A Quasi-Binomial Regression Model for Hurricane-Induced Power Outages during Early Warning

Prateek Arora[1,2,3] and Luis Ceferino[1,2]

[1]Civil and Urban Engineering, New York University, Brooklyn, NY, 11201.
[2]Center for Urban Science and Progress, New York University, Brooklyn, NY, 11201
[3]Correspondence Email: pa2178@nyu.edu

ABSTRACT

Hurricanes can cause devastating damage to overhead distribution lines leading to large power outages in electric grids. Power outage prediction models can help utilities to plan for an expedited power recovery by identifying the extent of power disruptions before the arrival of a hurricane. These models often use multiple input parameters, including early warning forecasts of hurricane characteristics, and environmental and demographic information. We propose a quasi-binomial regression model to advance power outage models and overcome their existing limitations, such as unbounded outage predictions, limited extrapolation, and high uncertainties at low and high winds. This paper shows that the quasi-binomial model allows us to better capture the mechanics of power system failures due to hurricanes. We fitted our model to power outage data for four historical hurricanes, Harvey (2017), Michael (2018), Isaias (2020), and Ida (2021). We validated our model for the outages in Florida during Hurricane Ian (2022). The quasi-binomial model outperformed existing random forest and negative binomial regression models with a 7% error versus 50% and 76%, respectively. To demonstrate the quasi-binomial model's good performance more comprehensively, we also tested a new beta regression model for outages. We show the quasi-binomial model had a smaller cross-validation Root Mean Squared Error of 0.23 compared to 0.28 for the beta model. Finally, we show that our model also captures that grids with more redundant components can be more resilient to hurricane-caused outages. Thus, our proposed quasi-binomial model advances the state of the art for power outage predictions.

INTRODUCTION

Extreme weather events such as hurricanes can cause large-scale devastation to electric infrastructure (Smith 2020). Often, millions of customers lose power from hurricanes in the United States (US), e.g., (a) Hurricane Ida (2021) left 1.2 million customers without power in Louisiana (AJOT 2021), (b) Hurricane Ian (2022) left more than 2.7 million customers without power in Florida (Florida Public Service Commission 2022; Giulia Carbonaro 2022; Cortes et al. 2022), (c) Hurricane Isaias (2020) caused more than 3 million across five states in US (twitter.com 2020; Arora and Ceferino 2023a). These long-lasting power outages can cause permanent detrimental effects for vulnerable communities since power is often required to operate other critical infrastructures, e.g., hospitals, to respond to post-disaster emergency operations after disasters (Ceferino et al. 2020).

The US Government has recognized the importance of security and resilience of critical infrastructure (Executive Office of the U.S. President 2022). For example, the US Senate passed the Grid Research and Security Act (2020) (Congress.gov 2020) to secure the power grid. Additionally, the US Department of Energy (DOE) has always put forward the benefits of increasing the resilience of power grid infrastructure to extreme weather events (Office and August 2014). To achieve resilience, utilities must identify the vulnerabilities of the power systems to prepare the grid to withstand changing conditions, such as extreme weather events. However, they are often unprepared and struggle to respond to sudden hurricane-induced large blackouts, resulting in prolonged outages which had historically lasted on average for three days (Zimmerman et al. 2017).

Power outage models can be combined with early warning hurricane forecasts (Cangialosi et al. 2020) to provide utilities with outage and grid damage predictions three to seven days in advance. These predictions can be critical for resilience since they can inform the pre-impact planning and an early onset of coordination of repair crews for the rapid recovery from power outages (National Academies of Sciences, Engineering, and Medicine 2017). Thus, researchers have developed probabilistic and machine learning (ML) power outage prediction models to forecast outages before the arrival of a hurricane, e.g., from three hours to three days (Liu et al. 2005; Liu et al. 2007; Han et al. 2009a; Guikema et al. 2014; McRoberts et al. 2018). These models utilize information on forecast hurricane winds, environmental conditions, demographics, and power system components to predict hurricane-induced outages.

Previously, researchers have developed generalized linear models (GLM), generalized additive models (GAM), and random forest-based power outage models (Liu et al. 2005; Han et al. 2009a; Guikema et al. 2014). However, limited attention has been paid to understanding the compatibility of outages with Negative Binomial GAM and random forest models with the mechanics of power infrastructure failures (Arora and Ceferino 2023b).

This paper presents a quasi-binomial GAM to predict power outages and overcome the above limitations. Unlike most previous models, the quasi-binomial distribution allows to model the fraction of customers without power in its appropriate range, from 0 to 100%. Additionally, the quasi-binomial model allows flexible variance modeling, *i.e.,* to represent high certainty in predictions for extremes at low and high winds. We also capture the effect of storm surges on power systems (Cruse and Kwasinski 2021) through a proxy for coastal flood exposure: distance of city to landfall (Pachev et al. 2023). We use data from Northeast, Southeast, and Southwest states on hurricane-induced power outages to make the power outage model generalizable across multiple cities at risk from hurricanes in the US.

LITERATURE REVIEW

Liu et al. (2005) introduced a negative binomial generalized linear model (GLM) to forecast storm-induced outages in North Carolina and South Carolina. They included four input parameters: maximum gust wind speeds, number of transformers, company indicator, and hurricane indicator. Including company and hurricane indicators limited the use of the forecast model to a particular region. Liu et al. (2008) also developed spatial generalized linear mixed models (GLMM) to account for any spatial correlation in power outages but did not observe a significant improvement in the predictions.

Later, Han et al. (2009b) worked on negative binomial GLM with outage data in the gulf coast and included extensive information on underground lines, overhead lines, switches, transformers, number of customers, 3-s wind gusts, duration winds over 20 m/s, hurricane indicators, pressure, time since last hurricane landfall, precipitation, soil moisture, and land cover. Han et al. (2009a) also proposed a model without any storm or company indicator to make the model generalizable across the gulf coast region. The same group also developed negative binomial generalized additive models (GAMs) to forecast storm-induced power outages (Han et al. 2009a). GAMs showed improvement in outage predictions compared to GLMs, as GAMs can account for non-linearity between input variables and predicted outages.

Guikema et al. (2010) and Quiring et al. (2011) used decision tree models with topology and soil parameters information to predict outages. Researchers have recently used other non-parametric random forest models to handle non-linearity in input parameters (Guikema et al. 2014; McRoberts et al. 2018; Shashaani et al. 2018). The random forest model grows multiple parallel trees to reduce the variance and make predictions robust to outliers and noise (Breiman 2001). Guikema et al. (2014) used 3-s wind gusts, duration of winds over 20 m/s, and population density to develop utility and hurricane-independent predictions of the fraction of customers without power. Nateghi et al. (2014) used the additional information on tree-trimming practices to improve the power outage predictions. However, the model had limited applicability since information on tree-trimming practices is generally not publicly available. Madrigano et al. (2015) included the information on tree species to further improve the power outage predictions with a random forest model.

Tonn et al. (2016) predicted power outages at different confidence intervals using quantile random forests. Wanik et al. (2017) used lidar-derived tree data to develop another random forest power outage model. McRoberts et al. (2018) developed a two-stage model to predict power outages. The first stage would make a binary prediction to identify outages or no outages in a region. The second stage would predict the fraction of customers without power for the regions with at least one outage. Shashaani et al. (2018) developed a three-stage power outage model with the first stage to identify outage or no-outage; the second stage to classify outage regions into low, moderate, and high class; the third stage to predict the total number of outages.

Researchers have also used artificial neural networks to predict outages (Xie et al. 2020). Haseltine and Eman (2017) used a neural network to predict the failure of individual power grid components before the arrival of a storm. However, high-resolution data on each grid component is generally not publicly available. Sun et al. (2016) used Twitter data to predict real-time outage, and Jaech et al. (2018) used repair logs to forecast repair times. However, information from Twitter and repair logs are not available before the arrival of a storm and could not be employed for pre-storm planning and coordination of resources.

Arora and Ceferino (2023b) studied the limitations of the state-of-the-art power outage models. The first limitation was that negative binomial GAMs could overpredict power outages. For example, they showed negative binomial could predict as high as 16 times more outages than the actual number of customers. Second, random forest regression has limited capability to extrapolate for outages at high winds due to sparsity of power failure data, i.e., predictions saturated at 70% of customers without power at wind speeds around 70 m/s. Third, both the Negative binomial GAM and the random forest did not capture the mechanistic behavior of infrastructure failures in their uncertainty estimates. Negative binomial GAM showed high variance when predicted outages were close to 100% at high winds, and random forest showed high variances when predicted outages were close to 0% at low winds. However, we expect close to 0% variance when outage predictions are close to 0% and 100%. For example, at very low winds, we expect little impacts to the power grid with high certainty, and at catastrophic wind levels, we expect large impacts also with high certainty. In addition, power system components near coastlines are heavily exposed to significant flooding, *e.g.,* due to storm surges (Cruse and Kwasinski 2021). Still, little attention has been given to incorporating this damage mechanism in power outage modeling.

This paper presents the quasi-binomial outage model to overcome the aforementioned limitations. We also included the proxy variable, the distance of a city to landfall, to model the damaging effects of storm surges on the power grid.

DATA DESCRIPTION

We included power outage data from five large historical hurricanes in developing the quasi-binomial regression power outage model. We acquired the power outage for Hurricane Harvey (2017) in Texas, Isaias (2020) in New Jersey and New York, and Ida (2021) in Louisiana from PowerOuage, an independent organization that keeps track of power outages based on information from utilities in the US (PowerOutage 2022). In addition, we obtained the power outage data for Hurricane Michael (2018) and Ian (2022) from Florida Public Service Commission (2022). We used Hurricane Harvey, Michael, Isaias, and Ida's data for model calibration and Ian's data for validation. For training, we filtered outliers as PowerOutage's reports are approximated in a few cases (PowerOutage 2022). We used the interquartile range criteria on the ratio of households and the number of customers without power to remove 255 out of the 2577 total data points (cities) from Hurricane Harvey, Michael, Isaias, and Ida for calibration.

Power outage predictions rely on input parameters, including hurricane, environmental, power system, and socio-demographic features. The input parameters are typically available at different scales. Thus, we adjusted the input data (e.g., aggregated or interpolated) to match the same spatial resolution of the outage data, as in our previous research (Arora and Ceferino 2023b). We developed the power outage model at the city-level spatial resolution consistent with the recent power outage models. Future research could develop an outage model with a finer resolution if input parameters and historical power outages at a finer spatial resolution are available.

In total, this study used data on 4.62 million outages in 2322 cities across four storms covering 11 utilities in New York and 5 in New Jersey, five in Florida, 39 in Texas, and nine in Louisiana. Figure 1 shows the number of outages across New Jersey after Hurricane Isaias (2020) to develop the quasi-binomial model. Additionally, Supplementary Figures S1, S2, S3, and S4 show the power outages in Texas during Hurricane Harvey (2017), Florida during Hurricane Michael (2018), New York during Hurricane Isaias (2020), and Louisiana during Hurricane Ida (2021). We further describe the response and input variables, also summarized in Table 1.

Response Variable

We developed the power outage model to predict the fraction of customers without power at the city level. We chose the fraction of customers without power as the response variable to quantify the impact of power outages across cities having variable numbers of customers.

