Estimating Power Outage Risk during Hurricanes in the Gulf Coast Region

Seung Ryong Han[i], Seth Guikema[ii], Steven Quiring[iii], Rachel Davidson[iv], Kyung Ho Lee[v], David Rosowsky[vi]

---

[i] Ph.D. Candidate, Zachry Department of Civil Engineering, Texas A&M University, College Station, Texas 77843

[ii] Assistant Professor, Zachry Department of Civil Engineering, Texas A&M University, College Station, Texas 77843

[iii] Assistant Professor, Department of Geography, Texas A&M University, College Station, Texas 77843

[iv] Assistant Professor, School of Civil and Environmental Engineering, Cornell University, Ithaca, New York 14853

[v] Senior Engineer, Energo Engineering, Inc. Houston, TX 77042

[vi] Department Chair, Zachry Department of Civil Engineering, Texas A&M University, College Station, Texas 77843

Abstract

Hurricanes have caused severe damage to the electric power system throughout the Atlantic and Gulf coast regions of the U.S. Electric power is critical to post-hurricane disaster response as well as to long-term recovery for impacted areas. Managing power outage risk and properly preparing for post-storm recovery efforts requires rigorous methods for estimating the number and location of power outages. This paper presents models for predicting the number of power outages in each grid cell of a utility company's service area based on Negative Binomial generalized linear regression models. The models were developed for a large, investor-owned utility company serving a sizeable portion of the Gulf Coast region. The most influential explanatory variables in the final regression model are the maximum gust wind speed, duration of strong winds, the miles of overhead line, the soil moisture before the storm, and other factors about the electric power distribution system. The outage count prediction model can provide a basis for choosing how many crews and materials to request from other utility companies in order to restore electric power as quickly as possible without incurring unnecessary expense.

INTRODUCTION

Hurricanes have caused severe power interruption throughout the Atlantic and Gulf Coast regions of the U.S. In recent years, for example, the central Gulf Coast region (i.e., Alabama, Mississippi, and Louisiana) has been significantly impacted by Hurricanes Opal (1995), Danny (1997), Georges (1998), Ivan (2004), Dennis (2005), and Katrina (2005). In addition to causing considerable direct repair and restoration costs for utility companies, hurricane-related power outages may result in loss of services from a number of other critical infrastructure systems such as water, transportation, communication, banking, and security systems.

This paper develops statistical models for predicting the number of power outages in each 12,000 foot by 8,000 foot grid cell of a company's service area for an approaching hurricane. These models are based on information about the hurricane, the power system, and the local geography. An outage is defined in this paper as the closure of protective device in the power system. A single outage may impact one customer or hundreds of customers.

Our models are based on data supplied by a large, investor-owned utility company serving the Gulf Coast region. We considered Poisson Generalized Linear Models and Negative

Binomial Generalized Linear Models for the predictive models. The Poisson Generalized Linear Model is appropriate for regression of count data, but it assumes that the variance of the event counts equals the mean the event counts. However, overdispersion, variance greater than the mean, often occurs. One way to manage overdispersion is to use the Negative Binomial GLM. These regression models provide outage count predictions in the service area of the utility company we are working with. These predictions can help the utility company better manage the effects of hurricanes by pre-positioning and deploying repair personnel and materials prior to a hurricane making landfall.

DESCRIPTION OF DATA

Power outage occurrence during hurricanes depends on a number of factors that influence the vulnerability of electric power system to outages. Some of these factors include the miles of distribution line and the number of poles, switches, transformers, customers and soil moisture levels in different geographic areas. Power outages during a hurricane are also dependent on the intensity and duration of the hurricane, and we use information from a hurricane wind field simulation model to capture these effects.

Our model is based on data provided by a large, investor-owned utility company in the Gulf Coast region. In our model, the service area of this utility company was covered with 12,000 foot by 8,000 foot grid cells, and these grid cells form the unit of analysis in our work. There are 6,681 grid cells in the service area. We used data on outages in the service area during the following 5 hurricanes: Danny (627 outages), Dennis (4,840 outages), Georges (1,705 outages), Ivan (13,568 outages), and Katrina (10,105 outages). This outage data was combined with information about the power system, geographic characteristics of the service area, and hurricane characteristics from a hurricane wind field simulation model and publicly available hurricane data.

Hurricane Characteristic Data

In order to capture the characteristics of the wind field during a given hurricane, we used estimates of the maximum 3-second gust wind speed and the length of time that the winds were above 20 m/s for each grid cell. We used 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 wind field 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 wind speed throughout the duration of a hurricane at the lower left corner of each grid cell. This estimated wind speed was then converted to a "surface wind speed", the wind speed estimated at a height of 10m in an assumed open exposure location, by using a multiplicative gradient-to-surface conversion factor. 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).

Based on the results of Liu et al. (2005), we also included hurricane indicator variables in our 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, regression models with hurricane indicator variables have limited usefulness for predicting outages during future hurricanes because the indicator variables can represent only past hurricanes. In this study, additional variables which represented hurricane characteristics beyond the wind speed have also been explored in an effort to remove the hurricane indicator variable as a significant regression parameter. These additional variables are: the central pressure of each hurricane, the Saffir-Simpson hurricane scale when each hurricane makes landfall, the maximum storm-total rainfall, the number of tornadoes during each hurricane, and the radius of maximum wind of each hurricane.

