

Article

# A Spatial Analytics Framework to Investigate Electric Power-Failure Events and Their Causes

Vivian Sultan \* and Brian Hilton

Center for Information Systems and Technology, Claremont Graduate University, Claremont, CA 91711, USA; brian.hilton@cgu.edu

\* Correspondence: vivian.sultan@cgu.edu

Received: 11 December 2019; Accepted: 13 January 2020; Published: 16 January 2020



**Abstract:** The U.S. electric-power infrastructure urgently needs renovation. Recent major power outages in California, New York, Texas, and Florida have highlighted U.S. electric-power unreliability. The media have discussed the U.S. aging power infrastructure and the Public Utilities Commission has demanded a comprehensive review of the causes of recent power outages. This paper explores geographic information systems (GIS) and a spatially enhanced predictive power-outage model to address: How may spatial analytics enhance our understanding of power outages? To answer this research question, we developed a spatial analysis framework that utilities can use to investigate power-failure events and their causes. Analysis revealed areas of statistically significant power outages due to multiple causes. This study's GIS model can help to advance smart-grid reliability by, for example, elucidating power-failure root causes, defining a data-responsive blackout solution, or implementing a continuous monitoring and management solution. We unveil a novel use of spatial analytics to enhance power-outage understanding. Future work may involve connecting to virtually any type of streaming-data feed and transforming GIS applications into frontline decision applications, showing power-outage incidents as they occur. GIS can be a major resource for electronic-inspection systems to lower the duration of customer outages, improve crew response time, as well as reduce labor and overtime costs.

**Keywords:** spatial analytics; power-failure; GIS; smart-grid

## 1. Introduction & Problem Definition

Electric power, in a short time, has become a necessity of modern life. Our work, healthcare, leisure, economy, and livelihood depend on the constant supply of electrical power. Even a temporary power outage can lead to relative chaos, financial setbacks, and the loss of life. Our cities depend on electricity and without the constant supply from the power grid, pandemonium would ensue. Power outages can be especially tragic for life-support systems in hospitals and nursing homes or systems in synchronization facilities such as in airports, train stations, and traffic control. The economic cost of power interruptions to U.S. electricity consumers is \$79 billion annually in damages and lost economic activity [1]. In 2017, the Lawrence Berkeley National Laboratory provided an update and estimated a power-interruption cost increase of more than 68% per year since their 2004 study [2].

Many reasons underlie power failures we are facing today. Among these reasons are severe weather, damage to electric transmission lines, shortage of circuits, and the aging of power-grid infrastructure. In examining some of these reasons, we found that severe weather is the leading cause of power outages in the United States [3]. Last year, weather events as a whole cost U.S. utilities \$306 billion, the highest figure ever recorded by the federal government [4].

The aging of the grid infrastructure is another noteworthy cause of power failure. In 2008, the American Society of Civil Engineers gave the U.S. power-grid infrastructure an unsatisfactory

grade [5]. They stated that the power transmission system in the United States requires immediate attention. Furthermore, the report mentioned that the U.S. electric power grid is similar to those of third-world countries. According to the Electric Power Research Institute, equipment such as transformers controlling power transmission need to be replaced, as they have exceeded their expected lifespan considering the materials' original design [6].

Electrical outages have three main causes [7], namely: (1) hardware and technical failures, (2) environment-related, and (3) human error. Hardware and technical failures (1) are due to equipment overload, short circuits, brownouts, and blackouts, to name a few [8–10]. These failures often align with unmet peak usages, outdated equipment, and malfunctioning back-up power systems. The environment-related (2) causes for power outages comprise weather, wildlife, and trees that come into contact with power lines. Lightning, high winds, and ice are common weather-related power interruptions. Also, squirrels, snakes, and birds that come in contact with equipment such as transformers and fuses can cause equipment to momentarily fail or shut down completely [8]. As for the third main cause for electrical outages, human error (3), the Uptime Institute estimated that human error causes roughly 70% of the problems that plague data centers. Hacking can be included in the human error category [11].

Analytics have been a popular topic in research and practice, particularly in the energy field. The use of analytics can help advance smart grid reliability through, for example, elucidating a root cause of power failure, defining a solution for a blackout through data, or implementing the solution with continuous monitoring and management. In this research paper, we unveil the novel use of analytics to investigate power failure events and their causes. As the objective in this research is to advance smart grid reliability, we specifically explore spatial analytics to offer a spatially enhanced predictive model for power outages.

## 2. Literature Review

The economic cost of power interruptions to U.S. electricity is \$79 billion annually [1]. 2017 was particularly bad for outages with wildfires in California and a number of hurricanes that plagued Texas, the Southeast, and Puerto Rico [12]. According to a report by the Electric Reliability Council of Texas [13], when Hurricane Harvey struck the Gulf Coast, about 280,000 people were without electricity at one point. The report specified that the storm caused six transmission lines and 91 circuits to fail, knocking out about 10,000 MW of generation.

When Hurricane Irma hit Florida, it impacted about five million customers in districts where Florida Power & Light operates [14]. Peter Maloney commented that "Miami-Dade County was hit hardest. At one point, more than 815,000 people, or 80% of FPL accounts in the county, were without power" [15]. According to Maloney, other jurisdictions in Florida, such as Palm Beach and Broward County, also showed loss of power in about 68–70% of accounts, due to the hurricane [15]. Figure 1 sketches the yearly total number of outages in the United States and people affected since 16 February 2008 [16] (p. 3).

In addition, Eaton's Blackout Tracker offered the following pie chart to breakdown 2017 reported power-outage incidents by cause [16] (p. 17). In the annual report, Eaton's Blackout Tracker grouped power-outage incidents into one of eight possible causes. The number next to the pie piece in Figure 2 is the number of outages associated with that cause.

Year	Total number of outages	People affected
2008*	2,169	25.8 million
2009	2,840	13.5 million
2010	3,149	17.5 million
2011	3,071	41.8 million
2012	2,808	25.0 million
2013	3,236	14.0 million
2014	3,634	14.2 million
2015	3,571	13.2 million
2016	3,879	17.9 million
2017	3,526	36.7 million

Figure 1. U.S. reported power outages by year. Source: Eaton’s Blackout Tracker [16].

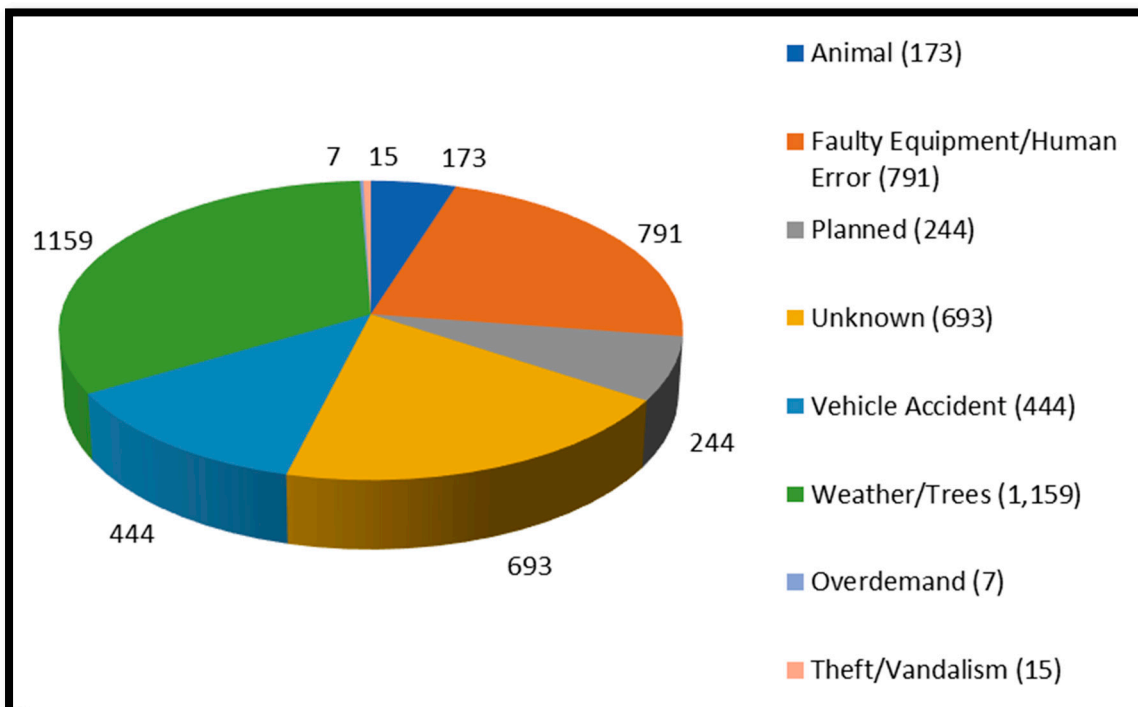


Figure 2. U.S. reported power outages by cause in 2017. Source: Eaton’s Blackout Tracker [16].