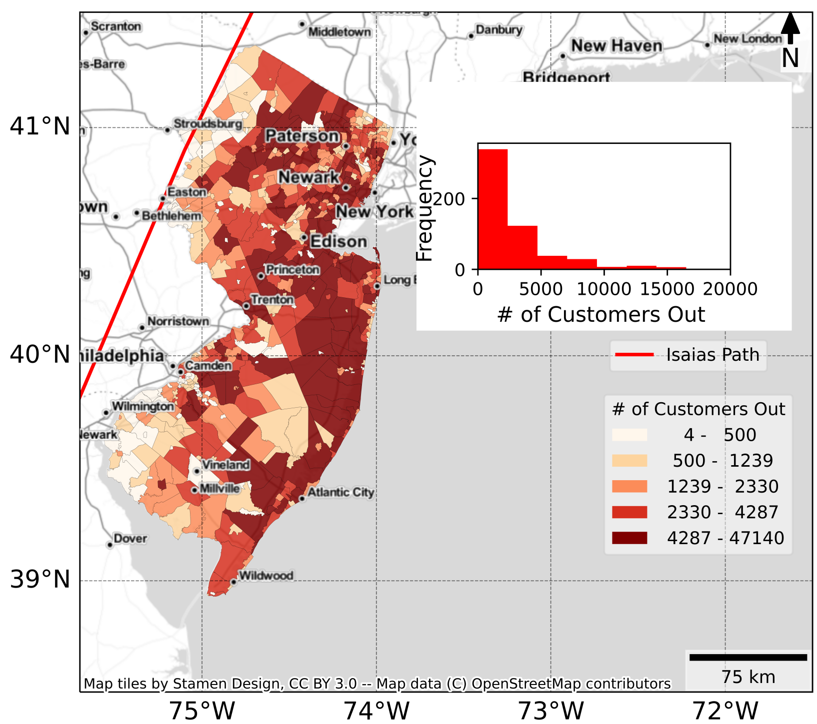

Fig. 1. Distribution of power outages at quantile breakpoints in the aftermaths of Hurricane Isaias (2020) in New Jersey, as reported by PowerOutage

Input Parameters

Researchers have used different explanatory variables to predict hurricane-induced power outages (Liu et al. 2005; Han et al. 2009b; Han et al. 2009a; Guikema et al. 2010; Guikema et al. 2014; Shashaani et al. 2018). The explanatory variables include hurricane winds, whose intensity and extent can determine the scale of damage to power systems (Bjarnadottir et al. 2013). Other environmental conditions (*e.g.,* land use, precipitation, soil moisture) can inform about more vulnerable sections of the power grid. For instance, power poles or trees in wet soil conditions are more susceptible to being torn out from soil under high wind conditions (McRoberts et al. 2018). Key power infrastructure data (*e.g.,* the number of transformers) can also inform on the redundancy of power systems. In grids with redundant components, people are more likely to access electricity even if a component fails, e.g., from strong hurricane winds (Brown 2002). Finally, socio-demographic information (*e.g.,* population density) can inform on the customer density and potential layout of the power grid in urban and rural areas. Thus, we divided the input parameters into four categories: hurricane features, environmental features, power system information, and socio-demographic features, which are described next.

Hurricane Features

We included 3-s gust wind speed and duration of strong winds, i.e., duration of winds over 20 m/s consistent with previous research (Liu et al. 2005; Han et al. 2009b; Han et al. 2009a; Guikema et al. 2010; Guikema et al. 2014; Shashaani et al. 2018). We used Best Track data for the hurricane path (Knapp et al. 2010) and a complete tropical cyclone wind field model by Chavas et al. (2015) to determine the axisymmetric winds and background wind model byLin et al. (2012) to obtain the complete wind structure. Often, background winds are neglected, which can lead to underestimated maximum winds. For example, we calculated that the maximum 3-s gust wind increased from 37.5 m/s to 49.5 m/s (32% more) in New Jersey after including background winds during Hurricane Isaias (2020). We consider the 3-s gust wind speed and duration of strong winds at the centroid of a city. We used the log of 3-s gust winds as an input parameter to better model power outages. We show the 3-s gust winds in New Jersey during Hurricane Isaias in Figure 2.

Historically, apart from high hurricane winds, significant storm surges can also impact the power systems. Poudyal et al. (2022) showed that the compound effect of hurricane winds and storm surge could lead to more power loss compared to damages by hurricane winds alone in Texas. Not only a hurricane can cause a significant storm surge when it transitions from sea to land. A hurricane can also cause flooding when transitioning from land to sea due to high waves (Bucci et al. 2022). Thus, we included the minimum distance between the city and the points where the hurricane crosses the coastal line as an input parameter to capture the compound risk to power systems from hurricane winds and storm surges. We normalized the distance with the radius of maximum winds to account for different hurricane sizes.

Other Environmental Features

We included various environmental features for outage modeling, e.g., land cover information, topography, soil conditions, and tree density. These environmental factors can help enhance outage predictions as described in the previous research (Liu et al. 2005; Han et al. 2009b; Han et al. 2009a; Guikema et al. 2010; Guikema et al. 2014; Shashaani et al. 2018; Arora and Ceferino 2023b).

We acquired land cover data available as National Land Cover Data (NLCD) from the US Geological Survey (USGS) on Multi-Resolution Land Characteristics Consortium (Jin et al. 2021). The land cover classes can inform about distribution grid layouts, resulting in different outage patterns. For example, rural areas mostly have a radial

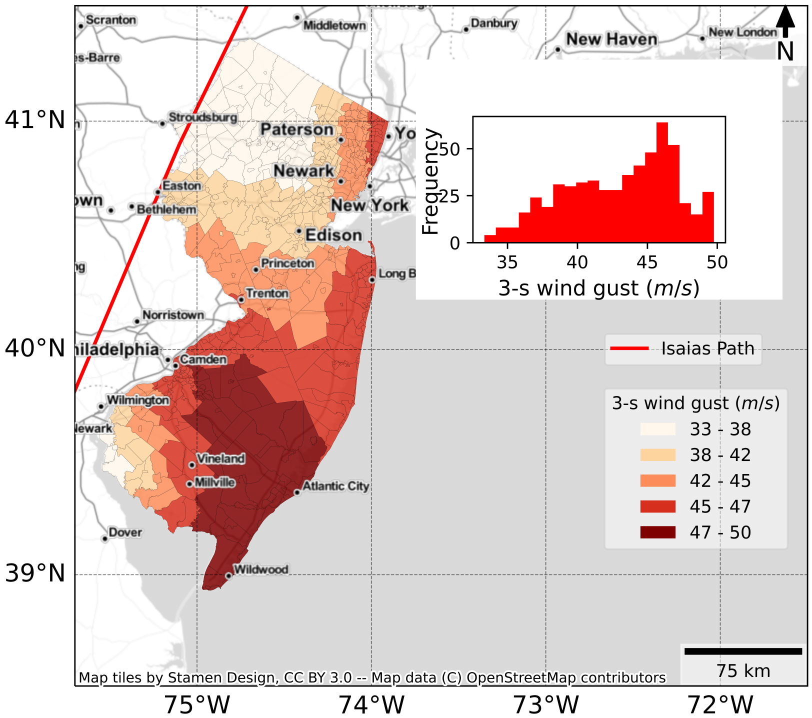

Fig. 2. Distribution of 3-s gust wind speeds during Hurricane Isaias (2021) in New Jersey. The wind gusts include both circulating and background components of the winds.

layout for the power grid where each customer is connected to a single feeder. Thus, the failure of a single feeder in a radial grid leads to the shutting down of all downstream lines until the repair is completed for downed power lines, which can result in more outage propagation due to failed power system components than in urban areas (Petersen 1982). USGS originally classified NLCD data into 20 classes, available in raster format with cells of size 30 $m \times 30\ m$. We reclassified the original 20 NLCD classes into nine major classes: developed, water, barren land, forest, scrub, grasslands, pasture land, cultivated cropland, and wetlands. We used following *python* packages: *rasterio*, *gdal*, *shapely*, *fiona*, *geopandas*, *zonal_statistics* to perform calculations on raster data. We calculated the percent of area occupied by each class of land cover in a city.

Han et al. (2009b), Nateghi et al. (2014), McRoberts et al. (2018) have highlighted the importance of including precipitation and soil moisture conditions in power outage predictions. Many distribution lines are close to the trees, posing an imminent threat to failure of distribution lines from trees falling due to strong hurricane winds. The deviations in precipitations and soil moisture from standard conditions might increase the likelihood of trees uprooting from high winds resulting in more outages. Trees are more likely to uproot due to strong hurricane winds under wet soil conditions (Han et al. 2009b; Nateghi et al. 2014). Additionally, persistent dry conditions can create gaps in soil, weakening tree roots and increasing the chance of tree uprooting from high winds.

We acquired the information on soil moisture and precipitation at an hourly scale for the past 42 years, starting in January of 1971 from Xia et al. (2012). Xia (2012) developed the Variable Infiltration Capacity (VIC) model available on National Land Data Assimilation System Phase 2 (NLDAS2). NLDAS2 data is available in a grid pattern with points spaced apart $0.125° \times 0.125°$. We obtained the precipitation and soil moisture at the centroid of a city using nearest neighbor interpolation.

We extracted the soil moisture available for three depths: $0 - 10$ cm, $10 - 40$ cm, and $40 - 100$ cm from NLDAS2 data. We converted the hourly soil moisture to daily soil moisture by taking an average of hourly soil moisture over a day. Soil moisture may vary significantly from one location to another because of ambient conditions. We normalized the soil moisture by computing the percentiles at each location after fitting the daily soil moisture data to Pearson Type III distribution using maximum likelihood estimate (MLE). Then, we evaluated the percentile of soil moisture a day before the storm would hit the city, denoted as CDF1: percentile of soil moisture for $0 - 10$ cm depth, CDF2: percentile of soil moisture for $10 - 40$ cm depth, and CDF3: percentile of soil moisture for $40 - 100$ cm depth.

Precipitation conditions are represented as standard precipitation index (SPI) (Wu et al. 2007; Guttman 1998; Casey 2016). SPI can represent the wet and dry conditions before a storm hits. We calculated the SPI for one month (SPI1), three months (SPI3), six months (SPI6), and 12 months (SPI12). SPI is obtained from precipitation time series data in three steps. First, we fitted the time series data to Pearson Type III distribution. Second, we calculated the percentile of precipitation for the month before the storm's impact. In the final third step, we obtained SPI using the inverse of cumulative standard normal distribution on the percentiles computed in the previous step. We also considered the expected precipitation after the hurricane as an input parameter, as heavy precipitation can lead to weakened roots and more trees uprooting (McRoberts et al. 2018).

We included the percentage of trees in a city to capture the falling tree hazard on electric lines and poles. We obtained the area covered by trees from the National Insect and Disaster Risk Maps (Krist Jr. et al. 2014) created by the US Department of Agriculture (USDA) available at a resolution of 240 $m \times 240$ m. We use zonal statics in ArcGIS to determine the total area covered by trees in a city.

The depth of the soil from which plants and trees can effectively derive water and nutrients for their growth is defined as the root zone (RZ) depth. Trees with more RZ depth can withstand strong hurricane winds (McRoberts et al. 2018). We obtained the RZ data from USDA under Gridded Soil Survey Geographic (Soil Survey Staff 2021) as raster data with a resolution of 30 $m \times 30$ m. We used zonal statistics in ArcGIS to obtain the mean RZ depth for a city. We included the RZ depth as an input parameter to power outage predictive models as RZ depth could additionally indicate falling hazards of trees on distribution lines and poles from strong hurricane winds.

Chapman (2000), Miller et al. (2013), Guikema et al. (2010), Quiring et al. (2011), McRoberts et al. (2018) have found variation in hurricane wind speeds with varying surface topography. Therefore, a city's mean and median elevation are typical topographic input parameters in outage predictions. We obtained mean and median elevation from a Digital Elevation Model (DEM) at a 30 m resolution in the Global Multi-Resolution Terrain Elevation Data (GMTED2010) by USGS (Danielson and Gesh 2011).