Soil Moisture Data

In order to capture the effects of soil moisture levels on the stability of trees and power poles, we estimated soil moisture levels in each grid cell at the times of the past hurricanes based on 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; Liang et al., 1996a; Liang et al., 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; Nijssen et al., 1997; Wood et al., 1997; Cherkauer and Lettenmaier, 1999), continental-scale simulations associated with the North American Land Data Assimilation System (NLDAS) (Maurer et al., 2002; Robock et al., 2003), and global-scale applications (Nijssen 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 VIC model was forced using station-based measurements of daily maximum and minimum temperature and precipitation. Daily 10m wind speeds from the 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 in the utility's service area. Modeled soil moisture data were initially reported as a depth (mm) and then were converted to a percentage of total capacity for each layer. Expressing soil moisture as a percentage of total capacity 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). The soil moisture data used in the outage model is based on the average soil moisture in the third layer (~40 to 100 cm) during the seven days prior to each hurricane landfall. Soil moisture at the other two levels was used in the statistical models but found to not be a statistically significant explanatory variable.

Summary of Input Data

In addition to information obtained from the hurricane wind field model and other characteristics of hurricanes, we included information about the power system obtained from the utility company. This includes the number of transformers, poles, switches, and customers in each grid cell and the miles of overhead line in each grid cell.

The explanatory variables in our statistical model together with their descriptive statistics are as follows:

- y_i: Number of outages in grid cell i (mean = 0.92, standard deviation = 3.60, minimum = 0, maximum = 156)
- $x_{i,t}$: Number of transformers in grid cell i (mean = 87.63, standard deviation = 145.69, minimum = 0, maximum = 1,525)
- $x_{i,p}$: Number of poles in grid cell i (mean = 234.90, standard deviation = 373.62, minimum = 1, maximum = 4,311)
- $x_{i,o}$: Miles of overhead line in grid cell i (mean = 8.58, standard deviation = 8.89, minimum = 0, maximum = 99)
- $x_{i,s}$: Number of switches in grid cell i (mean = 13.16, standard deviation = 28.42, minimum = 0, maximum = 447)
- $x_{i,m}$: Maximum 3-second gust wind speed in m/s (mean = 21.52, standard deviation = 12.28, minimum = 5.04, maximum = 53)
- $x_{i,d}$: Duration of strong winds (length of time the wind speed was above 20 m/s) in minutes (mean = 8.78, standard deviation = 8.83, minimum = 0, maximum = 42)
- $x_{i,Danny}$, $x_{i,Dennis}$, $x_{i,Georges}$, $x_{i,Ivan}$: 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 four hurricane indicator variables are zero.
- $x_{i\ Category}$: Saffir-Simpson hurricane scale when each hurricane makes landfall (1 to 5 rating)
- $x_{i\ Tornado}$: Number of tornadoes during hurricane
- $x_{i\ Pressure}$: Central pressure deficit (ΔP) in mb where $\Delta P = 1013 - Pc$ (Pc : Central pressure when making landfall)
- $x_{i\ Rainfall}$: Maximum storm-total rainfall in inches
- $x_{i\ MLS3}$: Soil moisture percentage at a depth of 40cm to 1min grid cell i (0 to 1)

In our model, each row of the data table corresponds to a single grid cell during a single hurricane. With five hurricanes and 6,681 grid cells, our data table 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 outages that occurred during Hurricane Katrina.

DESCRIPTION OF STATISTICAL MODELS

When modeling count data such as outages during hurricanes, ordinary least squares (OLS) regression models no longer apply. The assumed conditional distribution for the data is no longer correct, and the assumptions about the distributional form of the residuals no longer hold. For example, OLS regression assumes that the errors are homoscedastic (i.e., the magnitude of the error does not depend on the magnitude of the observed value), while with count data, the magnitude of the error usually increases with the magnitude of the observed value. Because of this and other violations of basic OLS assumptions, OLS regression is not appropriate for regression modeling of count data. However, a class of models called Generalized Linear Models (GLMs) has been developed for regression modeling of count data. In this section we give an overview of these models, drawing heavily on the descriptions given by Agresti (2002) and Cameron and Trivedi (1998).

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 (1), where $\vec{\beta}$ is the *n* x 1 vector of regression parameters (e.g., Cameron and Trivedi 1998).

$$E[y_i \mid \vec{x}_i] = \mu(\vec{\beta}, \vec{x}_i) \tag{1}$$

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

$$f(y_i \mid \vec{x}_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} \tag{2}$$

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 as given by:

$$\mu_i = \omega_i = \exp(\vec{x}_i^{'}\vec{\beta}) \tag{3}$$

This strong assumption of a conditional variance equal to the conditional mean is not a valid assumption for some count data sets, and, as shown below, this assumption does not hold for our hurricane power outage data set. In many cases the data is overdispersed relative to the Poisson model, meaning that the conditional variance is greater than the conditional mean. A number of models exist for dealing with this overdispersion.

The simplest method for accounting for overdispersion is to relax the assumption that the conditional variance equals the conditional mean and instead assume that:

$$\omega_i = \phi\mu_i \tag{4}$$

where ϕ is an overdispersion parameter to be estimated from the data. Cameron and Trivedi (1998) call this the NB1 model. As shown in Cameron and Trivedi (1998), the standard Poisson GLM estimates of the parameters $\vec{\beta}$ are still valid as maximum likelihood estimates, but the estimated standard errors of these estimates are inflated to properly account for the overdispersion in the data. This model leads to parameter estimates that are simple to interpret while accounting for the overdispersion in tests of parameter and model significance. However, this model is insufficient if the Poisson distribution is not a good assumption for the conditional likelihood – a situation that arises in many cases with considerable amounts of overdispersion. However, a Negative Binomial GLM, which makes different assumptions about the variance structure, provides a more flexible approach for handling overdispersion.

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 (5)

$$f(y_i \mid \vec{x}_i, \alpha) = \frac{\Gamma(y_i + \alpha)}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu} \right)^{\alpha^{-1}} \left(\frac{\mu}{\alpha^{-1} + \mu} \right)^{y_i} \qquad (5)$$

where $\mu_i = \exp(\vec{x}_i^{'}\vec{\beta})$ is a common link function, 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).