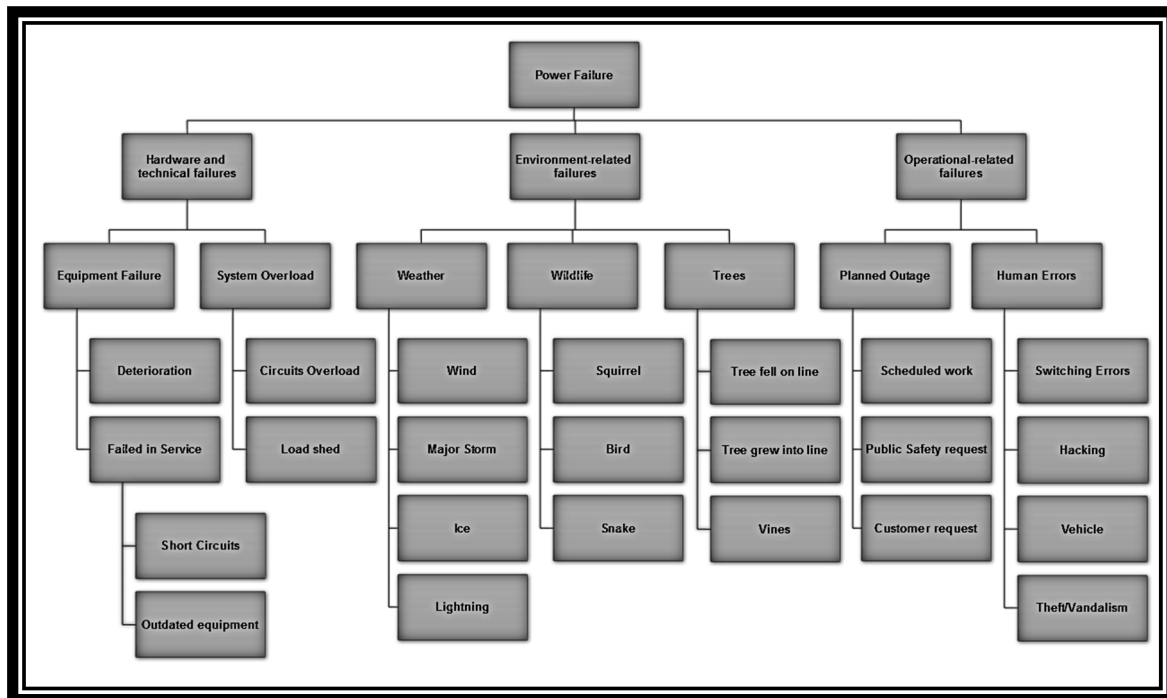
After looking into the Eaton’s Blackout Tracker and other similar reports that investigate power-outage incidents, we identified three key factors underlying these outages (Figure 3): (1) hardware and technical failures, (2) environment-related, and (3) operation-related failures.

Environment-related incidents comprise the largest portion of power-outage causes. Environment-related incidents can be classified into three distinct categories: weather, wildlife, and trees. Wisconsin Public Service [17] delineated the weather-related causes of power outages; a 2005 study by Davies Consulting for the Edison Electric Institute stated that 70% of power outages in the United States are weather-related [18]. Kenward and Raja [19] analyzed power-outage data over a 28-year period and pointed out that between 2003 and 2012, 80% of all outages were caused by weather. Similarly, Campbell [20] highlighted the damage to the electrical grid caused by seasonal storms, rain, and high winds.

According to the President’s Council of Economic Advisers and the U.S. Department of Energy’s Office of Electricity Delivery and Energy Reliability [3], severe weather is the leading cause of power

outages in the United States. “Between 2003 and 2012, an estimated 679 widespread power outages occurred due to severe weather” [3] (p. 3). Likewise, annual costs changed significantly and were greater due to of major storms such as Hurricane Ike in 2008. “Data from the U.S. Energy Information Administration show that weather-related outages have increased significantly since 1992” [3] (p. 8).

In addition to weather, other external forces cause power outages. Falling tree branches, for example, is another important cause for power disruption [21]. Animals coming into contact with power lines, such as large birds, are also important culprits of power outage in the United States [16]. Furthermore, human-error incidents cause power outages. Chayanam [7] indicated that training is essential for technicians and staff to battle outages with proper maintenance procedures.



**Figure 3.** Causes of Power Failure/Prepared by Vivian Sultan/Date Printed 7 January 2018.

Interrupted power supply is no longer deemed a mere inconvenience. As the duration and spatial extent of electricity-system outages increase, costs and inconvenience grow. Critical social services such as medical care, police, and other emergency services and communications systems depend on electricity to function at a minimum level. Failures can bring about catastrophic outcomes, and lives can be lost. Grid reliability is an area of research that will help better explain the causes of outages and prescribe interventions that will improve the reliability of the smart grid. In this manuscript, we use various spatial analytics tools to investigate U.S. power concerns and to answer the research question: How may spatial analytics enhance our understanding of power outages?

To date, there are several studies that delve into several causes of failure. For instance, Reed evaluated how power delivery system dealt with hurricanes [22]. Sun et al. discussed social media data in detecting power outages [23] and Guven et al. discussed how a GIS could help to analyze an electric distribution network [24]. These studies show that there is a palpable interest in using GIS in a data-driven way to deal with power outages. Our research is different and novel in that we provided a general framework to help researchers and practitioners deal with the multitude of data in analyzing power outages and integrated several outage events to detect regions where outage events should be investigated further.

### 3. Data Selection and Acquisition

#### 3.1. Infrastructure Data

The Electric Power Research Institute (EPRI) data repository includes the primary datasets we used to conduct this analysis. The datasets include data from advanced metering systems, supervisory control, and data acquisition (SCADA) systems, geospatial information systems (GIS), outage-management systems (OMS), distribution-management systems (DMS), asset-management systems, work-management systems, customer-information systems, and intelligent electronic-device databases. Access to datasets was provided as part of EPRI's data-mining initiative to provide a test bed for data exploration and innovation and to solve the top challenges faced by the utility industry [25].

When combined with clever analytical techniques, data provide the potential to transform the world into a smarter world, where the prevention of power outages may become a true reality, not merely a prediction. The SCADA/OMS/DMS archives at a power utility offer the required data to identify the parts of the system that contribute most to overall system downtime. OMS, for example, provides the data needed to calculate measurements of system reliability. OMS also provide historical data that can be mined to find common causes, failures, and damages. Since OMS have become more integrated with other operational systems such as GIS at the utility side, analysis has become more feasible, so this research may aim to improve grid reliability.

In general, EPRI's data consist of one main GIS data file, where this data is linked with seven databases within the EPRI's information systems. While some links are based on unique key association, several are linked spatially, via latitude and longitude. For example, the GIS data can be linked with the outage management system allowing for the analysis of the root cause of outages.

#### 3.2. Weather Data

1. Georgia spatial data infrastructure (GaSDI) and the Georgia GIS Clearinghouse is the data source for the monthly temperature and precipitation data we employed in this study. "This dataset contains contours that represent the average monthly temperatures (1960–1991) for the state of Georgia [and] display appropriate at least at regional scales and above" [26]. The data repository is displayed in Figure 4.

Search Results				
Records 1-50 of 52				1 next >>
Click the Access icon to download data. To preview, click the Title.				
Access	Extent	Title	Year	Scale
FREE	Georgia	Average April Maximum Temperatures : 1960-1991	1999	250,000
FREE	Georgia	Average April Mean Temperatures : 1960-1991	1999	250,000
FREE	Georgia	Average April Minimum Temperatures : 1960-1991	1999	250,000
FREE	Georgia	Average April Precipitation : 1960-1991	1999	250,000
FREE	Georgia	Average August Maximum Temperatures : 1960-1991	1999	250,000
FREE	Georgia	Average August Mean Temperatures : 1960-1991	1999	250,000
FREE	Georgia	Average August Minimum Temperatures : 1960-1991	1999	250,000
FREE	Georgia	Average August Precipitation : 1960-1991	1999	250,000
FREE	Georgia	Average December Maximum Temperatures : 1960-1991	1999	250,000
FREE	Georgia	Average December Mean Temperatures : 1960-1991	1999	250,000
FREE	Georgia	Average December Minimum Temperatures : 1960-1991	1999	250,000
FREE	Georgia	Average December Precipitation : 1960-1991	1999	250,000
FREE	Georgia	Average February Maximum Temperatures : 1960-1991	1999	250,000
FREE	Georgia	Average February Mean Temperatures : 1960-1991	1999	250,000
FREE	Georgia	Average February Minimum Temperatures : 1960-1991	1999	250,000
FREE	Georgia	Average February Precipitation : 1960-1991	1999	250,000
FREE	Georgia	Average January Maximum Temperatures : 1960-1991	1999	250,000
FREE	Georgia	Average January Mean Temperatures : 1960-1991	1999	250,000

**Figure 4.** Monthly Temperature and Precipitation Data Files. Source: Georgia Spatial Data Infrastructure [26].

2. The National Oceanic and Atmospheric Administration website (NOAA) is the data source for the storm events and storm details. The link to the NOAA Storm Events Database is <https://www.ncdc.noaa.gov/stormevents/>. According to NOAA's National Centers for Environmental

Information [27], this database contains records used to create the official NOAA Storm Data publication, detailing:

- a. The event of storms and other noteworthy weather phenomena;
- b. Odd, scarce, weather phenomena that generate media attention; and
- c. Other important meteorological events, such as record maximum or minimum temperatures or precipitation that occurs in connection with another event.

#### 4. Methodology

After the data has been obtained, preliminarily, data should be checked for inconsistencies, errors, and omissions. Since this type of analytical project requires data to be analyzed spatially, any tool used needs to be suitable for location analytics. One such tool is the ArcGIS platform, provided by the Environmental Systems Research Institute (ESRI). Therefore, we opt to use this tool as a demonstration of the analytical framework that others could adopt. In addition, the analytical framework needs a way to make sure that the steps are reusable. Thus, the ModelBuilder tool in the ArcMap software is utilized to create three models that could be exported and use in other similar analyses.

The following subsections detailed the creation of the geodatabase to store all acquired data, and the necessary steps to combine, sort, clean, and integrate data to arrive at the final processed sets of data, ready for analysis.

##### 4.1. Data Preparation

First, we created one geodatabase where all related data will reside with the WGS 1984 map projection, suitable for the study site in Georgia, United States. The geodatabase was then fed with all the aforementioned data and more general data, including (1) the imported data files from the EPRI's data repository, (2) Georgia's Topologically Integrated Geographic Encoding and Referencing (TIGER) road shapefile, (3) 2010 Georgia's county shapefile, (4) NOAA's storm and storm detail maps from 2013 to 2015, (5) 48 unzipped weather shapefiles, monthly temperatures, and precipitation data from GaSDI and the Georgia GIS Clearinghouse (four total files showing the maximum, minimum, and average temperature and the precipitation for each month of the year).

