.4093 <.0001	NA	2.0308 <.0001	-1.3711 0.0137	NA	NA	6.0567 0.2076	1787	1890.80

Table B.26. Model comparisons by likelihood ratio tests for the negative binomial GLM with principal components and alternative hurricane descriptors

Model comparison	Likelihood Ratio Test Statistic	Degrees of Freedom	Likelihood Ratio Test p-value	Conclusion
0 to 1	0.27	8	1	Models 0 and 1 are statistically indistinguishable
Null to 1	1200	18	0	Model 1 outperforms Null Model

Table B.27. Regression parameter estimates and p-values (second line of each cell) of customers out prediction models fitted by the negative binomial GLM with principal components for State C

Model	Intercept	Hurricane Cindy	Hurricane Dennis	Hurricane Frances	Hurricane Hanna	Hurricane Isidore	Hurricane Ivan	Hurricane Jeanne	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
0	-3.3545 <.0001	-1.2774 <.0001	-0.7539 <.0001	1.1710 <.0001	1.1225 <.0001	-0.3681 0.1066	-0.4522 0.1279	-0.0866 0.7704	0.7086 <.0001	0.0771 0.0097	-0.0744 0.2019	-0.1145 <.0001	0.0616 0.1760	-0.0341 0.2507	-0.1582 <.0001	-0.3529 <.0001	-0.1488 <.0001	0.1740 0.0004
1	-3.5512 <.0001	-1.2517 <.0001	-0.7810 <.0001	1.6089 <.0001	1.6193 <.0001	NA	-0.2370 0.0192	NA	0.7163 <.0001	0.1102 <.0001	NA	-0.1434 <.0001	NA	NA	-0.1564 <.0001	-0.3526 <.0001	-0.1431 <.0001	0.1758 <.0001

Model	PC11	PC12	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24	PC25	Dispersion Parameter	D.F	Deviance
0	-0.0837 0.0079	0.1279 0.0006	0.1995 <.0001	-0.1028 0.0178	0.2051 0.0027	0.3814 <.0001	-0.8615 <.0001	-0.5790 <.0001	-0.6898 <.0001	0.3398 <.0001	-0.5201 <.0001	0.2113 0.0730	0.7199 <.0001	0.6474 <.0001	-3.7056 0.7420	10.3729 0.2763	58607	9006.01
1	-0.0669 0.0249	0.1035 0.0043	0.2894 <.0001	NA	0.1841 0.0003	0.3553 <.0001	-0.8537 <.0001	-0.7334 <.0001	-0.6893 <.0001	0.3276 <.0001	-0.5257 <.0001	NA	0.6525 <.0001	0.6737 <.0001	NA	10.3721 0.2761	58615	9019.75

Table B.28. Model comparisons by likelihood ratio tests for the negative binomial GLM with principal components

Model comparison	Likelihood Ratio Test Statistic	Degrees of Freedom	Likelihood Ratio Test p-value	Conclusion
0 to 1	13.74	8	0.0888	Models 0 and 1 are statistically indistinguishable
Null to 1	20237	24	0	Model 1 outperforms Null Model

Table B.29. Regression parameter estimates and p-values (second line of each cell) of customers out prediction models fitted by the negative binomial GLM with principal components and alternative hurricane descriptors for State C

Model	Intercept	Pressure	Time	RMW	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
0	3.3640 0.0072	-0.0150 0.0073	0.0417 <.0001	-0.1723 <.0001	0.7039 <.0001	0.0515 0.0085	-0.2346 <.0001	-0.1336 <.0001	0.0843 0.0089	-0.0025 0.9242	-0.1764 <.0001	-0.3929 <.0001	-0.1832 <.0001	0.1760 <.0001	-0.1115 0.0002	0.1629 <.0001	0.0862 0.0353
1	3.3864 0.0059	-0.0151 0.0065	0.0418 <.0001	-0.1728 <.0001	0.7040 <.0001	0.0515 0.0085	-0.2349 <.0001	-0.1341 <.0001	0.0840 0.0089	NA	-0.1769 <.0001	-0.3935 <.0001	-0.1831 <.0001	0.1758 <.0001	-0.1114 0.0002	0.1627 <.0001	0.0864 0.0338