Feature	Abbreviation	Data Source
Outages	Outages	(PowerOutage 2022)
log of 3-s Gust Wind Speed	logvmax[a,b]	(Chavas et al. 2015; Knapp et al. 2010)
Duration of Strong Winds	Duration	
Distance to Hurricane crossing coast Normalized with Radius to Maximum winds	DistHurr[a,b]	
Percent Developed Area	Developed	(Jin et al. 2021)
Percent Water Area	Water	
Percent Barren Area	Barren[a]	
Percent Forest Area	Forest	
Percent Scrub Area	Scrub	
Percent Grassland Area	Grassland[a]	
Percent Pasture Area	Pasture	
Percent Crops Cultivated Area	Crops	
Percent Wetlands Area	Wetlands[a,b]	
Standard Precipitation Index 1 month	SPI1	(Xia et al. 2012)
Standard Precipitation Index 3 months	SPI3	
Standard Precipitation Index 6 months	SPI6[a,b]	
Standard Precipitation Index 12 months	SPI12	
Soil Moisture 1^{st} Layer	CDF1[a,b]	(Xia et al. 2012)
Soil Moisture 2^{nd} Layer	CDF2	
Soil Moisture 3^{rd} Layer	CDF3[a]	
7 day precipitation	precip[a,b]	(Xia et al. 2012)
Root Zone Depth	Rzone[a]	(Soil Survey Staff 2021)
Percent Treed Area	Trees[a,b]	(Krist Jr. et al. 2014)
Mean Elevation	Mean_Ele[a]	(Danielson and Gesh 2011)
Median Elevation	Median_Ele	
Substations	Substations	(data.gov 2022)
Substation Normalized to Population	Sub_Norm[a,b]	
Road Length Normalized to Substations	R_over_S	
Population Density	Pop_Den[a,b]	(American Community Survey 2019)

Table 1. Parameters to build the power outage prediction models: all variables are rescaled at the city level. Parameters are grouped into categories separated by horizontal lines. We selected one variable from each category from each group to minimize correlation across parameters.
a-variables selected after filtering of correlated features
b-finally selected variables after 5-fold cross-validation

Power System Information

Previously, Liu et al. (2005), Liu et al. (2007), Han et al. (2009a), Han et al. (2009b), Guikema et al. (2010), Quiring et al. (2011) have included information on power systems, such as density of transformers, type of protection device, number of poles, number of transformers, length of overhead lines, and length of the underground line, in outage predictions from hurricanes. The low number of protective devices and transformers, can indicate sparsity and low redundancy in power systems. The sparsity in power system protection devices, e.g., sectionalizing switches to separate the damaged part of power lines, profoundly affects power delivery to customers. The failure of one protection device in a sparse power network can lead to larger and longer power interruptions for many customers (Brown 2002). Also, National Academies of Sciences, Engineering, and Medicine (2017) reports that overhead distribution lines are the most vulnerable components of a power grid to strong hurricane winds. Power systems with many wood poles are particularly at high risk as these poles are generally designed to moderate wind speeds only up to 20 m/s (IEEE 2007).

Thus, including this information about power systems can help better capture the propagation of power outages. However, due to security concerns, the information on the power system's variables is not generally publicly accessible (States and Accountability 2020). To overcome this problem, Birchfield et al. (2017) , Pahwa et al. (2014), Schultz et al. (2014), Schweitzer et al. (2010), Valenzuela et al. (2019), Pisano et al. (2019), Zhai et al. (2021) proposed generating synthetic power grids using open public georeferenced data *e.g.,* roads, buildings. Yet, formulating and implementing these procedures can be a separate study.

To keep the analysis simple and still include key power system information in our model, we adopted a few simplified techniques of synthetic grid generation. To track the number of poles exposed to hurricane winds, we approximated the length of distribution lines with the length of local roads in a city since most lines are along roads (Winkler et al. 2010; Valenzuela et al. 2019). In addition, we gathered information on the location and number of transmission substations from the Homeland Infrastructure Foundation Level Database (data.gov 2022) to use it as a proxy for the power systems' sparsity or redundancy.

To illustrate why a small number and density of substations can be an important indicator of lack of redundancy, Figure 3 shows the power flow in a radial power grid. The power flow in a grid starts at generation substations. Then transmission lines deliver power to transmission substations. Sub-transmission power lines deliver power to commercial customers or distribution substations. Primary distribution lines deliver power to distribution transformers, and the secondary distribution system delivers power to feeders. Finally, each customer is connected to a single feeder for power delivery. The failure of a feeder will result in a power outage for all downstream customers. According to the flow of power supply in a radial grid as depicted in Figure 3, more transmission substations would indicate more distribution transformers (Brown 2002).

Thus, we included the number of substations, the number of substations normalized with population, and the number of transformers per unit road length (an indicator of power lines) in the power outage prediction model. Since some cities have no transmission substation, we calculated the variables representing power systems at a coarser resolution of the county level and assigned their values to all the cities within the county.

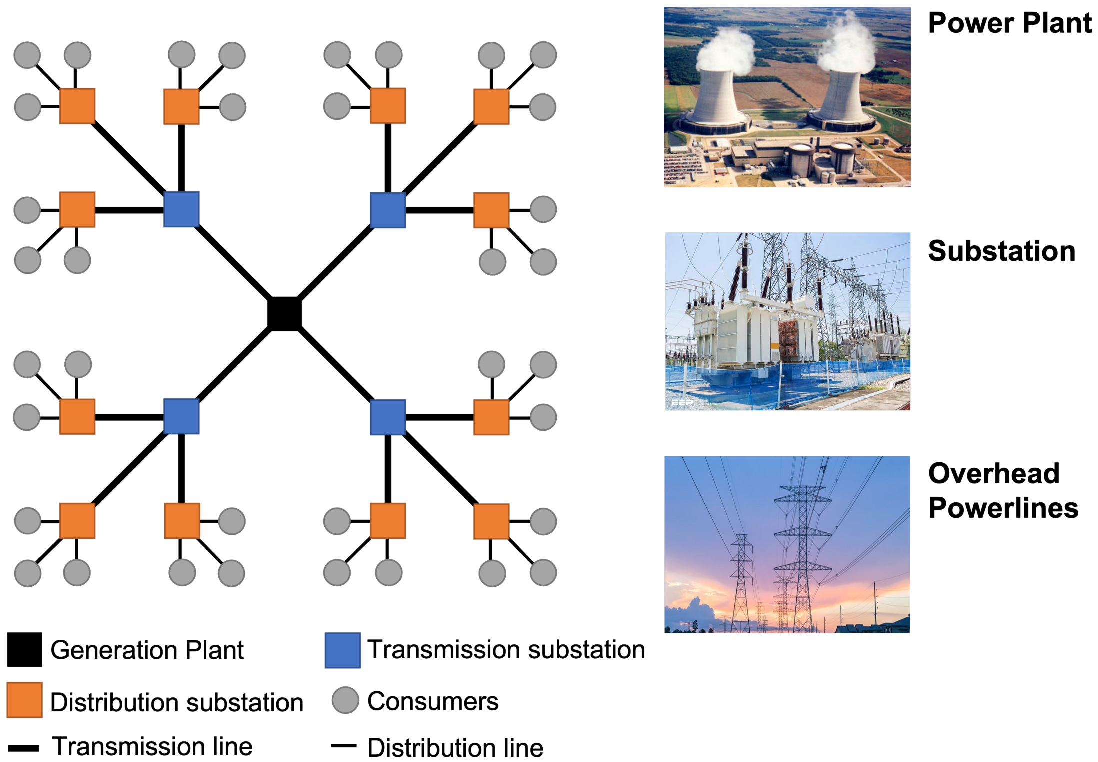

Fig. 3. Depiction of the flow of electricity in a radial power grid. First, power is delivered from a generation source (*e.g.,* coal plant) to a transmission substation. Then, high-voltage transmission power lines transmit power to the distribution substation. Finally, power is delivered to customers through distribution power lines. Each customer receives power from a single feeder. In a sparse radial grid with fewer feeders, the failure of a single feeder would result in outages for multiple downstream customers. Pictures for power plant, substation, and overhead powerlines are extracted from (USGS 2004), (Emergible Consulting Services, Accessed Online: 08/10/2023), and (JustEnergy, Accessed Online: 08/10/2023), respectively.

Density Information

Population density could also indicate the extent of lines and poles exposed to hurricane winds (Arora and Ceferino 2023b). Rural areas with a low population density generally have sparsely located power grid components where the failure of one component could result in a power outage for multiple customers compared to urban areas with redundant power components (Brown 2002). To capture this effect, population density has been used for power outage modeling (McRoberts et al. 2018). Thus, we also used population density from 5-year estimates in American Community Survey (2019) as an input parameter.

POWER OUTAGE MODELING

We develop the power outage model to predict the fraction of customers without power in a city. Beta, and binomial regressions (Olkin and Liu 2003) have been used in many fields, ranging from financial to ecological studies (Yee 2012; Olkin and Liu 2003; Douma and Weedon 2019), when the response variable is a fraction distributed from zero to one, such as the proportion of customers without power in the aftermath of a hurricane. Unlike the linear regression model, the beta, binomial, and quasi-binomial regression models do not assume homoscedasticity *i.e.,* constant variance for output variable with change in input parameters, and can model the variable dispersion in power outage predictions across different cities. Previously, researchers have applied methods other than linear regression to model power outages. Still, those models fail to capture the mechanics of power infrastructure failures as stated in Arora and Ceferino (2023b).

Thus, we explore beta regression and quasi-binomial regressions as models to enhance power outage predictions, better capturing the physics of power system failures. We present the quasi-binomial instead of the binomial regression because it better models large overdispersion, a feature extensively reported to be critical previously in outage prediction (Dunn and Smyth 2018; Han et al. 2009a). Also, notice that the quasi-binomial regression is a hierarchical model of the binomial regression; thus, it inherits its modeling strengths.

Beta Regression

Since the fraction of customers without power y in a city is a continuous variable with values ranging from zero to one, we can model y as beta distributed. The probability density function is given by

$$f(y|\alpha,\beta) = \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha,\beta)} \tag{1}$$

where

$$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \tag{2}$$

and α and β are the parameters to be estimated, and $\Gamma(.)$ is the gamma function. Correspondingly, the mean and variance for the fraction of customers without power are

$$E[y] = \frac{\alpha}{\alpha+\beta} \tag{3}$$

$$var[y] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{4}$$

Generally, the beta distribution is represented with an alternative form with parameters μ and ϕ so that the mean $E[y]$ and variance $var[y]$ of beta distribution are given as:

$$E[y] = \mu \tag{5}$$

$$var[y] = \frac{\mu(1-\mu)}{1+\phi} \tag{6}$$

This representation is known as mean-precision parameterization, where the previously defined μ is the mean, and ϕ is the precision parameter (inverse of dispersion). Thus, we can rewrite the probability distribution in Eq. 1 with μ and ϕ as

$$f(y;\mu,\phi) = \frac{\Gamma\{\phi\}}{\Gamma\{\mu\phi\}\Gamma\{(1-\mu)\phi\}} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} \tag{7}$$

where $0 < \mu < 1$, and $\phi > 0$. The wide range of μ and ϕ provides excellent flexibility to fit different probability distribution shapes, *e.g.,* skewed, symmetric, bell-shaped. Beta regression parameters are determined using Maximum Likelihood Estimates (MLE) similar to Generalized Linear Models (GLMs) (Dunn and Smyth 2018). The mean parameter μ is related to the input parameters via a link function

$$g(\mu) = \beta X \tag{8}$$

where β is the vector of the unknown parameters determined using MLE, and X is the matrix with input parameters. The link function $g(\mu)$ must be monotonic and twice differentiable, and μ is bounded between 0 and 1. (Ferrari and Cribari-Neto 2004). The most common choice for the link function is a logit function.

$$g(\mu) = log\left(\frac{\mu}{1-\mu}\right); \mu = \frac{e^{\beta X}}{1+e^{\beta X}} \tag{9}$$

The dispersion parameter ϕ can also be linked to X similar to μ in Eq.8. However, for the scope of this paper, we keep ϕ constant. Notice that under this assumption, the beta distribution still provides enough flexibility to fit different probability distribution shapes (*e.g.,* skewed, symmetric, bell-shaped). Finally, the log-likelihood function is maximized to estimate unknown parameters β and ϕ for n observations. We maximized the log-likelihood for the outage data obtained for Hurricane Harvey (2017), Michael (2018), Isaias (2020), and Ida (2021).