Second, we perused the outage data, and realizing that about 5% of the data (3992 records out of 80,839 records) were not spatially-enabled, we excluded them from the analysis. The final outage data has 76,848 records. Then, for each record, we create dummy variables to indicate the type of outage. Overall, there are four major types of outages have been identified: right of way, weather, equipment failure, and system overload, as shown in Table 1. With each of the outage events, we iteratively separate each outage into its own map layer. To associate the outages events, we utilize the average nearest-neighbor to identify the likelihood of data-forming clusters throughout the study sites.

**Table 1.** Four types of outages and their associate events.

Outage Type	Outage Events
Right of Way	Wind/Tree, Limb online, Tree Fell on Line, Tree Grew into Line, Vines
Weather	Wind, Ice, Major Storm, Lightning
Equipment	Failed in Service, Deterioration
System Overload	Thermal Overload, Overload, Load Shed

Third, to prepare the data, we convert the date time field in the NOAA's storm and storm event datasets into the day of the year so that the date component in all the data is synchronized.

Fourth, after setting up the data, we create data processing steps via the ArcMap ModelBuilder tool so that the steps could systematically go through all weather data layers within the geodatabase. We separate the steps into three models. The result of the models is an outage table with an additional

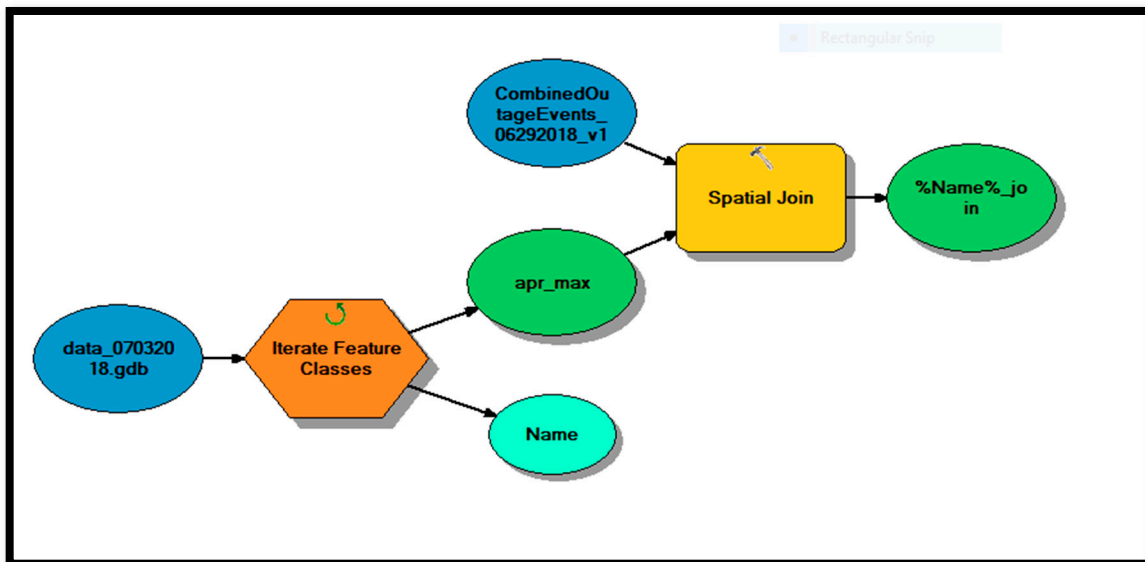
48 columns (four columns for each month displaying the maximum, mean, and minimum temperature, and another field for the precipitation for each outage event location).

#### 4.1.1. Interpreting a Model in ModelBuilder

First, ModelBuilder is a graphical workflow that helps to streamline all geoprocessing steps. All input data, geoprocessing tools, intermediary resulting data, and output data are displayed with specific colors and shapes. The blue oval in the ModelBuilder depicts input data and the green oval depicts resultant data. The teal oval depicts resultant value, the rectangle shows a specific processing tool, and the hexagon depicts an iterator, which is used to go through a specific list of items within a repository. The arrows are used to connect each component of a model within ModelBuilder. As a convention, the model workflow starts from left to right.

#### 4.1.2. Model 1

Model 1, depicted in Figure 5, shows that the workflow starts from selecting each weather feature iteratively, and spatially joins with the combined outage data file. This results in a new feature that contains data from each weather file and the outage event file.



**Figure 5.** ArcMap ModelBuilder (Model 1) to spatially join the outage events with the weather data.

#### 4.1.3. Model 2

The results of the Model 1 are 48 files corresponding to each weather dataset combined with the outage events. Therefore, Model 2 is used to rename the output field appropriately to reflect the month and type of weather data (whether max, min, mean temperature, or average precipitation). The process of Model 2 is displayed in Figure 6.

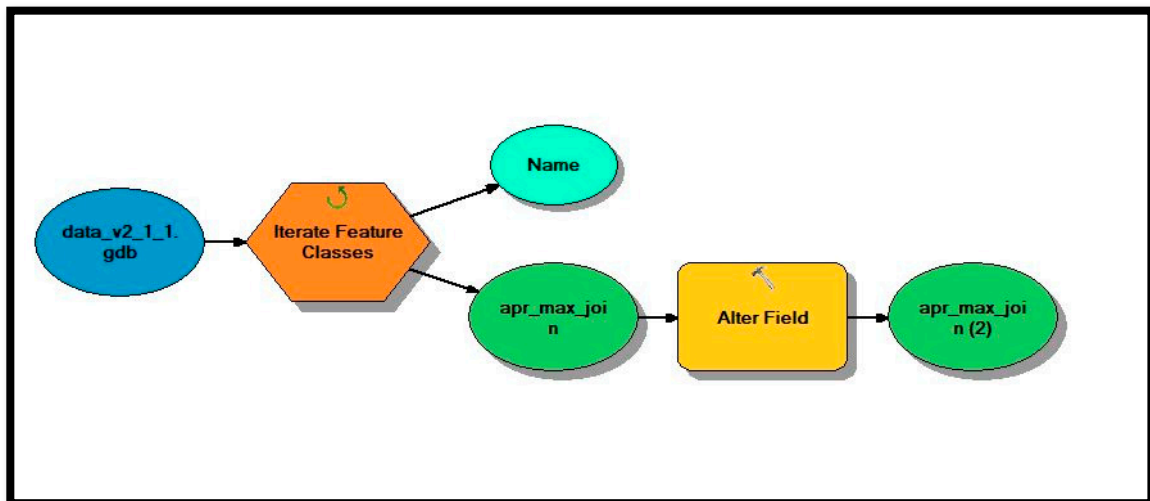


Figure 6. ArcMap ModelBuilder (Model 2) to rename the output field (contour field) from Model 1.

#### 4.1.4. Model 3

The last model, Model 3 (Figure 7), made sure that the processed data from the previous two models contained in one unified feature layer. For each of the 48 joined features, the model selects the appropriate join field and iteratively joins all data into the outage event layer.

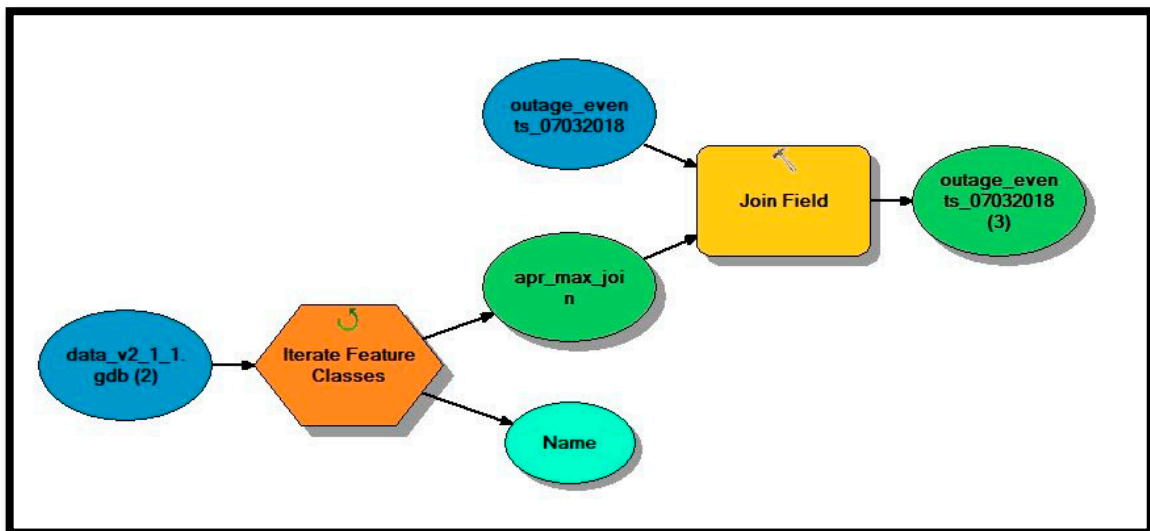


Figure 7. ArcMap ModelBuilder (Model 3) to join outage-events data with the 48 fields of weather data.

Fifth, after the data is processed through the ModelBuilder, we continue to enrich the data by creating four additional columns to the outage map attribute table to show the weather data for each outage event, considering the month of the year. For each outage event, we showed data for the maximum, mean, and minimum temperature, and precipitation. We followed this step by joining three additional data sources to the combined layer, namely the storm event and storm event details via the data transformation in the processing steps. Additionally, we added the following forestry data, showing how EPRI has maintained its infrastructure: “Forestry Expected Pruning Man Hours,” “Average Standard Tree Pruning Miles with Bucket,” “Average Mechanical Tree Pruning Miles,” “Average Climbing Tree Pruning Miles,” and “Actual Pruning Man Hours/Circuit Mileage.” Finally, we derived several equipment data, specifically transformer’s age and pole’s age to show how long a transformer and pole last since its first installation to its demise, respectively.