<u>Model Fitting and Measuring Goodness of Fit</u>

The Poisson GLM with NB1 variance was fit using iteratively weighted least squares implemented with the 'glmfit' command in Matlab's statistics toolbox and the 'glm' command in R. The Negative Binomial GLM was fit in SAS using the 'genmod' command and in R using the 'glm.nb' command. We used two different methods to compare fitted models for a data set. The first is the Akaike Information Criteria (AIC), and the second is a likelihood ratio test for comparing nested models. The AIC of a fitted model is defined as (Akaike 1974):

$$AIC = -2\left(Log\left(Likelihood\right)\right) + 2\left(\text{Number of Independent Parameters}\right) \qquad (6)$$

In comparing models, a smaller AIC is preferred. The AIC measure penalizes models that use larger numbers of independent parameters while also including the likelihood of the fitted model.

This is a similar idea to using an adjusted R^2 in ordinary least squares regression. Unlike the AIC measure, a likelihood ratio test is a formal hypothesis test using the difference in deviance between two nested models. This deviance has an approximately χ^2 distribution 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 one model contains a subset of the covariates included in the other model.

RESULTS

In this study, we used a Poisson GLM and a Negative Binomial GLM. We first fit a Poisson GLM with ϕ, the overdispersion parameter from equation (4), set to one using different combinations of explanatory variables. This yields the standard Poisson GLM model. Tables 1 and 2 summarize the Poisson GLM fits with the hurricane indicator variable and model fit diagnostics. The likelihood ratio test results in Table 2 coupled with the differences in model deviances suggest that models 0 and 1 are statistically indistinguishable and that model 0 outperforms (fits the data better than) model 2. In addition, the p-values for all parameters in model 2 are significant at a p-value of at least 0.01. Together, the likelihood ratio tests and the p-values lead us to the conclusion that model 1 is the preferred Poisson GLM for our power outage data. Tables 3 and 4 summarize the Poisson GLM fits without the hurricane indicator variable and model fit diagnostics. The likelihood ratio test results in Table 4 suggest that model 2 fits the data better than model 0 and model 1. A noteworthy result of the Poisson GLM in Tables 1 and 3 is that the deviance and AIC are much better with the hurricane indicator variable excluded than with the hurricane indicator variable included without adding additional hurricane parameters to the model. Moreover, the AIC of the best model in Table 3 is slightly better than the AIC of the best model in Table 1. Using the variables which are relatively easy to obtain, we can get a better fit for the outage data without the hurricane indicator variable. This can make the Poisson GLM a more useful predictive model for future hurricanes than a model that includes a hurricane indicator variable. However, we must also check to see if there is significant overdispersion that would suggest the use of a Negative Binomial model.

Table 1. Regression parameter estimates and p-values (second line of each cell) for the Poisson GLM with hurricane indicator variable.

Model	Intercept	Transformer	Pole	Mileage	Switch	Hurricane Danny	Hurricane Dennis	Hurricane Georges	Hurricane Ivan	Windspeed	Duration	MLS3	Deviance	D.F.	AIC
0	-0.6966 <.0001	0.0020 <.0001	-0.0000 0.6487	0.0417 <.0001	-0.0026 <.0001	-2.4585 <.0001	-1.0063 <.0001	-1.6182 <.0001	-1.6182 <.0001	0.0077 <.0001	0.0622 <.0001	-2.3716 <.0001	51276.84	33393	51300.84
1	-0.6948 <.0001	0.0019 <.0001	NA	0.0418 <.0001	-0.0026 <.0001	-2.4594 <.0001	-1.0059 <.0001	-1.6189 <.0001	-0.5103 <.0001	0.0077 <.0001	0.0622 <.0001	-2.3789 <.0001	51277.05	33394	51299.05
2	-2.4061 <.0001	0.0021 <.0001	-0.0000 0.006	0.0410 <.0001	-0.0024 <.0001	NA	NA	NA	NA	0.0617 <.0001	0.0173 <.0001	-1.9661 <.0001	57676.54	33398	57692.54

Table 2. Model comparisons by likelihood ratio tests for the Poisson GLM. with hurricane indicator variable.

Model comparison	Likelihood Ratio Test Statistic	Degrees of Freedom	Likelihood Ratio Test p-value	Conclusion
0 to 1	0.21	1	0.9384	Models 0 and 1 are statistically indistinguishable
0 to 2	6399.70	4	<0.0001	Model 0 outperforms Model 2

Table 3. Regression parameter estimates and p-values (second line of each cell) for the Poisson GLM without hurricane indicator variable.

Model	Intercept	Transformer	Pole	Mileage	Switch	Category	Tornado	Pressure	Rainfall	Windspeed	Duration	MLS3	Deviance	D.F.	AIC
0	-4.2307 <.0001	0.0020 <.0001	-0.0000 0.6487	0.0417 <.0001	-0.0026 <.0001	0.0570 0.6132	0.0047 <.0001	0.0346 <.0001	-0.0009 0.9103	0.0077 <.0001	0.0622 <.0001	-2.3716 <.0001	51276.84	33394	51300.84
1	-4.2782 <.0001	0.0020 <.0001	-0.0000 0.6458	0.0417 <.0001	-0.0026 <.0001	0.0693 0.0191	0.0046 <.0001	0.0346 <.0001	NA	0.0078 <.0001	0.0622 <.0001	-2.3684 <.0001	51276.86	33395	51298.86
2	-4.2783 <.0001	0.0019 <.0001	NA	0.0418 <.0001	-0.0026 <.0001	0.0704 0.0169	0.0046 <.0001	0.0346 <.0001	NA	0.0078 <.0001	0.0622 <.0001	-2.3753 <.0001	51277.07	33396	51297.07

Table 4. Model comparisons by likelihood ratio tests for the Poisson GLM without hurricane indicator variable.