So far, the beta distribution proposed in Eq. 7 and Eq. 8 is defined on an open interval, i.e., $0 < y < 1$. However, the fraction of customers without power in a city could also be equal to exactly zero or one. Fitting exact values of zeros and ones leads to poor estimates of β and ϕ. Previously researchers have introduced ways to model such boundary conditions for beta regressions (Smithson and Verkuilen 2006; Ospina and Ferrari 2012; Liu and Kong 2015; Swearingen et al. 2012). For example, Ospina and Ferrari (2012) nudged the output variables so that y would never reach values of zero and one. Furthermore, Ospina and Ferrari (2012), Liu and Kong (2015), Swearingen et al. (2012) developed zero, and one-inflated beta regressions to model the boundary conditions separately, but such models require additional parameters to be determined. Instead, we include a slack variable ϵ to bound the fraction of customers without power between open intervals (Smithson and Verkuilen 2006).

$$\tilde{y} = \begin{cases} y+\epsilon & if\ y=0 \\ y & if\ 0<y<1 \\ y-\epsilon & if\ y=1 \end{cases} \tag{10}$$

We chose $\epsilon = 10^{-8}$ to ensure no significant changes in the accuracy of predicted outages. For example, New York City (NYC) is the most populous city in the US, with 7.8 million people. If NYC has no outages, adding a value of 10^{-8} to the fraction of customers without power would still result in less than one outage, which is rounded off to zero outages.

The GLM-like formulation in Eq. 8 models a linear relationship between the logit of the mean of the fraction of customers without power and input parameters. However, this relationship can be non-linear, as demonstrated in previous research (Han et al. 2009a), which can be modeled with Generalized Additive Models (GAMs), an extension of GLMs. GAMs formulation using smoothing functions can capture this non-linear relationship.

$$g(\mu) = \beta_0 + \sum_{j=i}^{p} \beta_j f_j(x_j) \tag{11}$$

where $g(\mu)$ is the link function in Eq. 9, β_0 is the intercept, and f_j are smoothing splines functions which are fitted to each parameter j, where $j \in [1, \ldots, p]$, and p is the total number of input parameters. The most common examples of smoothing splines include P-splines, B-splines, and S-splines. Splines for each parameter could be fitted with polynomials of any degree. Thus, we modeled all parameters as polynomials of the S-spline, except wind gust (logvmax) and distance of the city to landfall (DistHurr) (Table 1). We modeled wind gusts with a polynomial of degree one because higher winds are expected to increase damage monotonically, e.g., zero damage at zero wind gusts and complete collapse of power systems with wind gusts from hurricanes of category 5 ($> 90\ m/s$). Also, cities closer to the point of the hurricane-crossing land tend to experience higher storm surges. Hence, we also modeled the parameter DistHurr as a degree one polynomial. We used *mgcv* library in the R program to perform MLE estimates of unknown parameters (Wood 2017).

Quasi-Binomial Regression

The binomial regression is another popular class of regressions to model fractions arising from count data, such as the fraction y of outages in a city with m total customers. Quasi-binomial regression inherits these binomial regression's modeling strengths and, in addition, can handle overdispersion in the outage predictions better (Dunn and Smyth 2018; Consul 2010). Quasi-binomial distribution is a simplified form of beta-binomial distribution. The beta-binomial distribution assumes that the mean rate of failure (μ in Eq. 12) is beta distributed (Dunn and Smyth 2018). Thus, the probability mass function of beta-binomial distribution is given as

$$P(y; m, \alpha, \beta) = \binom{m}{my} \frac{y^{\alpha-1}(1-y)^{\beta-1}}{B(\alpha, \beta)} \tag{12}$$

where $B()$ is the function from Eq. 2, and m is the number of customers in a city. The parameters α and β are determined using MLE. Unlike beta regression, MLE for a quasi-binomial regression is weighted based on the number of customers in a city. Correspondingly, the mean and variance for the fraction of customers without power are

$$E[y] = \frac{\alpha}{\alpha + \beta} \tag{13}$$

$$var[y] = \frac{m\alpha\beta(\alpha + \beta + m)}{(\alpha + \beta)^2(\alpha + \beta + 1)} \tag{14}$$

Similar to beta distribution, we can parametrize the beta-binomial in terms of μ and ϕ, resulting in the typical form of the quasi-binomial distribution (Dunn and Smyth 2018). In that case, $E[y] = \mu$ and $var[y] = \phi\mu(1 - \mu)$. Here, ϕ is the dispersion parameter that allows handling varying overdispersion in the proportions of outages.

Quasi-binomial regression can model the fractions in a closed interval, i.e., including zeros and ones. Thus, we do not need to apply the transformation in Eq. 10. Similar to the beta model, we used GAMs representation from Eq. 11 and logit link function from Eq. 9 for the quasi-binomial model. We use *mgcv* library in the R program to fit the quasi-binomial model for the fraction of customers without power.

FEATURE SELECTION

Initially, we considered a large number of input parameters, which can challenge developing meaningful regression models (Kira and Rendell 1992). High-dimensional input data can result in degraded performance with overfitting, *i.e.,* small errors for training data but high error on test data. Thus, feature selection is crucial in model development, especially for sparse datasets such as the case for natural hazards (Kumar and Minz 2014). Thus, we selected the most appropriate and sufficient number of features to predict outages accurately (Provost 1999), as in previous research for outage modeling (Han et al. 2009b).

Feature selection also helps address multi-collinearity, i.e., when highly correlated input features introduce spurious causation because of similar importance coefficients in the regression. One method used in previous research to avoid multi-collinearity is Principal Component Analysis (Han et al. 2009b; Han et al. 2009a; Markhvida et al. 2018). However, we did not use it to make the model generalizable by keeping the features in the original space, which also enabled us to study the effects of key hurricane characteristics, such as winds, on power outages.

Instead, we filtered out the highly-correlated input features. We provide the Pearson correlation ρ for all the variables in Supplementary Figure S5. First, we selected only one feature with highest ρ out of sets of highly-correlated input features with Pearson correlation coefficient (ρ) > 0.5, from Table 1. Then, we filtered out the input features with a correlation to the output (fraction of customers without power) below 0.1 to keep only the most important variables. For example, we filtered out the variable "Substations" as it had $|\rho| < 0.1$ with fraction of outages, but we still kept Substations normalized to populations since $|\rho| = 0.23$. So, our model will still include the effects of power system variables on outages. After filtering of variables, we removed 14 variables (out of 28) from further model development. The remaining variables for subsequent analysis are 1) logvmax, 2) DistHurr, 3) Barren, 4) Grassland, 5) Wetlands, 6) SPI6, 7) CDF1, 8) CDF3, 9) precip, 10) Rzone, 11) Treed_area, 12) Mean_Ele, 13) Sub_Norm, 14) Pop_Den.

MODEL SELECTION

We first conducted a qualitative assessment of the performance of beta and quasi-Bionmial GAM models. We computed their standardized quantile-residuals to assess model quality by evaluating how close they are to a standard normal distribution (Dunn and Smyth 1996; Pereira 2019; Klar and Meintanis 2012). Figure 4 shows Q-Q (quantile-quantile) plots for the quantile residuals in both models. Readers can find the information on calculating quantile residuals in the supplementary material. We observe high deviations for beta regression's quantile residuals at the tails, suggesting that this model cannot account for extreme outages. We hypothesize these deviations are high because beta regressions are not suited to include zeros and ones, as discussed earlier. In contrast, quantile residuals of binomial regression show a closer match to the standard normal distribution. We hypothesize that quasi-binomial regressions perform better because they can handle predictions across cities of various sizes, including zeros and ones. Quasi-binomial regressions assign different weights according to the cities' number of customers, unlike the beta regressions that assign the same weight for all (Douma and Weedon 2019).

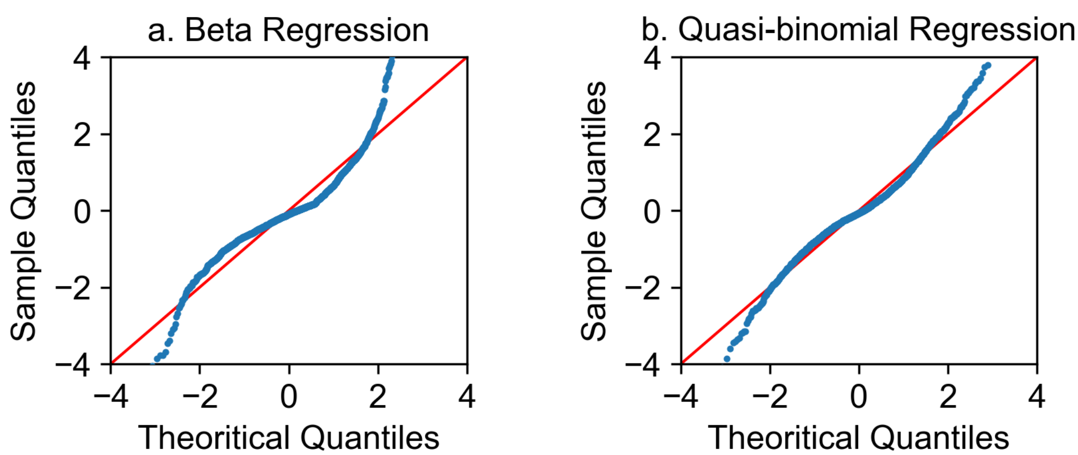

Fig. 4. Q-Q plots for quantile residuals, A: beta regression model, B: quasi-binomial regression model. Quantile residuals should be close to standard normal distribution for a good fit. However, plot A for beta regression shows high deviations for quantile residuals from standard normal distribution suggesting a poor fit for power outage modeling. Quantile residuals in plot B for quasi-binomial regression have a distribution closer to the standard normal distribution, suggesting a better fit for power outage modeling.

Next, we conducted a quantitative comparison of the models' performance through 5-fold cross-validation (Hastie et al. 2002), comparing the mean cross-validation root mean squared errors (RMSE) (Entekhabi et al. 2009). RMSE is given by

$$RMSE = \sqrt{\frac{1}{n}\sum(y-\hat{y})^2} \tag{15}$$

where y is the observed fraction of customers without power, $\hat{y}$ is the predicted fraction of customers without power, and n is the total number of observations. We obtained a 22% lower cross-validation RMSE for the quasi-binomial than for the beta regression (0.23 versus 0.28), further showing that the quasi-binomial regression can better model power outages. Hence, we use it for the rest of the paper.

Finally, we filtered out additional variables to control for overfitting on the quasi-binomial regression. We fitted different regressions with the variables selected previously and then dropped one variable at a time. We performed 5-fold cross-validation (Hastie et al. 2002) and compare the mean cross-validation RMSE for all the models (Figure 5). To obtain the feature importance, first, we subtracted the RMSE of each model and the RMSE of the base model with all variables from feature selection. Removing the most important feature will result in the highest increase in RMSE. Finally, we normalized the RMSE with the largest difference across all the models.