#### 4.2. Analysis Framework

After making sure that data are properly processed, we start the analysis process. The analysis consists of two types: non-spatial and spatial. First, the non-spatial analysis follows the traditional exploratory and confirmatory analysis. For instance, we explore the statistical relationships within the data using descriptive statistics and correlation analysis. Within this step, other analysis could be included, namely factor analysis or exploratory data analysis. This capability does not exist in the ArcGIS platform, so we opt to use SPSS for this task. In fact, any statistical tool would suffice. This step essentially provides an initial understanding of the data and how each component of the data could relate to one another in a non-spatial way.

Subsequently, a spatial analysis is conducted, showing how the outage could be analyzed through location analytics. In this instance, we select two analysis methods: hotspot analysis to show an initial analysis through space and emerging hotspot analysis to show how outage events are related to one another through space and time.

### 5. Results and Findings

#### 5.1. Initial Exploration

Initial data exploration indicated inadequate data for analysis and many null fields, as shown in Table 2. For example, the asset management folder showed inspection data for only two types of equipment. As for the age of asset data embedded in the GIS maps, analyses showed that “last date installed” and “original date installed” fields for equipment, but these were mostly null values. For instance, of the 4600 records for switches, only 106 records showed original date installed, which is equivalent to 2% of the total records. Thus, we decided to use pole-age data as a proxy for the rest of the equipment data in the analysis.

Not all files in the data set appeared useful considering the scope of this project work. For example, the Jets data file is about the field jobs. Another example is the circuit “load” data that do not include longitude/latitude data or any other georeferencing method to bring into ArcGIS. “load” data appeared to be overall feeder data and because distribution feeders are highly branched, data are not useful to draw conclusions on which branch is loaded and which is not. As for other data collections, such as SCADA data files, research indicated that it is only helpful if we need to dig into one of the operations that caused outage events.

**Table 2.** Power-outage cause category.

Outage Type	Percent of Outage Events Duration	Percent of Outage Events Count
Equipment	10.19%	6.03%
Right of Way	12.35%	7.02%
System Overload	1.00%	0.52%
Weather	16.97%	4.57%
<b>Total</b>	<b>40.51%</b>	<b>18.13%</b>

#### 5.2. Descriptive Statistics

In the final resultant dataset, the data contains the duration of the outage event, the number of notifications from the customer, the mean temperature, the precipitation, several forestry data attributes concerning tree pruning and maintenance, and age of the equipment, namely transformers and poles. The descriptive statistics in Table 3 show that there is a wide range of values between each data point. For instance, with the outage event customer calls, the mean is at 11.19, but the standard deviation is 85.07, with the max of 4888. This shows that, while there are many low call volumes, there are several instances whereby the notifications are numerous, thus skewing the data. It is also interesting to note that the equipment could fail as early as three years, and the transformers last up to eight years while poles can sustain for 93 years until succumbing to outage events.

**Table 3.** Descriptive statistics of study variables.

Variable.	M	SD	Min	Max
Outage event duration	89.15	204.81	0	3589
Outage event customer calls	11.18	85.07	0	4888
Temperature (mean)	62.52	13.72	40.25	80.75
Precipitation	4.29	0.66	2.8	5.80
Forestry expected pruning man hours	858.01	882.24	0	3300
Average standard tree pruning miles with bucket	6.63	6.48	0	20.49
Average mechanical tree pruning miles	3.00	2.94	0	9.28
Average climbing tree pruning miles	0.75	0.73	0	2.32
Actual pruning man hours/circuit mile	42.18	36.80	0	157
Transformer age	4.50	1.86	3	8
Pole age	23.90	16.76	3	93

Note:  $n = 76,847$ .

### 5.3. Correlation Results

To further explore the relationships between the variables, we ran the correlation analysis. The result is shown in Table 4. The correlation matrix shows that the duration of an outage event is significantly correlated with most of the variables in the dataset except whether the event is part of the forestry management route. Outage event duration is negatively correlated with temperature. While the relationship is significant, it is not strong. As for the number of calls from the customer, a somewhat opposite result could be observed. There are only four variables that are statistically significant with outage event customer calls: forestry management, actual pruning staff hours/circuit mile, transformer age, and pole age. In addition, the relationships are trivial.

It is interesting to observe that pole age is statistically significant with all other variables, indicating the importance of the age of the pole when it comes to perusing outages data. However, the relationship is minimally correlated. Meanwhile, transformer age is also statistically significant with all variables except temperature. Intriguingly, there is no relationship between transformer age and temperature. It is also expected to see that the forestry-related variables are highly positively correlated, with a high statistical significance power.

Overall, the correlation matrix shows that there are notable relationships between the variables. However, despite their statistical significance, the correlation relationships are marginal. Therefore, this creates a need to analyze data spatially. In the next section, we present the spatial analysis framework, and then perform several analyses to highlight the importance of incorporating a spatial component into any analysis task.

Table 4. Correlation results.

Variable	Outage Event Duration	Outage Event Customer Calls	Storm Event (1 = yes)	Temperature (mean)	Precipitation	Forestry Management (1 = yes)	Forestry Expected Pruning Staff Hours	Average Standard Tree-Pruning Miles with Bucket	Average Mechanical Tree-Pruning Miles	Average Climbing Tree-Pruning Miles	Actual Pruning Staff Hours/Circuit Mile	Transformer Age	Pole Age
Outage event duration	1.00												
Outage event customer calls	0.09 **	1.00											
Storm event (1 = yes)	0.08 **	0.01	1.00										
Temperature (mean)	−0.13 **	0.01	0.14 **	1.00									
Precipitation	0.08 **	0.00	−0.06 **	−0.37 **	1.00								
Forestry management (1 = yes)	0.01	−0.02 **	−0.01	0.00	−0.03 **	1.00							
Forestry expected pruning staff hours	0.01 **	0.00	0.00	0.00	−0.02 **	0.77 **	1.00						
Average standard tree-pruning miles with bucket	0.01 **	0.00	0.00	0.00	−0.02 **	0.81 **	0.96 **	1.00					
Average mechanical tree-pruning miles	0.01 **	0.00	0.00	0.00	−0.02 **	0.81 **	0.96 **	1.00	1.00				
Average climbing tree-pruning miles	0.01 **	0.00	0.00	0.00	−0.02 **	0.81 **	0.96 **	1.00 **	1.00	1.00			
Actual pruning staff hours/circuit mile	0.01 *	−0.01 **	−0.01	−0.01 **	−0.02 **	0.91 **	0.85 **	0.79 **	0.79 **	0.79 **	1.00		
Transformer age	0.01 *	0.01 **	0.01 **	0.00	0.02 **	−0.13 **	−0.11 **	−0.12 **	−0.12 **	−0.12 **	−0.11 **	1.00	
Pole age	0.02 **	−0.01 **	0.02 **	−0.01 **	0.03 **	0.02 **	0.04 **	0.03 **	0.03 **	0.03 **	0.03 **	−0.03 **	1.00

NOTE: \*  $p < 0.05$ ; \*\*  $p < 0.01$ , two-tailed tests.

5.4. Spatial Pattern Analysis in ArcGIS

Based on ArcGIS average nearest neighbor analyses reports displayed in Figure 8, the observed mean distance is the largest between system-overload outage events (727 m) compared to weather-related events (171 m), equipment failure (207 m), and right-of-way outage events (210 m). Weather-related outage showed the shortest mean distance between events. A clustered pattern appeared for all four outage-event types. Additionally, based on the analysis results, there is less than 1% likelihood that this clustered pattern could be the result of random chance, indicating statistical significance.

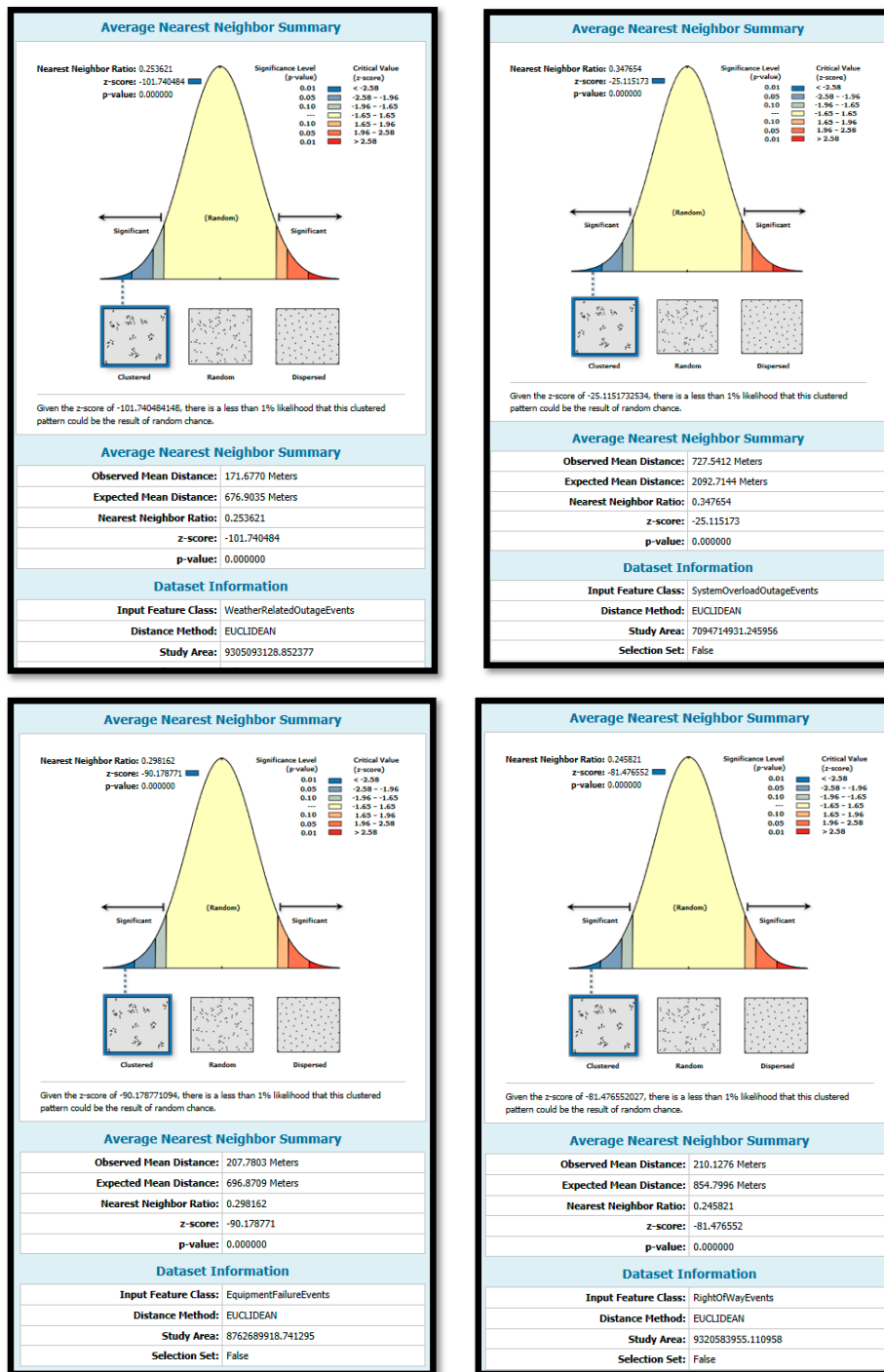


Figure 8. ArcGIS average nearest neighbor analyses results for four types of outages: weather (top left), system overload (top right), equipment failure (bottom left), and right of way (bottom right).