Model comparison	Likelihood Ratio Test Statistic	Degrees of Freedom	Likelihood Ratio Test p-value	Conclusion
0 to 1	0.02	1	0.9944	Models 0 and 1 are statistically indistinguishable
1 to 2	0.21	1	0.9384	Models 1 and 2 are statistically indistinguishable

There are a number of ways to check for overdispersion in Poisson GLMs. One is to refit the models with ϕ in equation (4) allowed to differ from 1.0, based on the data. When this was done for models 0 and 1, ϕ was found to be significantly higher than 1 (e.g., 1.29 with a small standard deviation for model 0). Another measure is the deviance divided by the degrees of freedom, also known as Pearson's chi-squared statistic. If this is not near 1, then the data may be overdispersed or underdispersed (Faraway 2006). With the Poisson GLMs fit for the power outage data, the deviance divided by the degrees of freedom is approximately 1.54, further suggesting that the data is overdispersed relative to a Poisson GLM.

Because overdispersion exists in the power outage data set, we fit a Negative Binomial GLM. Tables 5 and 6 summarize the results of fitting Negative Binomial GLMs with hurricane indicator variables. The deviance of the full model (model 0 in Table 3) is 18,678.95 on 33,393 degrees of freedom, suggesting no lack of fit. Moreover, the deviance values for the Negative Binomial GLM are smaller than the deviance values for the Poisson GLM, indicating a more reasonable fit to the data. All possible covariates are statistically significant at a p=0.05 significance level. The likelihood ratio test results in Table 6 suggest that model 0 fits the data better than model 1 and model 2. Negative Binomial GLM model 0 fits the data set better than any the Negative Binomial GLMs that include the hurricane indicator variables and better than the Poisson GLMs. The deviance of the best-fit Negative Binomial GLM divided by the degrees of freedom of this model is 0.5594. This suggests that the Negative Binomial Model accounts for the extra variability in the count data better than the Poisson GLMs do.

Table 5. Regression parameter estimates and p-values (second line of each cell) for the Negative Binomial GLM with hurricane indicator variable.

Model	Intercept	Transformer	Pole	Mileage	Switch	Hurricane Danny	Hurricane Dennis	Hurricane Georges	Hurricane Ivan	Windspeed	Duration	MLS3	α Std.Error	Deviance	D.F.	AIC
0	-2.2928 <.0001	-0.0046 <.0001	0.0011 <.0001	0.1653 <.0001	-0.0048 <.0001	-2.4932 <.0001	-1.0791 <.0001	-1.5010 <.0001	-0.6311 <.0001	0.0129 <.0001	0.0707 <.0001	-1.1146 <.0001	1.3962 0.0314	18678.95	33393	-30318.82
1	-4.2550 <.0001	-0.0045 <.0001	0.0009 <.0001	0.1769 <.0001	-0.0041 <.0001	NA	NA	NA	NA	0.0733 <.0001	0.0168 <.0001	-0.2503 0.2745	1.8108 0.0380	18702.76	33397	-28310.48
2	-4.3041 <.0001	-0.0045 <.0001	0.0009 <.0001	0.1768 <.0001	-0.0040 <.0001	NA	NA	NA	NA	0.0731 <.0001	0.0164 <.0001	NA	1.8130 0.0380	18693.91	33398	-28311.28

Table 6. Model comparison by likelihood ratio for the Negative Binomial GLM with hurricane indicator variable.

Model comparison	Likelihood Ratio Test Statistic	Degrees of Freedom	Likelihood Ratio Test p-value	Conclusion
0 to 1	23.81	4	<0.0001	Model 0 outperforms Model 1
1 to 2	8.85	1	0.0007	Model 2 outperforms Model 1
0 to 2	14.96	5	0.0081	Model 0 outperforms Model 2

The Negative Binomial GLMs in Table 5 share a limitation with the Negative Binomial GLMs developed in Liu et al. (2005); hurricane indicator variables are included. These variables likely indicate that there are characteristics of hurricanes besides gust wind speed and strong wind duration that matter in terms of our ability to predict the number of outages in each grid cell. The presence of these variables limits the usefulness of the model as a predictive tool for outage in future hurricanes because we do not know *a priori* which of the past hurricanes a future hurricane will be most like if we do not know what the relevant hurricane characteristics are that the hurricane indicator variables are not capturing. For this reason, we added explanatory variables to the model that would help to distinguish between particular hurricanes in an effort to remove the hurricane indicator variables as statistically significant variables. Tables 7 and 8 give the results of these model fits.

Of the models shown in Table 7, model 2 has the lowest α, model 13 has the lowest deviance, and model 14 has the lowest AIC. A low α is desirable because this indicates that the variables included in the model are accounting for more of overdispersion than the variables included in the other models. A low deviance is desirable because it indicates that the log likelihood of the fitted model is being maximized. A low AIC is desirable because it indicates that a good trade-off between likelihood maximization and finding the minimum set of needed parameters is being reached. Models 13 and 14 from Table 7 are very similar, differing only in whether or not the poles variable is included. Because a likelihood ratio test suggests that model 13 and 14 are statistically indistinguishable, model 14 is preferred as the more parsimonious model. Models 2 and 14 are candidates for recommendation for a predictive model based on our data set. Because these models are not nested models (i.e., the set of covariates in one of the models is not a subset of the covariates in the other model), a likelihood ratio test cannot be used to formally compare these models. Even though model 2 has a lower α, which would lead to lower variance outage predictions, the substantially lower AIC of model 14 suggests that this is the model that best fits the data set. Model 14 is the suggested predictive model. However, we are further investigating the predictive accuracy of these models through the use of a hold-out sample. Consequently, the parameter importance analysis in the next section will be done for both of these models.

Table 7. Regression parameter estimates and p-values (second line of each cell) for the Negative Binomial GLM without hurricane indicator variable.