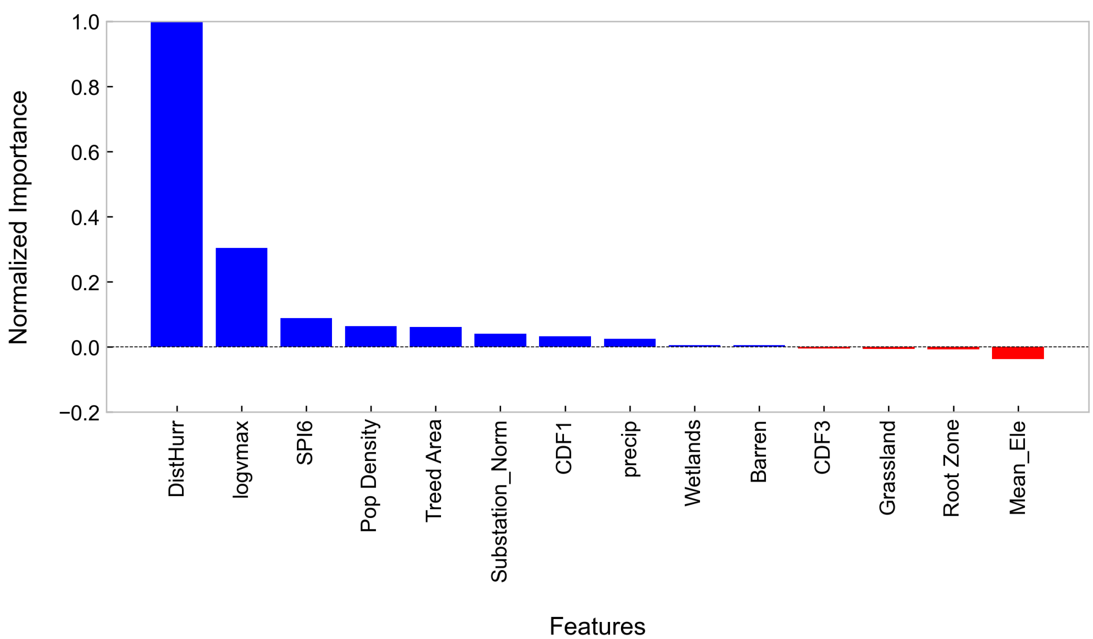

Fig. 5. Feature importance for all the variables selected after filtering in feature selection. Distance to the location of a hurricane crossing the coastline normalized by the radius of maximum winds and the logarithm of maximum winds observed are the most important features to determine power outages. Features with red colored bars were filtered out as removal of these features resulted in reduced cross-validation RMSE. Features with blue colored bars were selected for further analysis.

We show that distance to the hurricane crossing the coastline(DistHurr) and wind gusts (logvmax) are the most essential variables as their exclusion resulted in an RMSE increase of 8.4% and 24.6 % (Supplementary Table S1). The increase in RMSE after dropping DistHurr and logvmax underlines the importance of these variables directly correlated to the significant damage that winds and storm surges can cause to power grids. In contrast, excluding grassland area, rootzone, mean elevation, and soil moisture for the depth of 60-100 cm resulted in lower RMSE than a model with all the variables. Thus, we removed them.

Previous studies have shown that the reliability of power systems can be lower in inland water areas (Ankit et al. 2021), as water can corrode the electric equipment, and the flow of debris coming from disposals during the hurricane can further damage electrical equipment. We confirmed these are relevant features, as excluding land cover parameter wetlands leads to a higher estimate of RMSE. Also, population density and system information variables can indicate urban areas with more power system resources (Brown 2002). We also observed that these are critical variables in our analysis, as removing those two parameters led to a higher RMSE. We observed an RMSE value of 0.235 for 5-fold cross-validation, filtering out grassland area, rootzone, mean elevation, and (Supplementary Table S1). We obtained 261.5 for ϕ, which represents an overdispersion in outage data, a value that is consistent with previous research (Liu et al. 2005; Han et al. 2009a; Han et al. 2009b; Arora and Ceferino 2023b). The final selected variables are 1) logvmax, 2) DistHurr, 3) Barren, 4) Wetlands, 5) SPI6, 6) CDF1, 7) precip, 8) Treed_area, 9) Sub_Norm, 10) Pop_Den. All the selected variables are significant at a p-value of 0.01.

VALIDATION OF PROPOSED MODEL

We further validated the quasi-binomial model with outage data in Florida from Hurricane Ian (2022) (Florida Public Service Commission 2022). In addition, we also compared the predictions of quasi-binomial regression to two other state-of-the-art models for outage predictions: negative binomial regression (Liu et al. 2005; Han et al. 2009a; Han et al. 2009b) and random forest model (McRoberts et al. 2018; Shashaani et al. 2018). Figure 6, presents the predictions with quasi-binomial, negative binomial, random forest, and actual outages in Florida from Hurricane Ian (2022).

Since the resolution of our developed model is at the city level, and the actual outages report is available at the county level, we aggregated the predicted outages at the county level to compare them with the actual outages. For quasi-binomial and random forest regression models, we obtained the mean fraction of customers without power by estimating the population-weighted mean of the percentage of outages at the city level. We obtained the predictions for negative binomial as the count of outages. Hence, for negative binomial predictions, we summed the outages at the city level to obtain the total outages at the county resolution and calculated percentage outages by taking the ratio with total customers in a county. We report the total outages predicted with quasi-binomial, random forest, and negative binomial regression in Table 2. We also reported the percent error on the total outage counts estimated as

$$Error\ on\ total\ outages = \left| \frac{\sum m_i \hat{y}_i - \sum m_i y_i}{\sum m_i y_i} \right| \times 100 \qquad (16)$$

where $\sum m_i \hat{y}_i$ are the total outages predicted by a regression model and $\sum m_i y_i$ are the total outages in Florida from Hurricane Ian (2022). We also reported the RMSE on the fraction of customers without power in counties. The quasi-binomial model predicted the outages with the least error (7.04% error on total outages and RMSE of 0.13 on the fraction of outages) compared to Negative Binomial (61.2% error on total

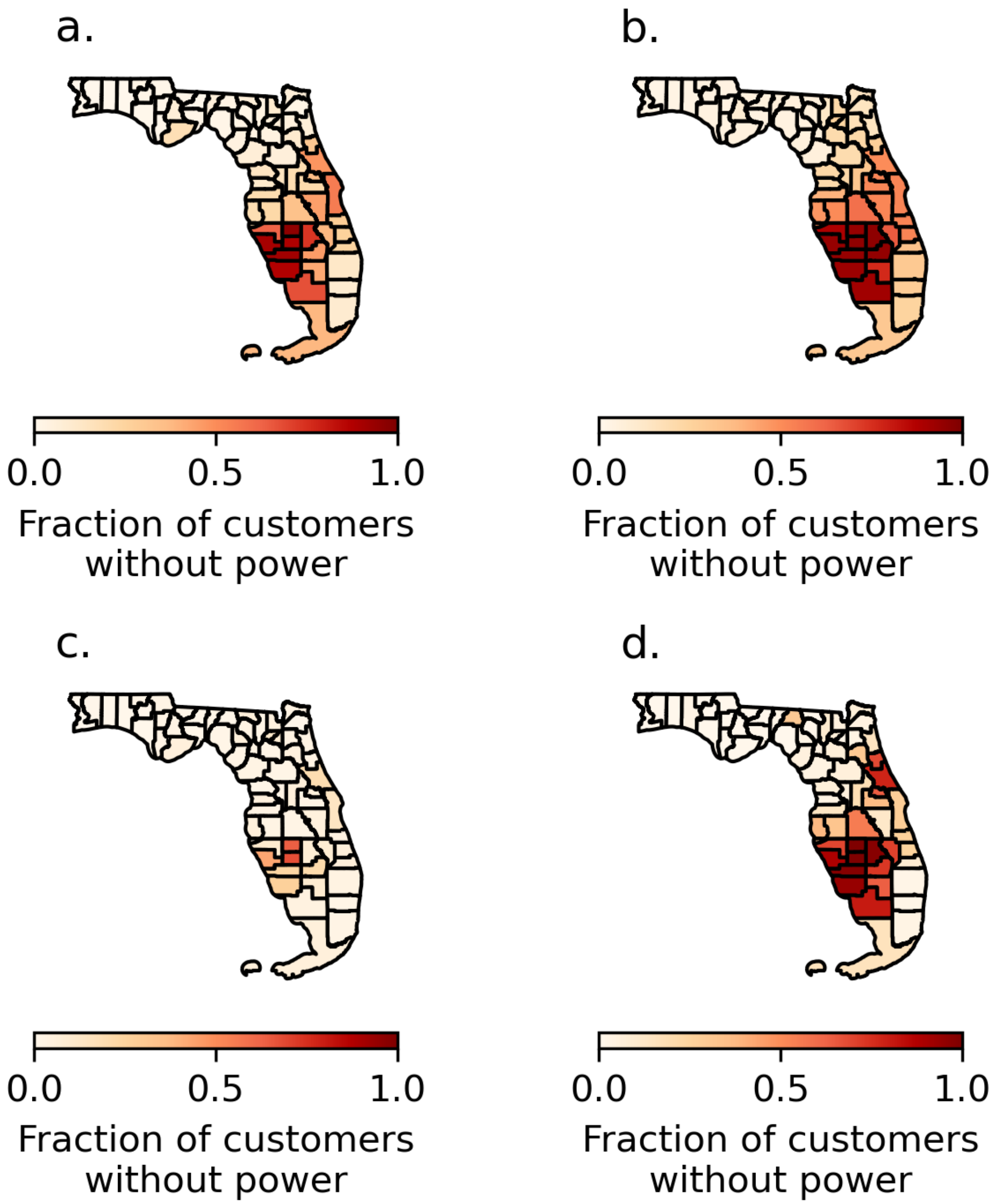

Fig. 6. Predicted outages (a, b, and c), and actual outages (d) in aftermaths of hurricane Ian (2022) in Florida. a.Quasi-binomial regression model predicted outages with an RMSE of 0.13 for percent outages in a county. b. Random forest regression model predicted outages with an RMSE of 0.15 c. The negative binomial regression model predicted outages with an RMSE of 0.28.

outages and RMSE of 0.28) and Random Forest Regression models (49.7% error on total outages and RMSE of 0.15).

Thus, the quasi-binomial regression showed the best model performance. Still, this proposed model can be improved as we observed outage over-predictions in Osceola, Brevard County, and nearby areas. After Hurricane Irma (2018), utilities have hardened the grid by undergrounding the vulnerable distribution lines (Pallone 2020). However, our model has mainly been fitted to power outages in regions with overhead lines. In the future, researchers can enhance power outage prediction with more specific information on the deployment of underground lines.

Regression Model	Total Outages	RMSE	Error on Total Outages
Quasi-Binomial	2,886,413	12.32%	7.04%
Random Forest	4,649,134	14.85%	49.72%
Negative Binomial	735,768	28.12%	76.30%
Actual Outages	3,105,294	-	-

Table 2. Reported Results for different models. Quasi-binomial regression model better predicted the outages across Florida during Hurricane Ian (2022) compared to Random Forest and Negative Binomial regression models. Percent error on total outages and RMSE on percent outages in a county are calculated using Eqs. 16 and 15 respectively.

OVERCOMING LIMITATIONS OF EXISTING OUTAGE MODELS

Arora and Ceferino (2023b) discussed the limitations of the existing state-of-the-art power outage models. Here, we show that the quasi-binomial model can overcome those limitations.

Unbounded predictions from negative binomial GAMs

Arora and Ceferino (2023b) showed that Negative Binomial can overpredict outages to values even larger than the number of customers since it had no upper bound for Hurricane Isaias (2020) in New Jersey. Similarly, negative binomial GAM predicted more outages than the number of customers in 7 cities in Florida during Hurricane Ian (2022), with an average overprediction ratio of 5.7. For example, the negative binomial regression model predicted 10,538 outages in Boca Grande, Florida, 7.6 times more than the number of customers, which is just 1,385.

Conversely, the quasi-binomial model makes bounded predictions between 0 and 1 for the fraction of customers without power. Therefore, we did not observe any outage predictions larger than the number of customers.

Lack of extrapolability with random forest

We expect more damage to the power systems with increased wind speeds, *i.e.*, a monotonic increase in outages with wind speeds. Arora and Ceferino (2023b) showed that random forest models can fail to capture these monotonic increases, leading to poor extrapolation of outage predictions for high and low winds when trained with limited outage data. To study this issue, we plotted how the random forest and quasi-binomial models capture the relationship between winds and outage predictions (Figure 7). We obtained the partial dependence plot by varying the wind speeds from 0 m/s to 120 m/s while keeping all other variables constant and taking the average outage predictions over all the observations in the training data. In this case, the wind data for training have a 5th and 95th percentile of 11.4 m/s and 62.5 m/s. For the random forest model, we observe an almost monotonic increase in the fraction of customers without power within this range. However, we observe poor extrapolations outside. For example, we do not see a decrease in outages below winds of 11.4 m/s, and predictions for wind speeds of 0 m/s is above 0. Also, the maximum outages saturate at 0.85 with the random forest at 62.5 m/s. It is rare to observe high wind events in the US, as even the most wind-affected areas, such as Florida, have a return period of 100 years for winds of 60 m/s. The sparseness of the power outage data from hurricanes makes it difficult for the non-parametric random forest model to follow the simple mechanics of power infrastructure failures with higher winds.