### 5.4.1. Spatial Analysis Framework

We developed the following framework (Figure 9) to guide the investigation and illustrate the various levels of analysis. At the first level, all features in the outage events are analyzed to provide a general sense of where in space the clusters of outages might be. As we get more refined in the granularity of the data, Level 2 indicates a more specific group input features, namely types of outage events. In this case, there are four: equipment failure, weather, right of way, and system overload. As the analysis drills down to finer details, Level 3 approaches each individual feature in each type and in turn, provides the analysis for a particular feature. Using this framework, the subsequent sections provide the implementation and the results of each level. Specifically, we use optimized hot spot analysis for all levels and emerging hot spot analysis for Level 1 and Level 2. These are meant to demonstrate how such a framework could apply to other similar scenarios.

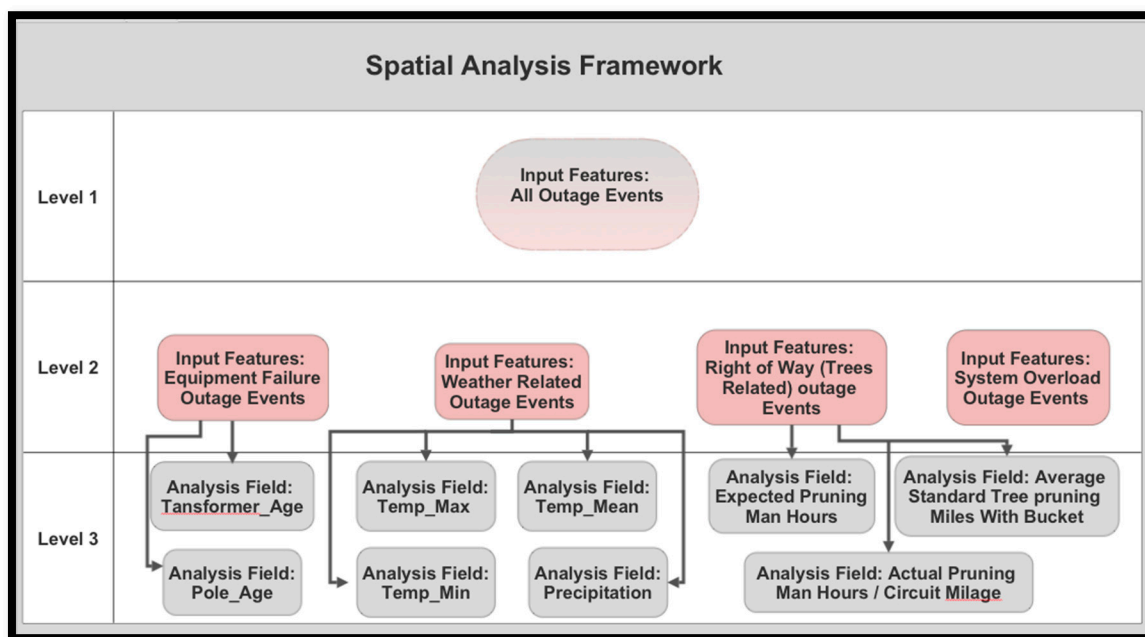


Figure 9. Spatial Analysis Framework.

### 5.4.2. Level 1 Spatial Analysis—Optimized Hot Spot Analysis

Based on Level 1 of the spatial analysis framework, we used ArcGIS optimized hot spot analysis tool to generate a map (Figure 10) of statistically noteworthy hot and cold spots using the Getis-Ord  $G_i^*$  statistic. Since we did not identify an analysis field, this tool assessed the characteristics of the input feature class (power outage events) to produce optimal results [27]. The tool showed one large area of hot spots in Clayton and Fulton County where power outage was statistically significant due to multiple causes. As for cold spots, they appeared in counties such as Coweta, mid and South Fayette, Butts, North Meriwether, and a majority of Henry County.

With a polygon cell size of 1319 m, there are 1296 weighted polygons on the study site. For each of the cells, there are an average of 59.29 incident counts, with the standard deviation of 81.23. The minimum and maximum of incidents counts are 1 and 598, respectively. Correspondingly, the optimal fixed-distance band based on the average distance to 30 nearest neighbors is 5738 m. There are 984 statistically significant output features, based on a false-discovery-rate correction for multiple testing and spatial dependence. Additionally, only 0.5% of features had fewer than eight neighbors.

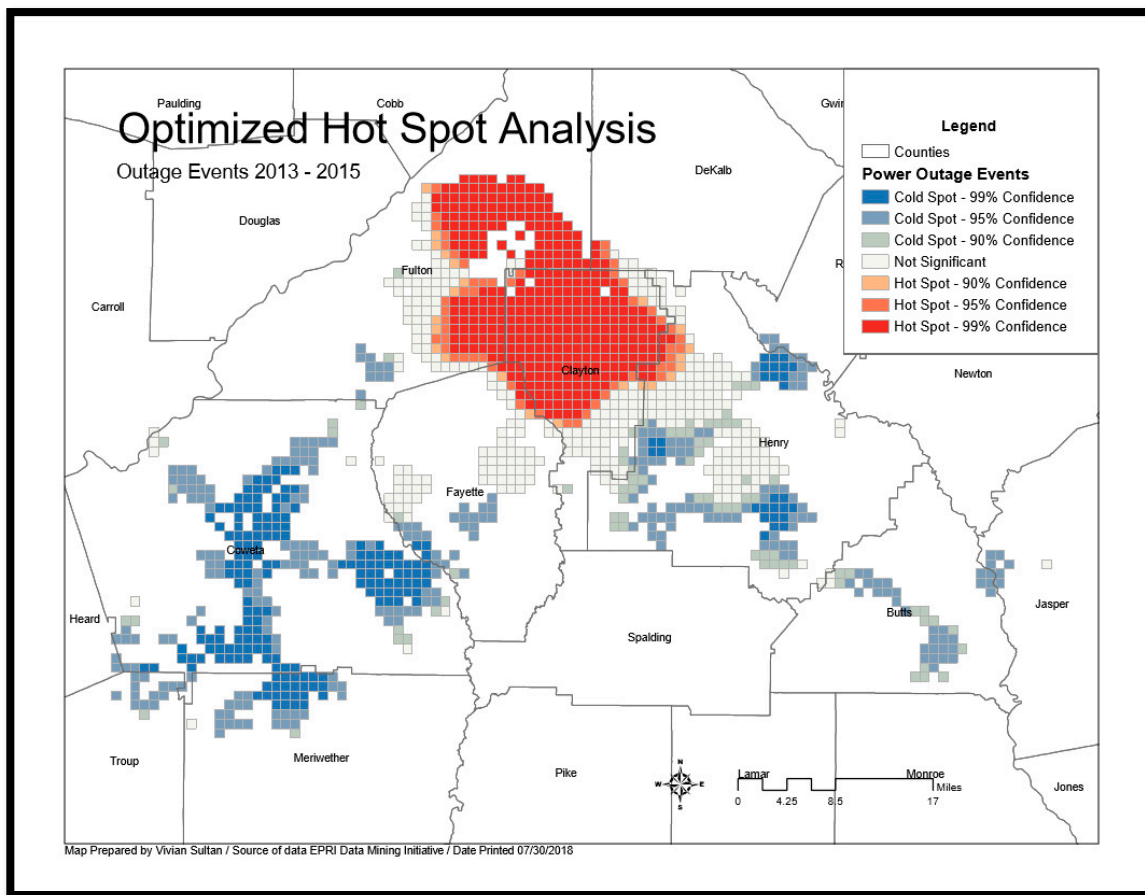


Figure 10. ArcGIS optimized hot-spot analysis Level 1 output map.

#### 5.4.3. Level 1 Spatial Analysis—Emerging Hot Spot Analysis

In addition to optimized hot spot analysis, emerging hot spot analysis is another candidate of the type of spatial analysis that could be run. Emerging hot spot analysis is similar to optimized hot spot analysis, with an added dimension of time. As a result, hot and cold spot categories expand to new types, including new, consecutive, intensifying, persistent, diminishing, sporadic, oscillating, and historical. The categories are defined as follows:

- New: the most recent time step interval is hot/cold for the first time
- Consecutive: a single uninterrupted run of hot/cold time-step intervals, comprised of less than 90% of all intervals
- Intensifying: at least 90% of the time-step intervals are hot/cold, and becoming hotter/colder over time
- Persistent: at least 90% of the time-step intervals are hot/cold, with no trend up or down
- Diminishing: at least 90% of the time-step intervals are hot/cold and becoming less hot/cold over time
- Sporadic: less than 90% of the time-step intervals are hot/cold
- Oscillating: the most recent time step interval is hot/cold, less than 90% of the time-step intervals are hot/cold and it has a history of reverse from hot to cold and vice versa.
- Historical: at least 90% of the time-step intervals are hot/cold, but the most recent time-step interval is not

The following section recaps ArcGIS emerging hot spot analysis results, and the interpretation of the results.