Model	Intercept	Transformer	Pole	Mileage	Switch	Category	Tornado	Pressure	Rainfall	Windspeed	Duration	MLS3	α Std.Error	Deviance	D.F.	AIC
0	-7.8926 <.0001	-0.0046 <.0001	0.0011 <.0001	0.1653 <.0001	-0.0048 <.0001	0.5207 0.0036	0.0026 <.0001	0.0359 <.0001	0.0414 0.0023	0.0129 <.0001	0.0707 <.0001	-1.1146 <.0001	1.3962 0.0314	18678.95	33393	-30328.82
1	-5.9057 <.0001	-0.0046 <.0001	0.0011 <.0001	0.1651 <.0001	-0.0049 <.0001	NA	0.0041 <.0001	0.0375 <.0001	0.0038 0.3634	0.0117 <.0001	0.0710 <.0001	-1.2521 <.0001	1.3945 0.0314	18696.26	33394	-30321.30
2	-5.7296 <.0001	-0.0046 <.0001	0.0011 <.0001	0.1650 <.0001	-0.0049 <.0001	NA	0.0041 <.0001	0.0367 <.0001	NA	0.0111 <.0001	0.0706 <.0001	-1.3774 <.0001	1.3941 0.0314	18699.03	33395	-30321.48
3	-10.3746 <.0001	-0.0046 <.0001	0.0011 <.0001	0.1661 <.0001	-0.0046 <.0001	1.1511 <.0001	NA	0.0342 <.0001	0.0890 <.0001	0.0158 <.0001	0.0722 <.0001	-0.9857 0.0002	1.4034 0.0315	18658.71	33394	-30312.56
4	-11.1201 <.0001	-0.0045 <.0001	0.0011 <.0001	0.1672 <.0001	-0.0047 <.0001	2.4260 <.0001	-0.0022 0.0002	NA	0.1028 <.0001	0.0240 <.0001	0.0404 <.0001	-2.5536 <.0001	1.5338 0.0336	18674.72	33394	-29636.72
5	-6.3265 <.0001	-0.0047 <.0001	0.0011 <.0001	0.1686 <.0001	-0.0050 <.0001	NA	NA	0.0392 <.0001	0.0162 0.0001	0.0196 <.0001	0.0803 <.0001	-1.4612 <.0001	1.4315 0.0319	18654.31	33395	-30172.66
6	-8.9193 <.0001	-0.0045 <.0001	0.0012 <.0001	0.1666 <.0001	-0.0049 <.0001	1.9126 <.0001	NA	NA	0.0610 <.0001	0.0218 <.0001	0.0378 <.0001	-2.7454 <.0001	1.5382 0.0337	18667.45	33395	-29623.58
7	-5.6988 <.0001	-0.0046 <.0001	0.0012 <.0001	0.1670 <.0001	-0.0052 <.0001	1.1868 <.0001	0.0015 <.0001	NA	NA	0.0195 <.0001	0.0385 <.0001	-3.3510 <.0001	1.5485 0.0338	18662.25	33395	-29578.18
8	-0.8291 <.0001	-0.0046 <.0001	0.0012 <.0001	0.1673 <.0001	-0.0055 <.0001	NA	0.0049 <.0001	NA	-0.0903 <.0001	0.0198 <.0001	0.0347 <.0001	-3.7504 <.0001	1.5797 0.0343	18672.78	33395	-29415.88
9	-5.6754 <.0001	-0.0047 <.0001	0.0011 <.0001	0.1680 <.0001	-0.0051 <.0001	0.1846 0.0007	NA	0.0322 <.0001	NA	0.0148 <.0001	0.0767 <.0001	-2.3605 <.0001	1.4277 0.0319	18677.34	33395	-30169.10
10	-5.6811 <.0001	-0.0046 <.0001	0.0012 <.0001	0.1681 <.0001	-0.0052 <.0001	1.2036 <.0001	NA	NA	NA	0.0204 <.0001	0.0426 <.0001	-3.6391 <.0001	1.5540 0.0339	18657.55	33396	-29556.86
11	-5.5717 <.0001	-0.0047 <.0001	0.0011 <.0001	0.1685 <.0001	-0.0051 <.0001	NA	NA	0.0356 <.0001	NA	0.0172 <.0001	0.0790 <.0001	-2.0265 <.0001	1.4316 0.0319	18668.92	33396	-30158.54
12	-1.0420 <.0001	-0.0048 <.0001	0.0012 <.0001	0.1721 <.0001	-0.0055 <.0001	NA	NA	NA	-0.0815 <.0001	0.0297 <.0001	0.0433 <.0001	-4.1056 <.0001	1.6305 0.0351	18643.97	33396	-29202.48
13	-1.8501 <.0001	-0.0045 0.0076	0.0011 <.0001	0.1714 <.0001	-0.0057 <.0001	NA	NA	NA	-0.0677 <.0001	0.0623 <.0001	NA	-3.1304 <.0001	1.6883 0.0360	18603.54	33397	-28973.00
14	-1.9216 <.0001	-0.0020 <.0001	NA	0.1745 <.0001	-0.0054 <.0001	NA	NA	NA	-0.0658 <.0001	0.0631 <.0001	NA	-2.9975 <.0001	1.7065 0.0362	18605.97	33398	-28880.58

Table 8. Model comparison by likelihood ratio for the Negative Binomial GLM without hurricane indicator variable.