Model	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24	PC25	Dispersion Parameter	D.F	Deviance
0	-0.2143 <.0001	0.2777 <.0001	0.3925 <.0001	-0.8897 <.0001	-0.3459 0.0004	-0.6988 <.0001	0.3613 <.0001	-0.5100 <.0001	0.3206 0.0040	0.8623 <.0001	0.6252 <.0001	-5.1540 0.6475	10.4552 0.2786	58611	9011.68
1	-0.2142 <.0001	0.2777 <.0001	0.3925 <.0001	-0.8899 <.0001	-0.3474 0.0004	-0.6994 <.0001	0.3606 <.0001	-0.5119 <.0001	0.3189 0.0041	0.8624 <.0001	0.6254 <.0001	NA	10.4557 0.2786	58613	9011.65

Table B.30. Model comparisons by likelihood ratio tests for the negative binomial GLM with principal components and alternative hurricane descriptors

Model comparison	Likelihood Ratio Test Statistic	Degrees of Freedom	Likelihood Ratio Test p-value	Conclusion
0 to 1	0.03	2	0.9851	Models 0 and 1 are statistically indistinguishable
Null to 1	20089	26	0	Model 1 outperforms Null Model

Table B.31. Regression parameter estimates and p-values (second line of each cell) of damaged pole estimation models fitted by the negative binomial GLM with principal components for State A

Model	Intercept	Hurricane Dennis	Hurricane Ivan	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
0	-6.7831 <.0001	-2.9295 <.0001	4.5204 <.0001	-0.0383 <.0001	0.3533 <.0001	0.4286 <.0001	0.3568 <.0001	0.6025 <.0001	0.0011 0.9436	0.6869 <.0001	-0.3823 <.0001	-0.6047 <.0001	0.2812 <.0001	-0.4088 <.0001	0.0327 0.1512	-0.1653 <.0001
1	-6.7504 <.0001	-2.9599 <.0001	4.4636 <.0001	-0.0382 <.0001	0.3474 <.0001	0.4192 <.0001	0.3544 <.0001	0.6020 <.0001	NA	0.6853 <.0001	-0.3801 <.0001	-0.6009 <.0001	0.2785 <.0001	-0.4099 <.0001	NA	-0.1661 <.0001
2	-6.7585 <.0001	-2.9548 <.0001	4.4844 <.0001	-0.0383 <.0001	0.3517 <.0001	0.4225 <.0001	0.3546 <.0001	0.6010 <.0001	NA	0.6886 <.0001	-0.3806 <.0001	-0.6034 <.0001	0.2808 <.0001	-0.4174 <.0001	NA	-0.1662 <.0001

Model	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24	PC25	PC26	Dispersion Parameter	D.F	Deviance
0	-0.0155 0.7929	0.9198 <.0001	-0.0344 0.4818	-0.2400 <.0001	0.9298 <.0001	0.2885 <.0001	2.1946 <.0001	1.0273 <.0001	0.0270 0.7824	0.0239 0.8451	0.3684 0.0066	0.0914 0.5659	4.2791 0.6034	0.5606 0.0181	2173	2489.38
1	NA	0.9181 <.0001	NA	-0.2363 <.0001	0.9354 <.0001	0.2911 <.0001	2.2039 <.0001	1.0315 <.0001	NA	NA	0.3646 0.0070	NA	NA	0.5622 0.0181	2181	2486.95
2	NA	0.9198 <.0001	NA	-0.2365 <.0001	0.9312 <.0001	0.2813 <.0001	2.2038 <.0001	1.0371 <.0001	NA	NA	NA	NA	NA	0.5631 0.0181	2182	2491.02

Table B32. Model comparisons by likelihood ratio tests for the negative binomial GLM with principal components

Model comparison	Likelihood Ratio Test Statistic	Degrees of Freedom	Likelihood Ratio Test p-value	Conclusion
0 to 1	2.43	8	0.9649	Models 0 and 1 are statistically indistinguishable
0 to 2	1.64	9	0.9901	Models 0 and 2 are statistically indistinguishable
Null to 1	353.54	20	0	Model 1 outperforms Null Model