On the contrary, we observed a monotonic increase in risk for outages with wind speeds in Figure 7a for the quasi-binomial model. The outages converge to zero at wind speeds of 0 m/s and continue to grow with increased wind speeds for the quasi-binomial model.

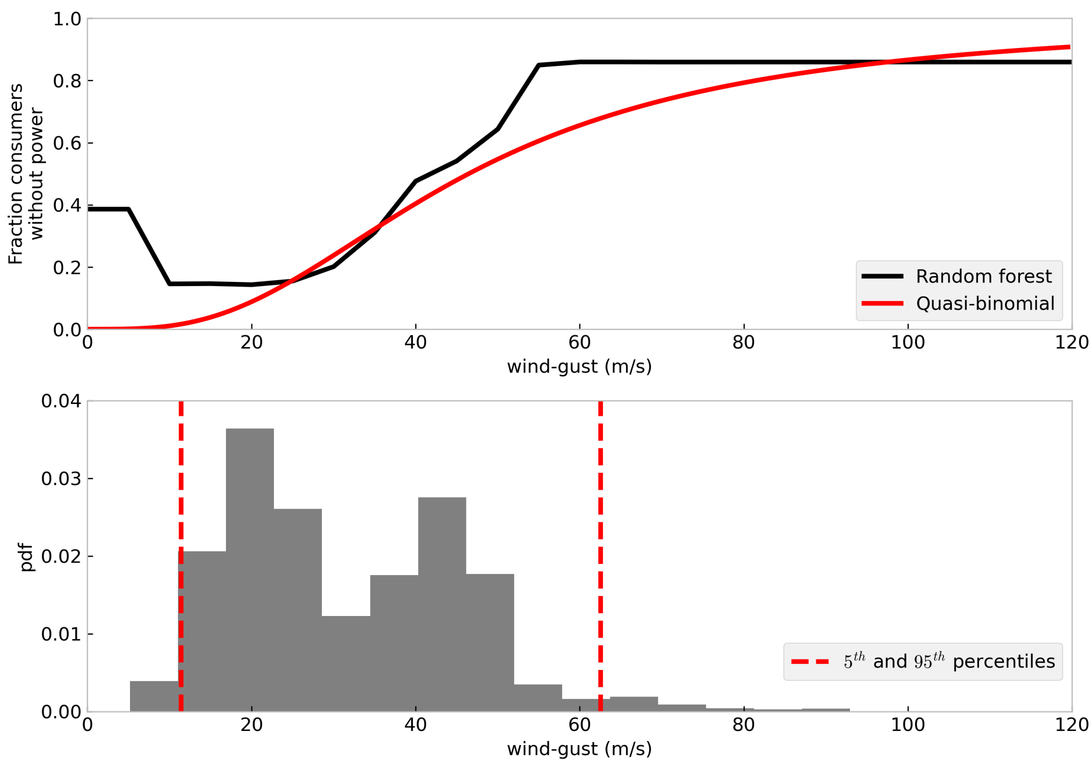

Fig. 7. a. Partial dependence of outages with winds for random forest and quasi-binomial models, b. Distribution of the wind speeds in the training data. Random forest fails to extrapolate for outages and winds outside the range of training data.

Uncertainty Quantification in Outage Predictions

Arora and Ceferino (2023b) also discussed the inability of negative binomial GAMs and random forest models to predict uncertainties in power outage predictions that follow the mechanics of infrastructure failures. Extreme events such as Hurricane Ian (2022), a category 5 hurricane with wind gusts over 70 m/s, are quite likely to devastate power systems. When Hurricane Ian (2022) made landfall, close to 100% customers in Charlotte County lost power. Another similar example is Hurricane Maria (2018) which destroyed power systems in Puerto Rico (Campbell 2018). On the contrary, we will likely observe no damage to power infrastructure at wind speeds closer to zero. The power system is structurally designed to resist at least wind speeds of 20 m/s (Bjarnadottir et al. 2013; IEEE 2007). Thus, we expect outage fractions of zero and close to one at low (close to zero) and high winds ($> 70m/s$), respectively, with very high certainty.

We explore how uncertainty in outage predictions is captured by quasi-binomial, random forest, and negative binomial models by analyzing predicted standard deviations for outages at different wind speeds in New Jersey (Figure 8). We vary winds keeping other variables constant for a clean comparison, and show the 90% confidence intervals on predictions on fractions of customers without power in Tuckerton Borough, New Jersey.

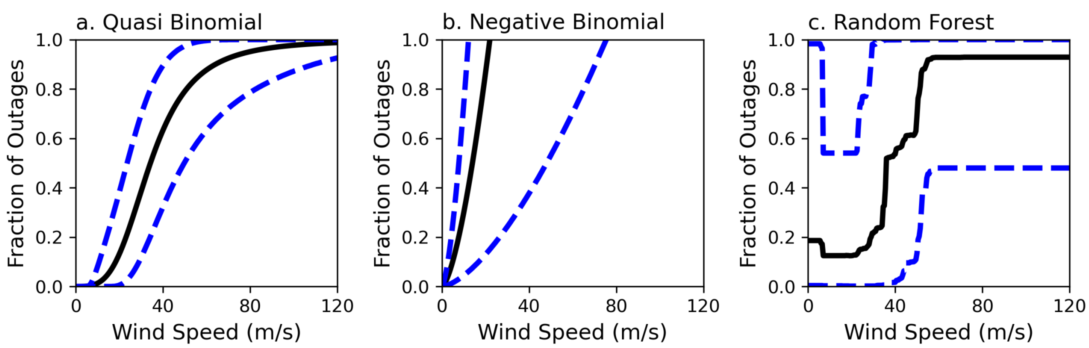

Fig. 8. Uncertainty quantification in power outages for different models. Solid black represents the mean fraction of customers without power, and blue dotted lines represent the 95% confidence intervals on outage predictions. a. Quasi-binomial regression method: shows lower uncertainty in outage predictions at low and high winds. b. Negative binomial regression method: shows lower uncertainty at low winds but large uncertainties at high winds. c.Random forest regression method: shows high uncertainties at low and high winds.

We rescaled the negative binomial predictions to compare the fraction of customers without power for all models. We truncated the predictions on the outage fraction for the negative binomial at 1 as there is no intrinsic upper bound in the model. For the negative binomial function, variance in outages grows as a function of the square of the mean outages (Arora and Ceferino 2023b) (Figure 8b). As discussed earlier, we expect small uncertainties (e.g., small confidence interval ranges) for predictions of outage fractions close to 0 (e.g., at winds close to $0m/s$) and 100% (e.g., at high winds of $> 70m/s$). For negative binomial, the 90% confidence interval's range is close to zero at low wind speeds, consistent with the expected little to no damage frequently observed in power systems at low wind speeds. However, we observed a very high variability with increasing wind speeds, e.g., the 90% confidence interval ranges from 0.12 to 0.62 for wind speeds of $80 m/s$. The high variability at high winds for negative binomial does not agree with the mechanics of infrastructure failures as complete damage is expected at high winds (>70m/s) with high certainty.

We used quantile random forest (QRF) regression to obtain confidence intervals on predictions for a fraction of outages. QRF gives confidence intervals purely from the data because the random forest is non-parametric and does not assume any underlying probability distribution. Random forest predicts the outage with a 90% confidence interval with a wide range (almost from 0 to 1) at winds close to zero and a range of 0.52 at high winds (from 0.48 to 1). Thus, random forest fails to account for the physics of the structural behavior of power systems at low and high winds. It is also important to note from Figure 8c risk of outages does not grow monotonically. Outage fraction saturates at 0.8 for the random forest regression, showing the limited capability to capture the risk correctly at high winds, consistent with observations in Arora and Ceferino (2023b).

In contrast, quasi-binomial predictions show a small range in the 90% confidence interval for outage predictions at low and high winds. Also, the predictions converge to 1 with a high probability at large winds. We also observed a monotonic increase in the risk of outages with wind speeds. Since quasi-binomial predictions are bounded between 0 and 1, predicted outages never exceed 100% of customers without power. Therefore, the quasi-binomial model overcomes the three limitations of state-of-art outage models for power outage predictions following closer the physics of infrastructure failures under hurricanes.

RESILIENCE WITH REDUNDANT POWER INFRASTRUCTURE

We investigated whether the quasi-binomial model captures how the power system configuration affects its vulnerability to hurricanes. To cleanly study the effect of redundancy, we explore two neighboring cities in New Jersey with similar population densities but different numbers of substations. We studied Upper Freehold Township, with 17 substations in the county, with 0.27 substations per 10,000 population and a population density of $0.06/1000m^2$, and Hamilton Township, with 25 substations in its county, with 0.67 substations per 10,000 population and a population density of $0.83/1000m^2$. These are neighboring cities where the DistHurr parameter (storm surge effect parameter) will be almost similar, and thus, variation in winds will determine the variation in the extent of damage to power systems. We plotted the fragility for Upper Freehold Township and Hamilton Township in Figure 10. We observe that more substations improve the resilience of power systems to hurricanes. For example, at a 100-year return wind gust of 32 m/s for these cities, the fraction of outages is 0.27 (40.7% less) versus 0.51 for Hamilton Township than for Upper Freehold Township (25 substations versus 17). More substations generally indicate more transformers and feeders supplying power to customers, which implies that the Hamilton Townships' grid has more redundancy. We assume four feeders are being served through one substation (Brown 2002). Thus, about one feeder and three feeders per 10,000 customers in Upper Freehold Township (0.27 substations per 10,000 customers) and Hamilton Township (0.67 substations per 10,000 customers), respectively.

To illustrate this point better, we draw the analogy to an example circuit shown in

Figure 9 with feeders connected to poles (yellow circles) to supply power to customers (red circles). In Figure 9a, all the customers receive power from one feeder. Thus, a 100% outage will occur when the feeder is inoperable. Similarly, 100% outage will occur for the circuit shown in Figure9b when all three feeders are inoperable. We obtained the probability of failure of one pole given wind speed from Bjarnadottir et al. (2013). If p is the probability of failure of one pole, and the failure is independent of other poles, then the probability of failure of k poles is p^k. We presented the probability of outages for the example circuit in Figure 10b. We observe that redundancy in a circuit with more feeders reduces the circuit's vulnerability to outages. For example, a community with one feeder has 24% probability of 100% outages at 32 m/s compared to 0% chances with three feeders. Note that both Townships in Figure 10a have non-zero outages at 20 m/s winds, as during a hurricane, other factors such as trees can also cause power outages. Thus, our presented power outage model can capture redundancy effects even with limited information on the power systems, e.g., the number of substations.