Model comparison	Likelihood Ratio Test Statistic	Degrees of Freedom	Likelihood Ratio Test p-value	Conclusion
0 to 2	20.08	2	<0.0001	Model 0 outperforms Model 2
0 to 3	20.24	1	<0.0001	Model 3 outperforms Model 0
0 to 4	4.23	1	0.0742	Models 0 and 4 are statistically indistinguishable
3 to 5	4.40	1	0.0641	Models 3 and 5 are statistically indistinguishable
5 to 12	10.34	1	0.0001	Model 12 outperforms Model 5
12 to 13	40.43	1	<0.0001	Model 13 outperforms Model 12
0 to 13	75.41	4	<0.0001	Model 13 outperforms Model 0
0 to 14	72.98	5	<0.0001	Model 14 outperforms Model 0
13 to 14	2.43	1	0.2976	Models 13 and 14 are statistically indistinguishable

DISCUSSION OF THE MODELS

In order to use the statistical models for planning hurricane response, estimates of the number of power outages are needed for each grid cell. These outage estimates are a function of the statistically significant variables describing the hurricane characteristics and the electric power distribution system. Because the Poisson GLM is not adequate due to overdispersion, a Negative Binomial GLM is recommended. In order to estimate the mean number of outages in each grid cell (μ_i), the values of the explanatory variables for grid cell *i* are input into equations (7) – (9) below.

Equation (7) gives the equation for the mean count in a given grid cell according to the Negative Binomial GLM that includes the hurricane indicator variables. While the Negative Binomial models without the hurricane indicator variables provide better fits to the data, this model is shown here for comparison.

$$\begin{aligned}\mu_i &= \exp(-2.29 - 0.005x_{i,t} + 0.001x_{i,p} + 0.165x_{i,o} - 0.005x_{i,s} \\ &- 2.49x_{i,Danny} - 1.08x_{i,Dennis} - 1.50x_{i,Georges} - 0.631x_{i,Ivan} + 0.013x_{i,m} + 0.071x_{i,d} - 1.11x_{i,MLS3})\end{aligned} \quad (7)$$

Equation (8) gives the equation for the mean count in a given grid cell according to model 2 in Table 7, and equation (9) gives this equation according to model 14 from Table 2.

$$\begin{aligned}\mu_i &= \exp(-5.73 - 0.005x_{i,t} + 0.001x_{i,p} + 0.165x_{i,o} - 0.005x_{i,s} \\ &+ 0.004x_{i,Tornado} + 0.037x_{i,\Pr essure} + 0.011x_{i,m} + 0.071x_{i,d} - 1.38x_{i,MLS3})\end{aligned} \quad (8)$$

$$\mu_i = \exp(-1.92 - 0.002x_{i,t} + 0.175x_{i,o} - 0.005x_{i,s} - 0.066x_{i,Ra\inf all} + 0.063x_{i,m} - 3.00x_{i,MLS3}) \quad (9)$$

For example, the grid cell level estimates of power outages during Hurricane Danny could be obtained from equation (7) by setting $x_{i,Dennis} = x_{i,Georges} = x_{i,Ivan} = 0$ and $x_{i,Danny} = 1$. This can be done in the same way for hurricanes Dennis, Georges, and Ivan. The intercept of the statistical model is the expected value for Hurricane Katrina for given values of the other covariates with $x_{i,Danny} = x_{i,Dennis} = x_{i,Georges} = x_{i,Ivan} = 0$. For the Negative Binomial GLM without the hurricane indicator variable, the estimates of power outages during a certain hurricane could be obtained by plugging the values for each variable representing the aspects of the hurricane into equation (8) or (9).

In order to evaluate the relative impacts of the different explanatory variables on the mean number of counts, the relative rate of change in $\mu(x)$ with respect to a change in x_j can be written as, (Cameron and Trivedi 1998);

$$\delta_j = \left(\frac{1}{\mu(x)}\right)\frac{\partial \mu(x)}{\partial x_j} = \beta_j \tag{10}$$

For discrete indicator variables such as the hurricane indicator variable, the interpretation of the derivative must be different, but we retain the same formula as in Cameron and Trivedi (1998). Due to the difference of the units and variability of each explanatory variable, the use of the parameter δ_j to evaluate each relative effect is not ideal. It is helpful to consider the relative effect from a change proportional to each standard deviation as suggested by Liu et al. (2005) and given by

$$\delta_{\sigma j} = \beta_j \sigma_j \quad (11).$$

The parameters δ_j and $\delta_{\sigma j}$ provide an indication of the impacts of changes in each explanatory variable on the expected number of outages. This provides a measure of the relative importance of the different explanatory variables (Cameron and Trivedi 1998). Figure 1 shows the relative rate of change of the mean with respect to a change in x_j and from a change proportional to each standard deviation of the fixed effects (Transformer, Pole, Mileage, Switch, Customer, Windspeed, Duration, and MLS3) and hurricane indicators. As shown in Figure 1, the relative impacts of the hurricane indicators are much higher than the relative impacts of the other explanatory variables when basing this conclusion on the relative rate of change with respect to a change in x_j.

Unlike Figure 1, Figure 2 shows that the relative effect of the explanatory variables of the Negative Binomial GLM without the hurricane indicator variables are more concentrated on the variable MLS3. This suggests that when the hurricane indicator variables were removed from the model, the effects corresponding to the hurricane indicator variables are redistributed among the additional variables which represent hurricane characteristics and other variables such as the MLS3.