Table B.33. Regression parameter estimates and p-values (second line of each cell) of damaged pole estimation models fitted by the negative binomial GLM with principal components and alternative hurricane descriptors for State A

Model	Intercept	Pressure	Time	RMW	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12
0	-21.2502 <.0001	0.1543 <.0001	0.1202 <.0001	0.0000 <.0001	-0.0383 <.0001	0.3533 <.0001	0.4286 <.0001	0.3568 <.0001	0.6025 <.0001	0.0011 0.9436	0.6869 <.0001	-0.3823 <.0001	-0.6047 <.0001	0.2812 <.0001	-0.4088 <.0001	0.0327 0.1512
1	-21.3121 <.0001	0.1553 <.0001	0.1197 <.0001	0.0000 <.0001	-0.0382 <.0001	0.3474 <.0001	0.4192 <.0001	0.3544 <.0001	0.6020 <.0001	NA	0.6853 <.0001	-0.3801 <.0001	-0.6009 <.0001	0.2785 <.0001	-0.4099 <.0001	NA

Model	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24	PC25	PC26	Dispersion Parameter	D.F	Deviance
0	-0.1653 <.0001	-0.0155 0.7929	0.9198 <.0001	-0.0344 0.4818	-0.2400 <.0001	0.9298 <.0001	0.2885 <.0001	2.1946 <.0001	1.0273 <.0001	0.0270 0.7824	0.0239 0.8451	0.3684 0.0066	0.0914 0.5659	4.2791 0.6034	0.5606 0.0181	2173	2489.38
1	-0.1661 <.0001	NA	0.9181 <.0001	NA	-0.2363 <.0001	0.9354 <.0001	0.2911 <.0001	2.2039 <.0001	1.0315 <.0001	NA	NA	0.3646 0.0070	NA	NA	0.5622 0.0181	2181	2486.95

Table B.34. Model comparisons by likelihood ratio tests for the negative binomial GLM with principal components and alternative hurricane descriptors

Model comparison	Likelihood Ratio Test Statistic	Degrees of Freedom	Likelihood Ratio Test p-value	Conclusion
0 to 1	2.43	8	0.9649	Models 0 and 1 are statistically indistinguishable
Null to 1	353.54	20	0	Model 1 outperforms Null Model

Table B.35. Regression parameter estimates and p-values (second line of each cell) of damaged transformer estimation models fitted by the negative binomial GLM with principal components for State A

Model	Intercept	Hurricane Dennis	Hurricane Ivan	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13
0	-6.1642 <.0001	-2.2920 <.0001	3.7386 <.0001	-0.0133 0.1864	0.2961 <.0001	0.4754 <.0001	0.3714 <.0001	0.5087 <.0001	0.0163 0.4008	0.6039 <.0001	-0.3213 <.0001	-0.2441 <.0001	0.2090 <.0001	-0.4163 <.0001	0.0570 0.0284	-0.1683 <.0001
1	-6.0901 <.0001	-2.4810 <.0001	3.7353 <.0001	NA	0.3204 <.0001	0.4682 <.0001	0.3769 <.0001	0.5216 <.0001	NA	0.6256 <.0001	-0.3279 <.0001	-0.2532 <.0001	0.2186 <.0001	-0.4366 <.0001	0.0661 0.0121	-0.1803 <.0001
2	-6.0910 <.0001	-2.4567 <.0001	3.7137 <.0001	NA	0.3133 <.0001	0.4654 <.0001	0.3714 <.0001	0.5174 <.0001	NA	0.6221 <.0001	-0.3230 <.0001	-0.2473 <.0001	0.2132 <.0001	-0.4341 <.0001	NA	-0.1775 <.0001