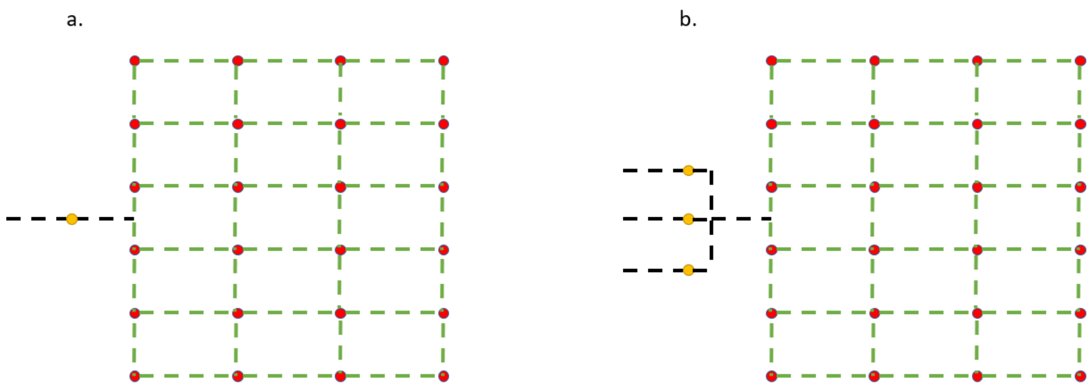

Fig. 9. Example Circuit to study the change in risk of hurricane-induced outages to communities by introducing redundancy in power systems a. power is supplied to customers through one feeder b. power is supplied to customers through three feeders

CONCLUSION

This paper presented a quasi-binomial regression-based hurricane-induced power outage prediction model to improve pre-hurricane response planning to outages. We also tested the beta regression model for the power outage model, but the beta model had 22% higher cross-validation RMSE than the quasi-binomial model. We sought to capture the underlying mechanics of power infrastructure failures with the presented quasi-binomial regression model. The presented outage model addresses the shortcomings of existing power outage models: (1) unbounded outage predictions, (2) poor extrapolation for higher winds, (3) variance estimates in outage predictions unrepresentative of the mechanics of infrastructure failures, and (4) limited ability to account for storm surge vulnerability to power systems. The presented outage model can predict the outage for large power infrastructures at the city level with a lower bound of zero to an upper bound of one for the fraction of customers without power. The quasi-binomial regression model

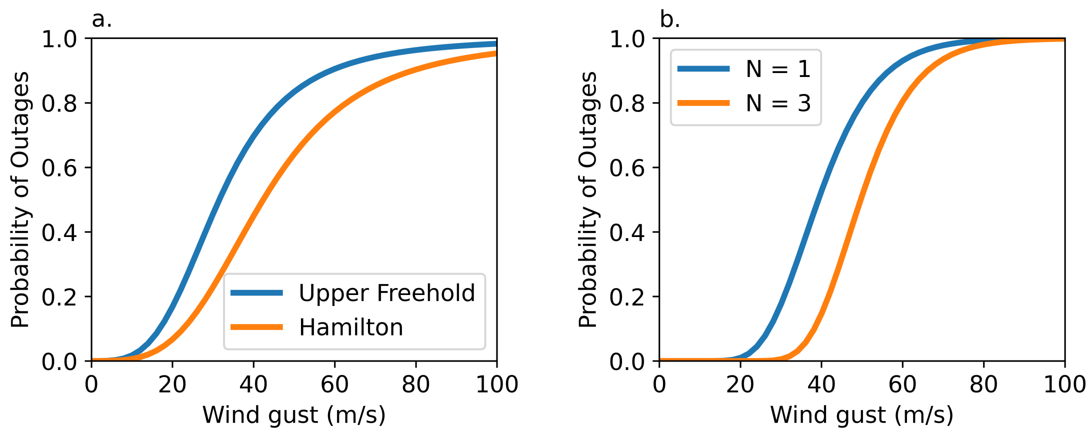

Fig. 10. Fragility comparison with the different number of power components: a. Upper Freehold Township with 17 substations in its county, and Hamilton Township with 25 substations in its county. b. fragility curve for one and three numbers of feeders for the example circuits in Figure 9

also showed a monotonic increase in the risk of outages with gust winds speeds. Finally, the quasi-binomial regression model shows small uncertainty for outage predictions at low and high winds.

We utilize the historical outage datasets from the states of New Jersey, New York, Florida, Texas, and Louisiana in the United States to develop an outage model applicable to multiple cities prone to hurricane-induced outages. The outage modeling depends on different input parameters, including hurricane, environmental, demographic conditions, and power system information. Since many input parameters are only available after the hurricane, we studied outage predictions with limited parameters to enable outage predictions during hurricanes' early warning, including only wind speed, the minimum distance of city to hurricane crossing land normalized with radius to maximum winds, precipitation, barren area, wetlands area, soil moisture, standard precipitation index, percent treed area, substation information, and population density. The most important variables for the outage predictions were wind speeds and the minimum distance of city to hurricane crossing land normalized with radius to maximum winds, as these two factors are correlated to damage effects to power systems from hurricane winds and storm surges.

We compared the performance of the proposed quasi-binomial regression outage model with state-of-the-art outage models, i.e., the negative binomial and random forest models. We compared the models' performance for Hurricane Ian (2022) in Florida. While the negative binomial model and random forest predicted outages with an error of 76.30% and 49.72%, respectively, the Quasi-Binomial regression model had an error of 7.04%, predicting 2,886,413 outages for Hurricane Ian in Florida, close to the actual 3,105,294 outages.

We illustrated how our model captures the effect of the redundancy of power systems on hurricane-induced outages. We investigated the variation of risk of outages in cities with different numbers of substations as a proxy for redundancy. We observed that cities with more substations have fewer customers without electricity. For example, Mercer County has 25 substations, and Monmouth County has 17. At 32 m/s of wind speed, Hamliton Township in Mercer County will have $\sim$ 41% lesser percent of customers without power than Upper Freehold Township in Monmouth County. Thus, power grids with redundant components such as parallel or meshed grids (Brown 2002) can increase the resilience of power systems to extreme weather events.

SUPPLEMENTARY INFORMATION

Supplementary Information for the article is provided in a separate file.

DATA AVAILABILITY

All codes supporting this study's findings are available from the corresponding author upon reasonable request. Power outage data is obtained from PowerOutage (PowerOutage 2022). The source for all other data is provided in the article, and data is available from authors upon reasonable request.

ACKNOWLEDGEMENTS

The authors are thankful for the financial support provided by the NYU Tandon School of Engineering Fellowship and NYU CUSP Dissertation Fellowship. This research is also partially supported by Coalition for Disaster Resilient Infrastructure Fellowship (Grant No. 201924669).

REFERENCES

AJOT (2021). "Hurricane Ida caused at least 1.2 million electricity customers to lose power | AJOT.COM, <https://ajot.com/news/hurricane-ida-caused-at-least-1.2-million-electricity-customers-to-lose-power> (9).

American Community Survey (2019). "American Community Survey, <https://www.census.gov/programs-surveys/acs>.

Ankit, A., Liu, Z., Miles, S. B., and Choe, Y. (2021). "U.S. Power Resilience for 2002–2017." *preprint*.

Arora, P. and Ceferino, L. (2023a). "Could rooftop solar panels and storage have enhanced the electricity resilience during Hurricane Isaias (2020)?." *ICASP 14 Proceedings*.

Arora, P. and Ceferino, L. (2023b). "Probabilistic and machine learning methods for uncertainty quantification in power outage prediction due to extreme events." *Natural Hazards and Earth System Sciences*, 23(5), 1665–1683.

Birchfield, A. B., Gegner, K. M., Xu, T., Shetye, K. S., and Overbye, T. J. (2017). "Statistical Considerations in the Creation of Realistic Synthetic Power Grids for Geomagnetic Disturbance Studies." *IEEE Transactions on Power Systems*, 32(2), 1502–1510.

Bjarnadottir, S., Li, Y., and Stewart, M. G. (2013). "Hurricane Risk Assessment of Power Distribution Poles Considering Impacts of a Changing Climate." *Journal of Infrastructure Systems*, 19(1), 12–24.

Breiman, L. (2001). "Random forests." *Machine Learning*, 45:5–32.

Brown, R. E. (2002). “Electric Power Distribution Reliability, <http://www.dekker.com>.

Bucci, L., Alaka, L., Hagen, A., Delgado, S., and Beven, J. (2022). “National hurricane center tropical cyclone report: Hurricane ian.” *National Hurricane Center*, 1–72.

Campbell, A. F. (2018). “Puerto Rico power restored 11 months after Hurricane Maria - Vox, <https://www.vox.com/identities/2018/8/15/17692414/puerto-rico-power-electricity-restored-hurricane-maria>.

Cangialosi, J. P., Blake, E., DeMaria, M., Penny, A., Latto, A., Rappaport, E., and Tallapragada, V. (2020). “Recent progress in tropical cyclone intensity forecasting at the national hurricane center.” *Weather and Forecasting*, 35(5), 1913–1922.

Casey, S. (2016). “The United States.” *The Ashgate Research Companion to the Korean War*, (August), 49–60.

Ceferino, L., Mitrani-Reiser, J., Kiremidjian, A., Deierlein, G., and Bambarén, C. (2020). “Effective plans for hospital system response to earthquake emergencies.” *Nature Communications 2020 11:1*, 11(1), 1–12.

Chapman, L. (2000). “Assessing topographic exposure.” *Meteorological Applications*, 7(4), 335–340.

Chavas, D. R., Lin, N., and Emanuel, K. (2015). “A model for the complete radial structure of the tropical cyclone wind field. Part I: Comparison with observed structure.” *Journal of the Atmospheric Sciences*, 72(9), 3647–3662.

Congress.gov (2020). “H.R.5760 - 116th Congress (2019-2020): Grid Security Research and Development Act | Congress.gov | Library of Congress, <https://www.congress.gov/bill/116th-congress/house-bill/5760>.

Consul, P. C. (2010). “Communications in Statistics-Theory and Methods On some properties and applications of quasi-binomial distribution ON SOME PROPERTIES AND APPLICATIONS OF QUASI-BINOMIAL DISTRIBUTION.” 1(2), 477–504.

Cortes, M., Arora, P., Ceferino, L., Ibrahim, H., Istrati, D., Reed, D., Roueche, D., Safiey, A., Tomiczek, T., Zisis, I., Alam, M., Kijewski-Correa, T., Prevatt, D., and Robertson, I. (2022). “Steer: Hurricane ian preliminary virtual reconnaissance report (pvrr).

Cruse, G. and Kwasinski, A. (2021). “Statistical Evaluation of Flooding Impact on Power System Restoration Following a Hurricane.” *2021 Resilience Week, RWS 2021 - Proceedings*.

Danielson, J. and Gesh, D. (2011). “Global multi-resolution terrain elevation data 2010 (GMTED2010): U.S. Geological Survey Open-File Report 2011–1073, <https://www.usgs.gov/publications/global-multi-resolution-terrain-elevation-data-2010-gmted2010>.

data.gov (2022). “Substations - Overview, <https://www.arcgis.com/home/item.html?id=755e8c8ae15a4c9a

Douma, J. C. and Weedon, J. T. (2019). “Analysing continuous proportions in ecology and evolution: A practical introduction to beta and Dirichlet regression.” *Methods in Ecology and Evolution*, 10(9), 1412–1430.

Dunn, P. K. and Smyth, G. K. (1996). “Randomized quantile residuals.” *Journal of Computational and Graphical Statistics*, 5(3), 236–244.

Dunn, P. K. and Smyth, G. K. (2018). *Generalized Linear Models With Examples in R, https://link.springer.com/book/10.1007/978-1-4419-0118-7*.

Emergible Consulting Services, Accessed Online: 08/10/2023. “Emergible Consulting Services, <https://ecsind.com/project-execution.php>.

Entekhabi, D., Reichle, R. H., Koster, R. D., and Crow, W. T. (2009). “Performance Metrics for Soil Moisture Retrievals and Application Requirements, <http://www.esa.int/esaLP/LPsmos.html>.

Executive Office of the U.S. President (2022). “Federal Register :: Critical Infrastructure Security and Resilience Month, 2022, <https://www.federalregister.gov/documents/2022/11/03/2022-24148/critical-infrastructure-security-and-resilience-month-2022>.

Ferrari, S. and Cribari-Neto, F. (2004). “Beta regression for modelling rates and proportions.” *Journal of Applied Statistics*, 31(7), 799–815.

Florida Public Service Commission (2022). “Florida Public Service Commission, <https://www.floridapsc.com/>.

Giulia Carbonaro (2022). “Florida Power Outage Map, Update as Hurricane Ian Leaves 2M Without Power, <https://www.newsweek.com/florida-power-outage-map-update-hurricane-ian-leaves-2-million-without-power-1747378>.

Guikema, S. D., Nateghi, R., Quiring, S. M., Staid, A., Reilly, A. C., and Gao, M. (2014). “Predicting Hurricane Power Outages to Support Storm Response Planning.” *IEEE Access*, 2(September 2015), 1364–1373.

Guikema, S. D., Quiring, S. M., and Han, S. R. (2010). “Prestorm Estimation of Hurricane Damage to Electric Power Distribution Systems.” *Risk Analysis*, 30(12), 1744–1752.