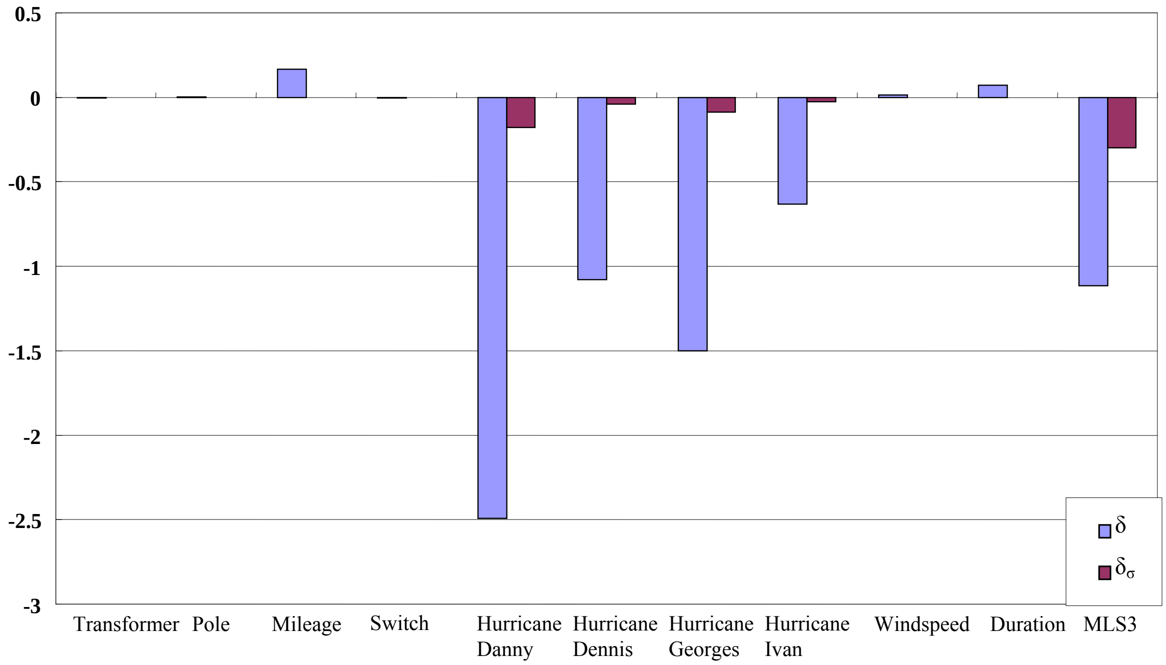

Figure 1. Relative effect of fixed effects and of hurricane indicators for the model given by equation (8).

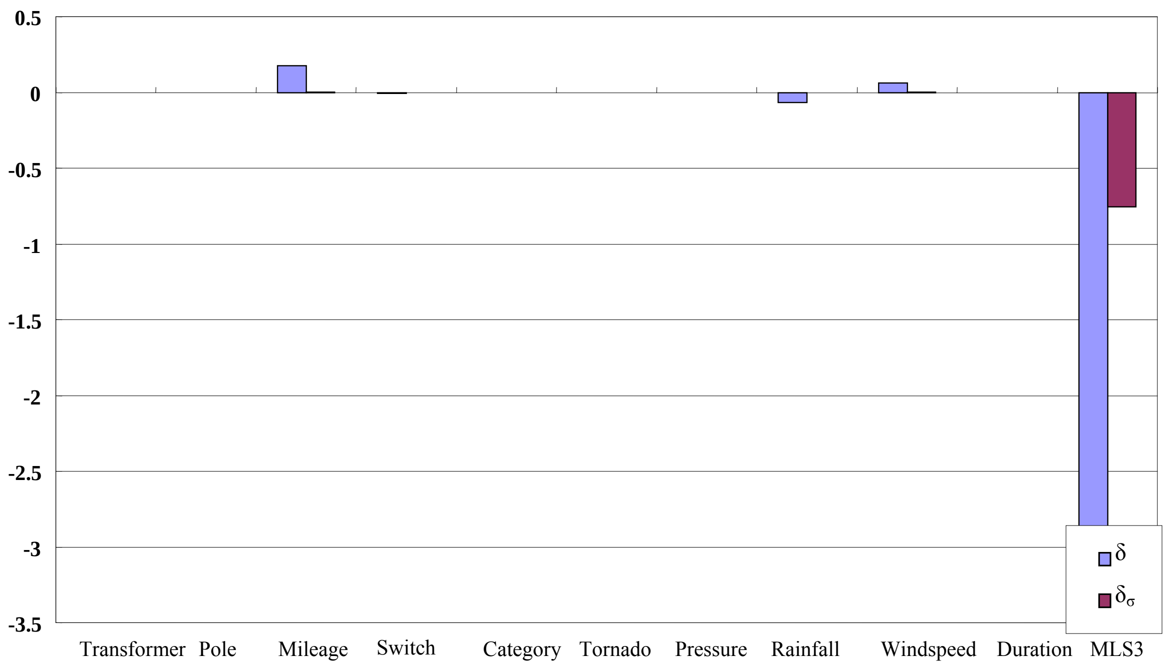

Figure 2. Relative effect of fixed effects and of hurricane indicator substitutes for the model given by equation (9).

Figure 2 and Table 7 indicate that soil moisture (MLS3) has a high relative importance in the models that do not include the hurricane indicator variable. It is also evident from Table 7 that the presence/absence of variables that describe hurricane characteristics, specifically category and pressure, seem to have a large impact on the magnitude of the soil moisture regression parameter. By just removing the pressure variable, the estimated soil moisture regression parameter changes from -1.11 (Model 0) to -2.55 (Model 4), and if the category and tornado variables are also removed the soil moisture parameter is -4.11 (Model 12). There is not significant correlation between these variables, so collinearity does not seem to explain these results. The deepest (third) layer of soil moisture in the model is strongly linked to the annual cycle, and soil moisture in this layer responds relatively slowly to recent weather conditions. It reflects the cumulative influence of weather conditions over the past several months. In ongoing research we are investigating additional soil moisture variables that would measure the deviation from long-term soil moisture trends in each grid cell in order to investigate the unexpected soil moisture results.

MODEL LIMITATIONS: COLLINEARITY PROBLEMS

A GLM regression model is built on the assumption that the explanatory variables are statistically independent of each other. If this is not the case, the model fit and parameter estimates will be highly sensitive to small changes in the input data. Regression modeling based on highly correlated input data (i.e., collinear data) can lead to poor estimation. Our data set suffers from this problem. Some of the hurricane variables are highly correlated (e.g., the correlation coefficient for Pressure-Rainfall is -0.88), as are some of the electric power system variables (e.g., the correlation coefficient for Transformers-Poles is 0.96). In the Negative Binomial models presented in Table 7 we have addressed this by removing those variables that are highly correlated to reduce the collinearity problem. This further suggests that model 14 , which minimal collinearity problems, is preferred to model 2, which may have significant residual collinearity problems. However, this approach of choosing variables to remove to reduce collinearity is ad hoc. In ongoing research we are exploring the use of Principal Components Analysis to transform the explanatory variables to a smaller set of orthogonal input variables. This would eliminate the collinearity problem.