Model	PC14	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24	PC25	PC26	Dispersion Parameter	D.F	Deviance
0	-0.0824 0.2122	0.8877 <.0001	-0.3837 <.0001	-0.3617 <.0001	0.3881 <.0001	-0.0729 0.2603	1.9180 <.0001	0.9563 <.0001	0.0055 0.9591	0.7834 <.0001	-0.0307 0.8402	-0.1738 0.3185	-1.3852 0.8831	0.7135 0.0228	2173	2556.11
1	NA	0.9341 <.0001	-0.3722 <.0001	-0.3474 <.0001	0.4229 <.0001	NA	1.9979 <.0001	1.0273 <.0001	NA	0.7852 <.0001	NA	NA	NA	0.7151 0.0228	2181	2556.91
2	NA	0.9293 <.0001	-0.3674 <.0001	-0.3577 <.0001	0.4237 <.0001	NA	1.9871 <.0001	1.0150 <.0001	NA	0.7940 <.0001	NA	NA	NA	0.71651 0.0228	2182	2560.36

Table B.36. Model comparisons by likelihood ratio tests for the negative binomial GLM with principal components

Model comparison	Likelihood Ratio Test Statistic	Degrees of Freedom	Likelihood Ratio Test p-value	Conclusion
0 to 1	0.8	8	0.9992	Models 0 and 1 are statistically indistinguishable
0 to 2	4.25	9	0.8942	Models 0 and 2 are statistically indistinguishable
Null to 1	271.53	20	0	Model 1 outperforms Null Model

Table B.37. Regression parameter estimates and p-values (second line of each cell) of damaged transformer estimation models fitted by the negative binomial GLM with principal components and alternative hurricane descriptors for State A

Model	Intercept	Pressure	Time	RMW	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14
0	-17.5913 <.0001	0.1218 <.0001	0.0973 <.0001	0.0000 <.0001	-0.0133 0.1864	0.2961 <.0001	0.4754 <.0001	0.3714 <.0001	0.5087 <.0001	0.0163 0.4008	0.6039 <.0001	-0.3213 <.0001	-0.2441 <.0001	0.2090 <.0001	-0.4163 <.0001	0.0570 0.0284	-0.1683 <.0001	-0.0824 0.2122
1	-18.2926 <.0001	0.1301 <.0001	0.1003 <.0001	0.0000 <.0001	NA	0.3204 <.0001	0.4682 <.0001	0.3769 <.0001	0.5216 <.0001	NA	0.6256 <.0001	-0.3279 <.0001	-0.2532 <.0001	0.2186 <.0001	-0.4366 <.0001	0.0661 0.0121	-0.1803 <.0001	NA

Model	PC15	PC16	PC17	PC18	PC19	PC20	PC21	PC22	PC23	PC24	PC25	PC26	Dispersion Parameter	D.F	Deviance
0	0.8877 <.0001	-0.3837 <.0001	-0.3617 <.0001	0.3881 <.0001	-0.0729 0.2603	1.9180 <.0001	0.9563 <.0001	0.0055 0.9591	0.7834 <.0001	-0.0307 0.8402	-0.1738 0.3185	-1.3852 0.8831	0.7135 0.0228	2173	2556.11
1	0.9341 <.0001	-0.3722 <.0001	-0.3474 <.0001	0.4229 <.0001	NA	1.9979 <.0001	1.0273 <.0001	NA	0.7852 <.0001	NA	NA	NA	0.7151 0.0228	2181	2556.91

Table B.38. Model comparisons by likelihood ratio tests for the negative binomial GLM with principal components and alternative hurricane descriptors

Model comparison	Likelihood Ratio Test Statistic	Degrees of Freedom	Likelihood Ratio Test p-value	Conclusion
0 to 1	0.8	8	0.9992	Models 0 and 1 are statistically indistinguishable
Null to 1	271.53	20	0	Model 1 outperforms Null Model

VITA

Seung Ryong Han received his Bachelor of Science degree in civil engineering from KonKuk University at Seoul in February 2000. He entered the Civil & Environmental Engineering Department at Korea University in March 2000 and received his Master of Science degree in February 2002. His research specialty includes ductility of composite columns in structural engineering. He entered the Civil Engineering Ph.D. program at Texas A&M University in August 2003 and received his Doctor of Philosophy degree in August 2008. His research interests include hurricane risks, risk analysis, and structural reliability.

Name: Seung Ryong Han

Address: Structural Engineering Division
3136 TAMU
College Station, TX 77843-3136

Email Address: hoanung@hanmail.net