In Figure 11, the space–time cube aggregated 76,847 points into 6000 fishnet grid locations over 32 time-step intervals. Each location is 1319 m by 1319 m square. The entire space–time cube spans an area 131,900 m west to east and 79,140 m north to south. Each time-step interval is 1 month so the entire time period covered by the space–time cube is 32 months. Of the 6000 total locations, 1296 (21.60%) contain at least one point for at least one time-step interval. These 1296 locations comprise 41,472 space–time bins of which 18,973 (45.75%) have point counts greater than zero. No statistically significant increase or decrease emerged in point counts over time. The summary of results is displayed in Table 5.

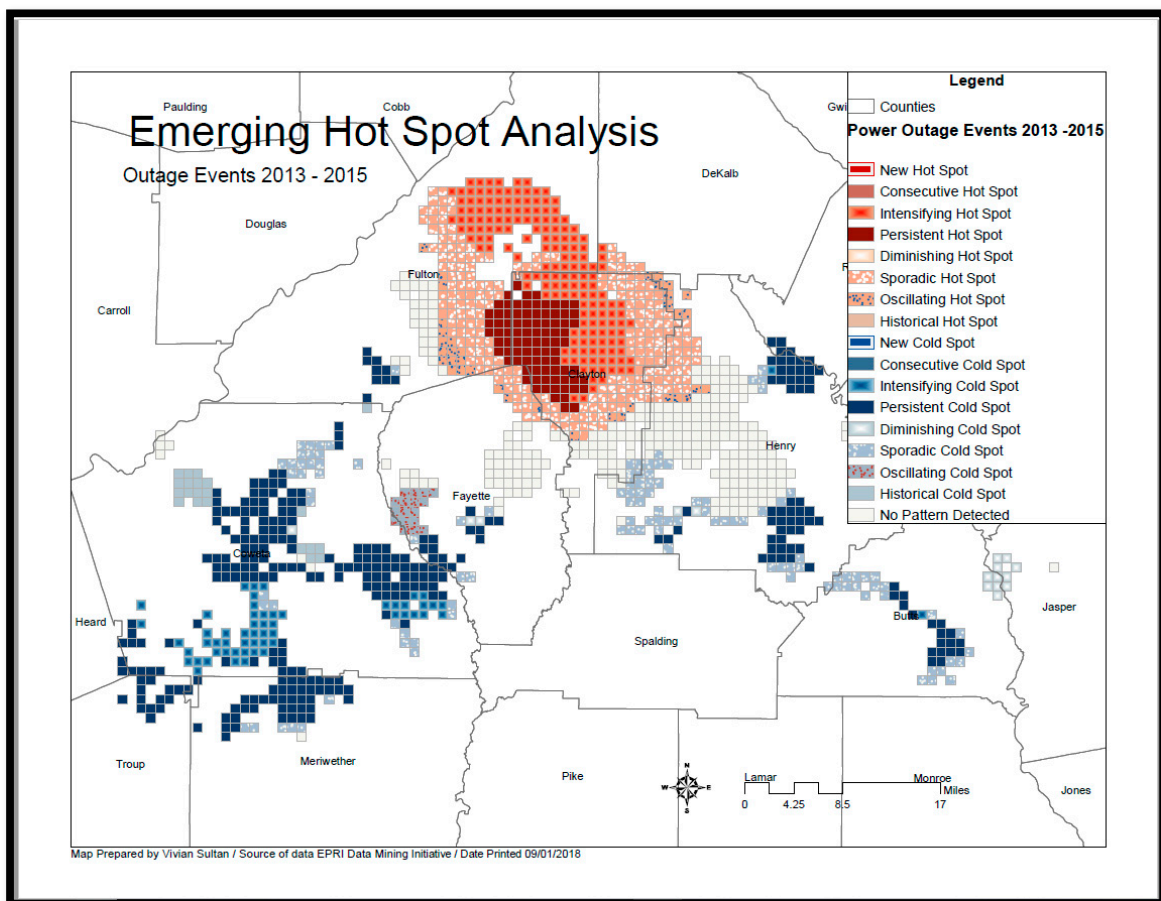


Figure 11. ArcGIS Emerging Hot Spot Analysis Level 1.

Overall, the emerging hot spot analysis elucidates a more drill-down look at the cold and hot spots observed in the optimized hot spot analysis. With the addition of time, another dimension, we can observe that there are many areas are persistent hot and cold spots, indicating where chronic issues and stability are respectively. Persistent hot spots cluster in the north west of Clayton county while persistent cold spots can be found in Henry, Butte, Merriweather, Troup, Fayette, and Coweta counties. These spots indicate that, through 90% of the time, the locations are identified as hot and cold spots consistently. Furthermore, there are some intensifying hot spots near north of Clayton and Fulton county, signifying a need for further investigation. In the west of Fayette county, we can find some oscillating cold spots, the locations where they used to be hot spots but have now transitioned to cold spots. These are fascinating because it could be studied further to see what has supported the improvement of reducing outage events. Such understandings could be used to improve other locations, especially those which are classified as persistent hot spots and intensifying hot spots.

**Table 5.** Analysis summary of emerging hot spot level 1 results.

	HOT	COLD
New	0	0
Consecutive	0	0
Intensifying	169	57
Persistent	86	298
Diminishing	0	16
Sporadic	227	113
Oscillating	37	17
Historical	0	21

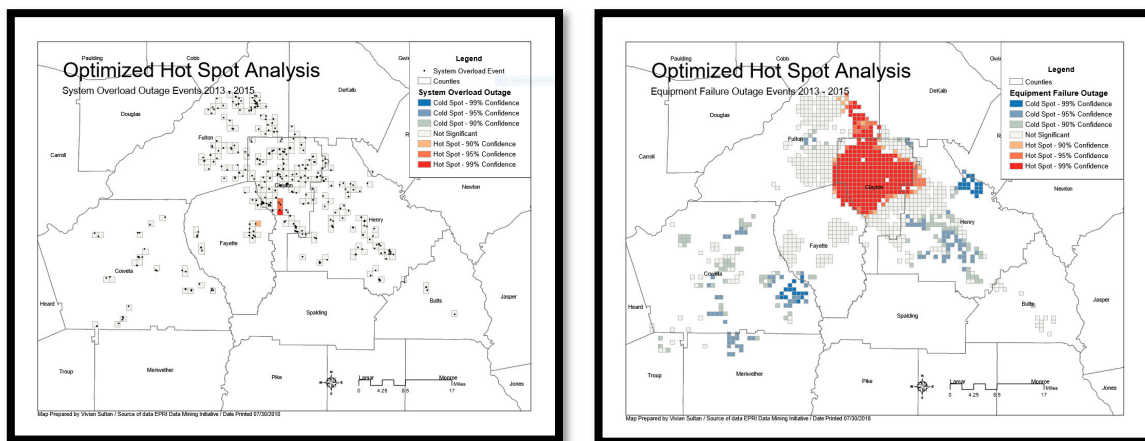
5.4.4. Level 2 Spatial Analysis—Optimized Hot Spot Analysis

Based on Level 2 of the spatial analysis framework, we generated an additional four map layers (Figure 12) using the optimized hot spot analysis tool by consecutively selecting the input feature classes “System overload power outage events,” “equipment failure outage events,” “weather related outage events,” and “right of way outage events.”

The map on the top left of Figure 12 indicated the outages based on system overload. There is only one small region in the south of Clayton county that is identified as a hot spot. The relationships of all other spots are insignificant, signifying a uniformity in outage events based on system issues. As a result, this hot spot should investigate further and making sure to bring down the high level of outages in this region due to system overloading.

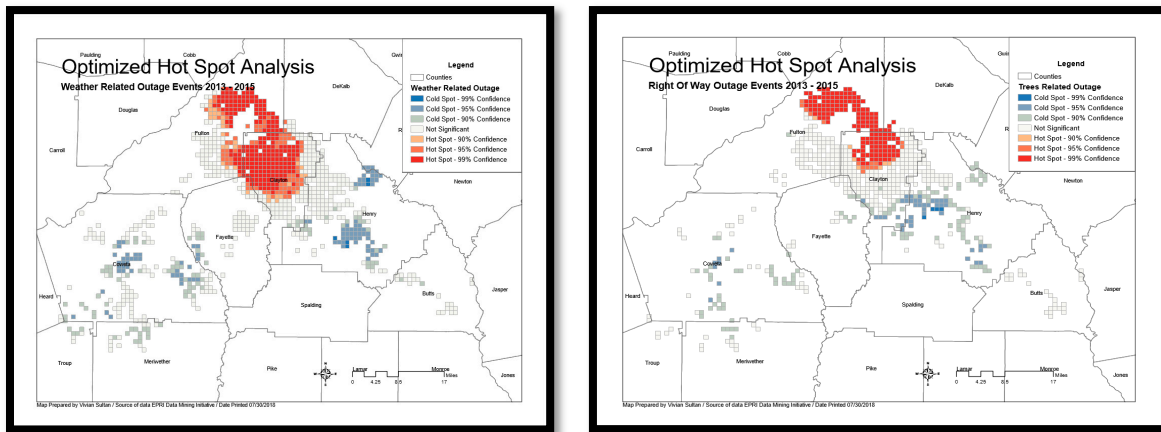
The map on the top right of Figure 12 provided the result of an optimized hot spot analysis for equipment failure. As we can see, there is a large hot spot resided in Clayton county, located in the middle to the north of the county, and extended further north into Fulton county. It seems like the equipment used in this region would fail faster than the rest of the state. Therefore, more resources should be placed to routinely check and maintain a high level of operability and redundancy to make sure that the equipment failure is brought down. Furthermore, there are several cold spots observed in this map. More specifically, Henry county has a cold spot in the north, and sporadic cold spots around middle and south of the county. Additionally, Coweta county has a cluster of cold spots in the southeast while Meriwether county has a cold spot in the north. These locations provide another good location for investigation. These investigations could reveal best practices and standards that could be used to improve the equipment failure event in the hot spot locations.