CONCLUSIONS

For better risk management of power outages during hurricanes, models for predicting the number and location of outages were developed using Poisson Generalized Linear Models and Negative Binomial Generalized Linear Models. Because overdispersion existed in the power outage data set, the Negative Binomial GLM gave the best fit to the data set. To get the better fits, factors that can influence the vulnerability of electric power system to outages are considered. Of those factors, the windspeed and duration of the wind were obtained using the hurricane wind field model developed by Huang et al. (2001). Also, factors about the electric power distribution system provided by a large, investor-owned utility company were used. To capture hurricane characteristics, hurricane indicator variables were adopted based on the results of Liu et al. (2005). Due to the limitation of the usefulness of a model with the hurricane indicator variables, additional variables which represent hurricane characteristics beyond the wind speed have also been explored in an effort to remove the hurricane indicator variable. These additional variables are: the central pressure of each hurricane, the Saffir-Simpson hurricane scale when each hurricane makes landfall, the maximum storm-total rainfall, the number of tornadoes during hurricane, and the radius of maximum wind.

The most influential explanatory variables in the final regression model are the maximum gust wind speed, the soil moisture before the storm, maximum storm-total rainfall during the storm, and factors about the electric power distribution system (e.g. the miles of overhead line, the number of transformers, and the number of switches). In order to evaluate the relative impacts of the explanatory variables on the mean number of counts, the relative rate of change in $\mu(x)$ was calculated. With fixed effects and of hurricane indicators, the relative impacts of the hurricane indicators are much higher than the relative impacts of the other explanatory variables. Therefore, we conclude the hurricane indicators explain the influence of hurricanes in the model, but they do not provide any understanding of what aspects of the hurricanes they are measuring. However, the model without hurricane variables can provide an equivalent or better fit while using only information about hurricanes that can be easily measured. The outage count prediction model can provide a basis for choosing how many crews and materials to request from other utility companies in order to restore electric power as quickly as possible without incurring unnecessary expenses.

ACKNOWLEDGEMENTS

This work was funded by a grant from the large, investor-owned utility company in the Gulf Coast region that provided data. While they wish to remain anonymous, their support is gratefully acknowledged. We would also like to thank Jery Stedinger for helpful discussions about the modeling approach and Shawn Marino for his assistance in calculating soil moisture levels.

REFERENCES

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

Akaike, H. 1974. A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, Vol. AC-19, No. 6, pp. 716-723.

Abdulla, F. A., D. P. Lettenmaier, E. F. Wood, and J. A. Smith (1996), Application of a macroscale hydrologic model to estimate the water balance of the Arkansas-Red river basin, *Journal of. Geophysical Research*, Vol. 101, pp. 7449–7459.

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

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

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

Faraway, J.J. 2006. *Extending the Linear Model with R: Generalized Linear, Mixed Effects, and Nonparametric Regression Models*, Chapman & Hall/CRC Press, Boca Raton, FL.

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

Guikema, S.D., R.A. Davidson, and H. Liu. 2006. Statistical Models of the Effects of Tree Trimming on Power System Outages, *IEEE Transactions on Power Delivery*, Vol. 21, No. 3, pp. 1549-1557.

Hansen, M. C., R. S. DeFries, J. R. G. Townshend, and R. Sohlberg (2000), Global land cover classification at 1 km spatial resolution using a classification tree approach, *International Journal of Remote Sensing*, Vol. 21, pp. 1331–1364.

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

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges (1994), A simple hydrologically based model of land surface water and energy fluxes for GCMs, *Journal of Geophysical Research*, Vol. 99, No. 14, pp. 415–414,428.

Liang, X., D. P. Lettenmaier, and E. F. Wood (1996a), One-dimensional statistical dynamical representation of subgrid spatial variability of precipitation in the two-layer variable infiltration capacity model, *Journal of Geophysical Research*, Vol. 101, No. 21, pp. 403–421,422.

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

Liu, H., R.A. Davidson, D.V. Rosowsky, and J.R. Stedinger. 2005. Negative binomial regression of electric power outages in hurricanes, *Journal of Infrastructure Systems*, Vol. 11, No. 4, pp. 258-267.

Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, and B. Nijssen (2002), A long-term hydrologically-based data set of land surface fluxes and states from the conterminous United States, *Journal of Climatology*, Vol. 15, pp. 3237–3251.

Nijssen, B., D. P. Lettenmaier, X. Liang, S. W. Wetzel, and E. F. Wood (1997), Streamflow simulation for continental-scale river basins, *Water Resources Research*, Vol. 33, pp. 711–724.

Nijssen, B., R. Schnur, and D. P. Lettenmaier (2001), Global retrospective estimation of soil moisture using the variable infiltration capacity land surface model, 1980–93, *Journal of Climatology*, Vol. 14, pp. 1790–1808.

Robock, A., L. Luo, E. F. Wood, F. Wen, K. E. Mitchell, P. R. Houser, J. C. Schaake, D. Lohmann, B. Cosgrove, J. Sheffield, W. Duan, R. W. Higgins, R. T. Pinker, J. D. Tarpley, J. B. Basara, and K. C. Crawford (2003), Evaluation of North American Land Data Assimilation System over the southern Great Plains during the warm season, *Journal of Geophysical Research*, vol. 108, pp. 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, Clemson, S.C.

Vickery, P.J. and L.A. Twisdale, 1995. Wind-field and filling models for hurricane wind- speed prediction, *Journal of Structural Engineering*, Vol. 121, No. 11, pp. 1700–1709.

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