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

Han, S. R., Guikema, S. D., and Quiring, S. M. (2009a). “Improving the predictive accuracy of hurricane power outage forecasts using generalized additive models.” *Risk Analysis*, 29(10), 1443–1453.

Han, S. R., Guikema, S. D., Quiring, S. M., Lee, K. H., Rosowsky, D., and Davidson, R. A. (2009b). “Estimating the spatial distribution of power outages during hurricanes in the Gulf coast region.” *Reliability Engineering and System Safety*, 94(2), 199–210.

Haseltine, C. and Eman, E. E. S. (2017). “Prediction of power grid failure using neural network learning.” *Proceedings - 16th IEEE International Conference on Machine Learning and Applications, ICMLA 2017*, 2017-December, 505–510.

Hastie, T., Tibshirani, R., and Friedman, J. (2002). “Springer Series in Statistics The Elements of Statistical Learning Data Mining, Inference, and Prediction.

IEEE (2007). “National Electrical Safety Code, ANSI/IEEE Standard C2-2007.

Jaech, A., Zhang, B., Ostendorf, M., and Kirschen, D. S. (2018). “Real-Time Prediction of the Duration of Distribution System Outages.” *preprint*.

Jin, S., Dewitz, J., Li, C., Sorenson, D., Zhu, Z., Shogib, R., Danielson, P., Granneman, B., Costello, C., Case, A., et al. (2021). “National land cover database 2019: A comprehensive strategy for creating the 1986-2019 forest disturbance date product.

JustEnergy, Accessed Online: 08/10/2023. “JustEnergy, <https://justenergy.com/blog/power-to-choose-texas-electricity/>.

Kira, K. and Rendell, L. A. (1992). “A Practical Approach to Feature Selection.”

Machine Learning Proceedings 1992, 249–256.

Klar, B. and Meintanis, S. G. (2012). “Specification tests for the response distribution in generalized linear models.” *Computational Statistics*, 27(2), 251–267.

Knapp, K. R., Kruk, M. C., Levinson, D. H., Diamond, H. J., and Neumann, C. J. (2010). “The international best track archive for climate stewardship (IBTrACS).” *Bulletin of the American Meteorological Society*, 91(3), 363–376.

Krist Jr., F. J., Ellenwood, J. R., Woods, M. E., Mcmahan, A. J., Cowardin, J. P., Ryerson, D. E., Sapio, F. J., Zweifler, M. O., and Romero, S. A. (2014). “2013–2027 National Insect and Disease Forest Risk Assessment.” (January).

Kumar, V. and Minz, S. (2014). “Smart Computing Review Feature Selection: A literature Review.” *Smart Computing Review*, 4(3).

Lin, N., Emanuel, K., Oppenheimer, M., and Vanmarcke, E. (2012). “Physically based assessment of hurricane surge threat under climate change.” *Nature Climate Change*, 2(6), 462–467.

Liu, F. and Kong, Y. (2015). “zoib: An R Package for Bayesian Inference for Beta Regression and Zero/One Inflated Beta Regression.” *The R Journal*, 7(2), 34–51.

Liu, H., Davidson, R. A., and Apanasovich, T. V. (2007). “Statistical forecasting of electric power restoration times in hurricanes and ice storms.” *IEEE Transactions on Power Systems*, 22(4), 2270–2279.

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.” *Reliability Engineering and System Safety*, 93(6), 897–912.

Liu, H., Davidson, R. A., Rosowsky, D. V., and Stedinger, J. R. (2005). “Negative Binomial Regression of Electric Power Outages in Hurricanes.” *Journal of Infrastructure Systems*, 11(4), 258–267.

Madrigano, J., Ito, K., Johnson, S., Kinney, P. L., and Matte, T. (2015). “A case-only study of vulnerability to heat wave–related mortality in New York City (2000–2011).” *Environmental Health Perspectives*, 123(7), 672–678.

Markhvida, M., Ceferino, L., and Baker, J. W. (2018). “Modeling spatially correlated spectral accelerations at multiple periods using principal component analysis and geostatistics.” *Earthquake Engineering & Structural Dynamics*, 47(5), 1107–1123.

McRoberts, D. B., Quiring, S. M., and Guikema, S. D. (2018). “Improving Hurricane Power Outage Prediction Models Through the Inclusion of Local Environmental Factors.” *Risk Analysis*, 38(12), 2722–2737.

Miller, C., Gibbons, M., Beatty, K., and Boissonnade, A. (2013). “Topographic speed-up effects and observed roof damage on Bermuda following Hurricane Fabian (2003).” *Weather and Forecasting*, 28(1), 159–174.

Nateghi, R., Guikema, S., and Quiring, S. M. (2014). “Power Outage Estimation for Tropical Cyclones: Improved Accuracy with Simpler Models.” *Risk Analysis*, 34(6), 1069–1078.

National Academies of Sciences, Engineering, and Medicine (2017). “Enhancing the Resilience of the Nation’s Electricity System.” *Enhancing the Resilience of the Nation’s Electricity System*.

Office, E. and August, P. (2014). “Economic benefits of increasing electric grid resilience to weather outages.” *Climate, Energy, and Environment: Issues, Analyses,*

and Developments, 2(August), 89–118.
Olkin, I. and Liu, R. (2003). “A bivariate beta distribution.” *Statistics and Probability Letters*, 62(4), 407–412.
Ospina, R. and Ferrari, S. L. (2012). “A general class of zero-or-one inflated beta regression models.” *Computational Statistics & Data Analysis*, 56(6), 1609–1623.
Pachev, B., Arora, P., del Castillo-Negrete, C., Valseth, E., and Dawson, C. (2023). “A framework for flexible peak storm surge prediction.
Pahwa, S., Scoglio, C., and Scala, A. (2014). “Abruptness of Cascade Failures in Power Grids.” *Scientific Reports 2014 4:1*, 4(1), 1–9.
Pallone, G. (2020). “FPL Crews Install Underground Power Lines for Pilot Program, <https://www.mynews13.com/fl/orlando/news/2020/10/27/florida-power-and-light-underground-power-lines-pilot-program>.
Pereira, G. H. (2019). “On quantile residuals in beta regression.” *Communications in Statistics-Simulation and Computation*, 48(1), 302–316.
Petersen, H. C. (1982). “Electricity Consumption in Rural Vs. Urban Areas.” *Western Journal of Agricultural Economics*, 07(01), 13–18.
Pisano, G., Chowdhury, N., Coppo, M., Natale, N., Petretto, G., Soma, G. G., Turri, R., and Pilo, F. (2019). “Synthetic models of distribution networks based on open data and georeferenced information.” *Energies*, 12(23).
Poudyal, A., Wertz, C., Nguyen, A., Mahmud, S. U., Gunturi, V., and Dubey, A. (2022). “Spatiotemporal impact analysis of hurricanes and storm surges on power systems.
PowerOutage (2022). “POWEROUTAGE.US, <www.poweroutage.us>.
Provost, F. (1999). “Distributed Data Mining: Scaling up and beyond.
Quiring, S. M., Zhu, L., and Guikema, S. D. (2011). “Importance of soil and elevation characteristics for modeling hurricane-induced power outages.” *Natural Hazards*, 58(1), 365–390.
Schultz, P., Heitzig, J., and Kurths, J. (2014). “A random growth model for power grids and other spatially embedded infrastructure networks.” *European Physical Journal: Special Topics*, 223(12), 2593–2610.
Schweitzer, E., Member, S. S., Scaglione, A., Monti, A., Member, S. S., Pagani, G. A., Wang, Z., Scaglione, A., and Thomas, R. J. (2010). “Medium Voltage Radial Distribution Systems.” *IEEE Transactions on Smart Grid*, 1(1), 28–39.
Shashaani, S., Guikema, S. D., Zhai, C., Pino, J. V., and Quiring, S. M. (2018). “Multi-Stage Prediction for Zero-Inflated Hurricane Induced Power Outages.” *IEEE Access*, 6, 62432–62449.
Smith, A. B. (2020). “U.S. Billion-dollar Weather and Climate Disasters, 1980 - present (NCEI Accession 0209268).” *National Centers for Environmental Information*.
Smithson, M. and Verkuilen, J. (2006). “A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables.” *Psychological methods*, 11(1), 54–71.
Soil Survey Staff (2021). “Gridded Soil Survey Geographic Database (gSSURGO) | Ag Data Commons. Available online. Accessed [08/01/2021], <https://data.nal.usda.gov/dataset/gridded-soil-survey-geographic-database-gssurgo>.
States, U. and Accountability, G. (2020). “Critical infrastructure protection: Actions

needed to address significant cybersecurity risks facing the electric grid." *Key Government Reports. Volume 56*, (August), 153–239.

Sun, H., Wang, Z., Wang, J., Huang, Z., Carrington, N., and Liao, J. (2016). "Data-Driven Power Outage Detection by Social Sensors." *IEEE Transactions on Smart Grid*, 7(5), 2516–2524.

Swearingen, C. J., Melguizo Castro, M. S., and Bursac, Z. (2012). "Inflated Beta Regression: Zero, One, and Everything in Between." *SAS Global Forum*, 325–2012.

Tonn, G. L., Guikema, S. D., Ferreira, C. M., and Quiring, S. M. (2016). "Hurricane Isaac: A Longitudinal Analysis of Storm Characteristics and Power Outage Risk." *Risk Analysis*, 36(10), 1936–1947.

twitter.com (2020). "PowerOutage.us on Twitter: "Over 3 million electric customers are now without power in the #USA from #Isaias as it continues to move up the east coast. With 1.3 million customers out in #NewJersey. Check out https://t.co/8cAFt3zGJe for detailed #PowerOutage data.[2020-08-04 4:03 PM EDT] https://t.co/eybtCatm23" / Twitter, <https://twitter.com/PowerOutage_us/status/1290744180956901379>.

USGS (2004). "SCIENCE AND TECHNOLOGY TO SUPPORT FRESH WATER AVAILABILITY IN THE UNITED STATES.

Valenzuela, A., Inga, E., and Simani, S. (2019). "Planning of a resilient underground distribution network using georeferenced data." *Energies*, 12(4), 1–20.

Wanik, D. W., Parent, J. R., Anagnostou, E. N., and Hartman, B. M. (2017). "Using vegetation management and LiDAR-derived tree height data to improve outage predictions for electric utilities." *Electric Power Systems Research*, 146, 236–245.

Winkler, J., Dueñas-Osorio, L., Stein, R., and Subramanian, D. (2010). "Performance assessment of topologically diverse power systems subjected to hurricane events." *Reliability Engineering & System Safety*, 95(4), 323–336.

Wood, S. N. (2017). "Generalized additive models: An introduction with R, second edition." *Generalized Additive Models: An Introduction with R, Second Edition*, 1–476.

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." *International Journal of Climatology*, 27(1), 65–79.

Xia, Y. (2012). "NLDAS VIC Land Surface Model L4 Hourly 0.125 x 0.125 degree V002, <https://doi.org/10.5067/ELBDAPAKNGJ9>.

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D., Koren, V., Duan, Q., Mo, K., Fan, Y., and Mocko, D. (2012). "Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products." *Journal of Geophysical Research: Atmospheres*, 117(D3), 3109.

Xie, J., Alvarez-Fernandez, I., and Sun, W. (2020). "A review of machine learning applications in power system resilience." *IEEE Power and Energy Society General Meeting*, 2020-August.

Yee, T. W. (2012). *Package "VGAM" (Vector generalized linear and additive models)*, <http://www.springer.com/series/692>.

Zhai, C., Chen, T. Y. j., White, A. G., and Guikema, S. D. (2021). "Power outage prediction for natural hazards using synthetic power distribution systems." *Reliability Engineering and System Safety*, 208(October 2020).

Zimmerman, R., Zhu, Q., Francisco De Leon, ., and Guo, Z. (2017). "Conceptual Modeling Framework to Integrate Resilient and Interdependent Infrastructure in Extreme Weather." *Journal of Infrastructure Systems*, 23(4), 04017034.