Coincidentally, the optimized hot spot analysis for weather (bottom left of Figure 12) is almost identical with the one for equipment failure. This finding should be investigated further, and authorities could formulate a plan to battle weather related issues while also addressing equipment failure ones.



**Figure 12.** Cont.

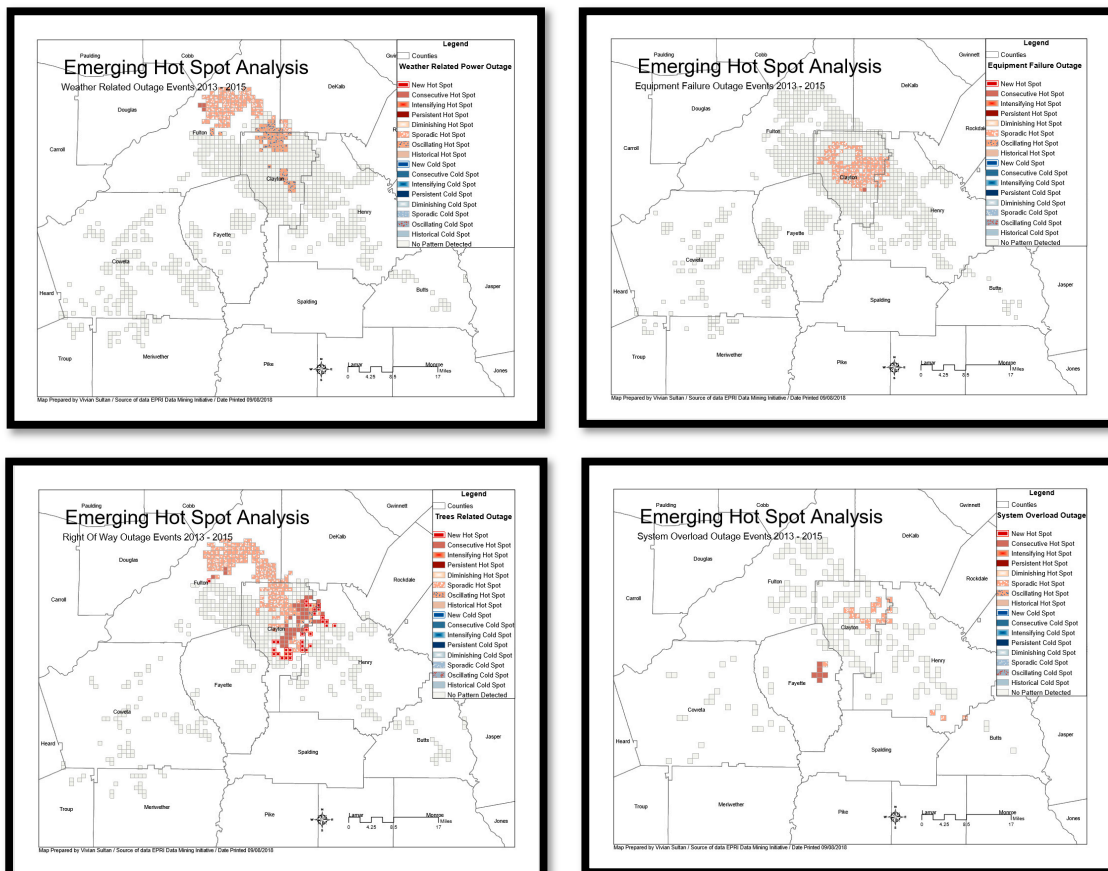




**Figure 12.** Spatial Analysis Level 2—Optimized Hot Spot Analysis output maps per type of outage causes. **Top left:** system overload; **top right:** equipment failure, **bottom left:** weather; **bottom right:** right of way.

### 5.4.5. Level 2 Spatial Analysis—Emerging Hot Spot Analysis

Based on Level 2 of the spatial analysis framework, we generated an additional four map layers (Figure 13) using the emerging hot spot analysis tool by consecutively selecting the input feature classes “system overload power outage events,” “equipment failure outage events,” “weather related outage events,” and “right of way (trees related) outage events.”



**Figure 13.** Spatial Analysis Level 2—Emerging Hot Spot Analysis output maps per type of outage13causes. **Top left:** weather; **top right:** equipment failure, **bottom left:** right of way; **bottom right:** system overload.

Analysis results, illustrated in Figure 13, show that the right of way (trees-related) outages had the highest number of locations with hot trends (259 total count of locations) compared to weather-related outages (160 locations), equipment-failure outage (129 locations), and system overload (27 count of locations with hot trends). Thus, a utility company can use this intelligence to reduce the risk of power outages and plan accordingly.

In this instance, we have identified trees/forestry that need pruning to be the leading cause of outages and we identified 259 locations with hot trends. These 259 locations include the 40 consecutive locations with a single uninterrupted run of statistically significant hot spots.

The utility company can use this information to reduce the risk of wildfire and keep customers safe. The electric utility would accelerate its vegetation-management work and prioritize tree-pruning fieldwork to tackle these 40 consecutive locations first. Also, considering the availability of weather forecasts, this analysis can help a utility firm to prepare, should a storm be anticipated. Priorities would be given to staging equipment and restoration workers at those 160 locations with weather-related hot trends in the event of a storm.

In future work, a solution will be developed using Insights for ArcGIS to demonstrate how a utility company can prioritize locations that need inspection or infrastructure work and detect regions where new components such as distribution switches may provide net benefits. Through Insights for ArcGIS, a utility analyst has the capability to work with interactive maps and charts at the same time. The objective of the next step is to provide an instantiation of the GIS model and the spatial analysis framework developed in this study.

#### 5.4.6. Level 3 Spatial Analysis

The  $G_i^*$  statistic was also designed for an Analysis Field with a variety of values. The statistic is not appropriate for binary data. Therefore, we used the tool to check the analysis field to ensure the values had at least some variation [27].

As we are interested in analyzing weather-related outage events associated with the temperature, we identified “temperature” for the analysis field. Based on Level 3 of the spatial analysis framework, we generated four additional map layers (Figure 14) by consecutively selecting “maximum temperature,” “mean temperature,” “minimum temperature,” and “precipitation” for the analysis field.

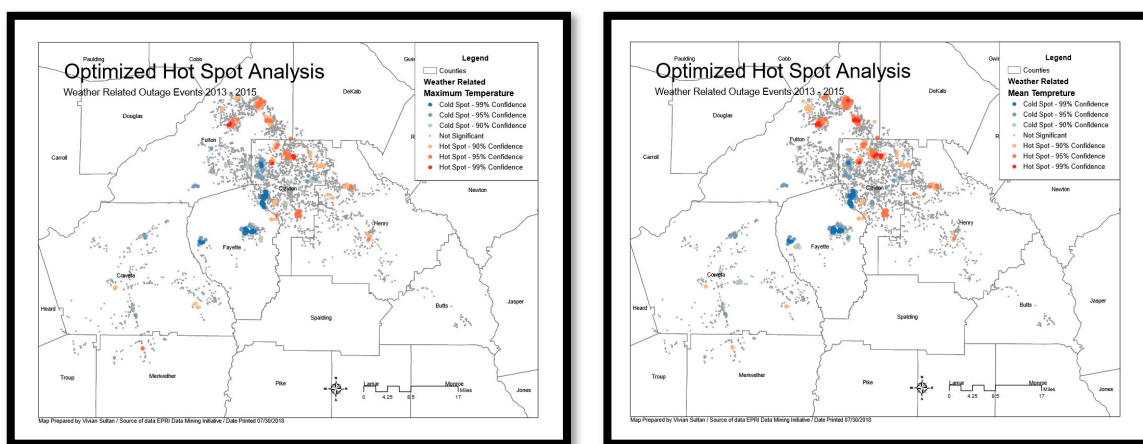
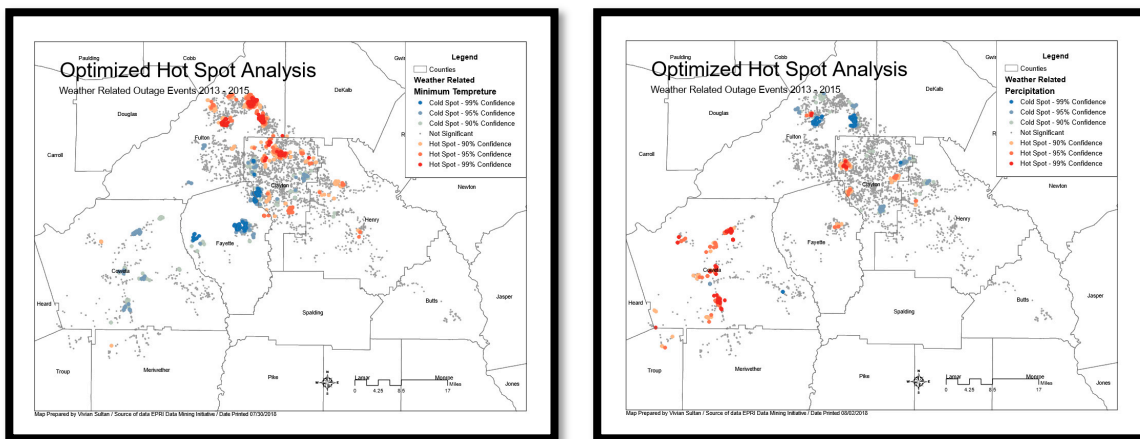
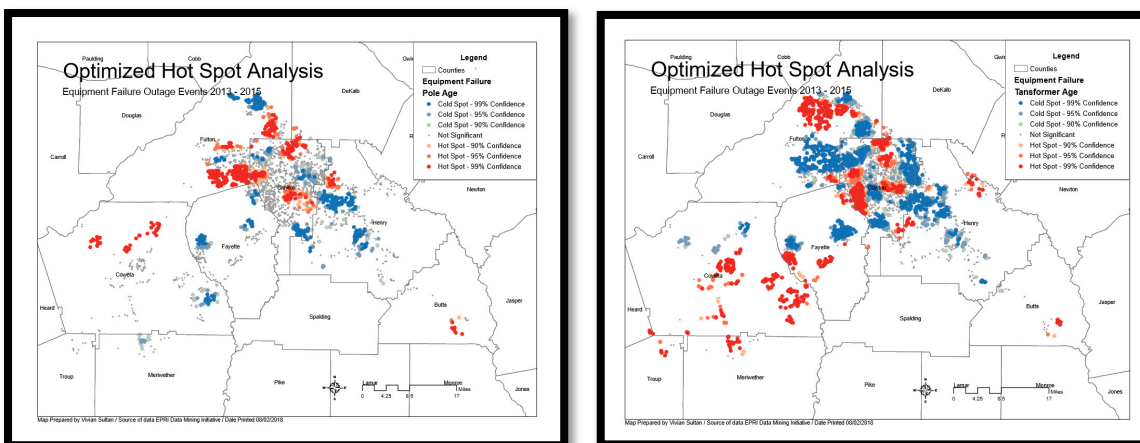


Figure 14. Cont.



**Figure 14.** Spatial Analysis Level 3—Optimized Hot Spot Analysis output maps for each weather type outage events. **Top left:** maximum temperature; **top right:** mean temperature, **bottom left:** minimum temperature; **bottom right:** precipitation.

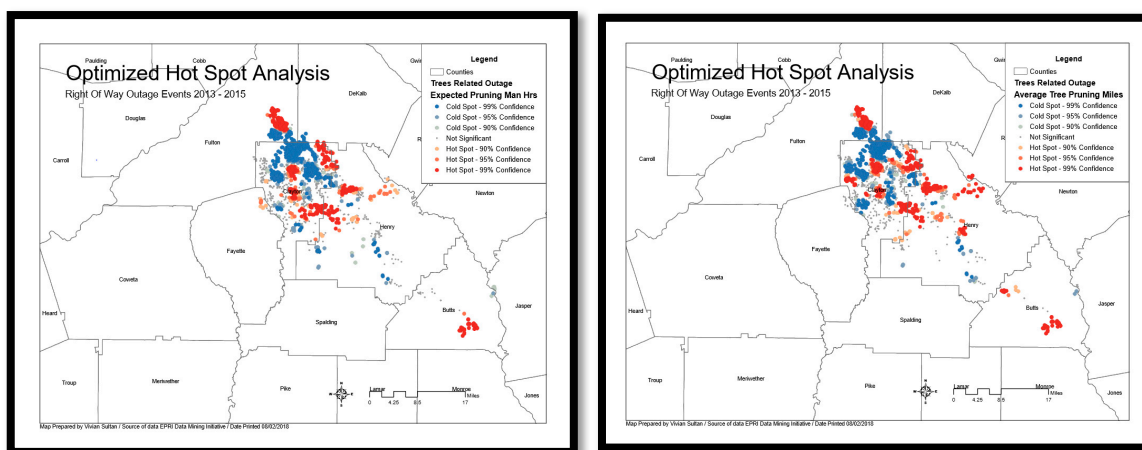
Additionally, we generated two map layers (Figure 15) by submitting an equipment failure outage-events map layer in the input feature and consecutively selecting “transformer age” and “pole age” for the analysis field in the optimized hot spot analysis tool. The purpose of this step was to examine equipment failure outage events that align with the infrastructure age.



**Figure 15.** Spatial Analysis Level 3—Optimized Hot Spot Analysis output maps for each equipment failure type outage events. **Left:** pole age; **right:** transformer age.

Based on Level 3 of the spatial analysis framework, we generated two supplementary map layers (Figure 16) using the Optimized Hot Spot Analysis tool by selecting Right Of Way (trees related) outage events in the input feature and consecutively selecting “expected pruning staff hours,” “average standard tree pruning miles with bucket,” and “actual pruning staff hours/circuit mileage” for the analysis field.

As illustrated by the maps in Figure 16, the variables of average standard tree-pruning miles with bucket and forestry expected pruning staff hours nearly perfectly correlated with each other. Further analysis in SPSS indicated that the variables average climbing tree-pruning miles, average standard tree-pruning miles with bucket, average mechanical tree-pruning miles, and forestry expected pruning staff hours were either perfectly correlated ( $r = 1.00$ ) or nearly perfectly correlated ( $r > 0.90$ ) with each other.



**Figure 16.** Spatial Analysis Level 3—Optimized Hot Spot Analysis output maps for right of way outage events. **left:** Expected Pruning Staff Hours; **right:** Average Standard Tree Pruning Miles with Bucket; **bottom:** Actual Pruning Staff Hours/Circuit Mileage.

### 6. Conclusions

This study aimed to answer: How may spatial analytics enhance our understanding of power outages? To answer the research question, we developed a spatial analysis framework that can be used effectively in the utility industry to investigate power-failure events and their causes. Analysis revealed areas where power outage was statistically significant due to multiple causes.

The GIS model presented in this study can help advance Smart Grid reliability through, for example, elucidating a root cause of power failure, defining a solution for a blackout through data, or implementing a solution with continuous monitoring and management. In this study, we unveiled the novel use of location analytics to enhance understanding of power outages.

One limitation of this research is that we used poles-age data as a proxy for infrastructure age and the rest of the equipment data. Research showed many null fields, missing, and incomplete data in the electric-utility-system database.

Future research should include analysis in ArcGIS Pro, which is Esri’s next-gen desktop GIS product that provides professional 2D and 3D mapping and added tools to advance visualization, analytics, and imaging. Also, ArcGIS GeoEvent Server is another tool to accommodate the multiple streams of data flowing continuously through filters and processing steps that one may define. Thus, identifying failures in the network by performing real-time analytics on streams of data can become feasible. Future work may involve connecting to virtually any type of streaming data feed and transforming GIS applications into frontline decision apps, showing power-outage incidents as they occur.

From this research, we conclude that GIS offers a solution to analyze the electric-grid distribution system. Our model provides evidence that GIS can perform the analysis to investigate power-failure events and their causes. If additional funds and data are made available, we can expand this analysis, build on ArcMap source code, and create a custom solution for the utility industry to control and forecast power outages. GIS can be a major resource to assist electronic inspection systems, lower the duration of customer outages, improve crew response time, and reduce labor and overtime costs.

**Author Contributions:** Vivian Sultan completed the research, developed the solution, and wrote the paper. Brian Hilton was the chair of Vivian Sultan’s dissertation committee, and designed the paper structure, reviewed the findings, and proofread/edited the final paper. All authors have read and agree to the published version of the manuscript.

**Funding:** This research is based on the first author’s doctoral dissertation earned from Claremont Graduate University. No additional funding was used for this project.

**Acknowledgments:** The first author thanks her entire dissertation committee for their helpful advice. The first author also thanks Jim Horstman, IT Utility Consultant, and Jared Green, Project Manager at EPRI, who guided and supported the first author in this research initiative.

**Conflicts of Interest:** The authors declare that they have no competing interests.

## References

1. LaCommare, K.; Eto, J. *Understanding the Cost of Power Interruptions to U.S. Electricity Consumers*; Lawrence Berkeley National Laboratory: Berkeley, CA, USA, 2004.
2. Eto, J. *The National Cost of Power Interruptions to Electricity Consumers—Revised Update*; Lawrence Berkeley National Laboratory: Berkeley, CA, USA, 2017.
3. President’s Council of Economic Advisers and the U.S. Department of Energy’s Office of Electricity Delivery and Energy Reliability, with Assistance from the White House Office of Science and Technology. Economic Benefits of Increasing Electric Grid Resiliency to Weather Outages. 2013. Available online: [http://energy.gov/sites/prod/files/2013/08/f2/GridResiliencyReport\\_FINAL.pdf](http://energy.gov/sites/prod/files/2013/08/f2/GridResiliencyReport_FINAL.pdf) (accessed on 19 November 2015).
4. Porter, J. The \$306 Billion Question: How to Make Outage Management Better? 12 June 2018. Available online: <https://www.power-grid.com/2018/06/12/the-306-billion-question-how-to-make-outage-management-better/#gref> (accessed on 13 August 2019).
5. American Society of Civil Engineers. 2009 Infrastructure Fact Sheet. 2009. Available online: [http://www.infrastructurereportcard.org/2009/sites/default/files/RC2009\\_rail.pdf](http://www.infrastructurereportcard.org/2009/sites/default/files/RC2009_rail.pdf) (accessed on 7 October 2015).
6. Stone, D. It’s the Electric Grid, Stupid. 9 September 2011. Available online: <http://www.thedailybeast.com/articles/2011/09/09/major-power-outage-shows-weakness-of-aging-electric-infrastructure.html> (accessed on 5 October 2015).
7. Chayanam, K. Analysis of Telecommunications Power Outages—Inline. Director of Thesis: Andrew P. Snow. June 2005. Available online: [https://etd.ohiolink.edu/rws\\_etd/document/get/ohiou1125024491/inline](https://etd.ohiolink.edu/rws_etd/document/get/ohiou1125024491/inline) (accessed on 15 September 2017).
8. Westar Energy. What Causes Power Outages? Working to Improve Service Reliability. 2017. Available online: <https://www.westarenergy.com/outage-causes> (accessed on 11 September 2017).
9. Diesel Service and Supply. Causes of Power Failures & Power Outages | Diesel Service. 2017. Available online: [http://www.dieselserviceandsupply.com/Causes\\_of\\_Power\\_Failures.aspx](http://www.dieselserviceandsupply.com/Causes_of_Power_Failures.aspx) (accessed on 11 September 2017).
10. Rocky Mountain Power. Key Causes of Power Outages. 2017. Available online: <https://www.rockymountainpower.net/ed/po/or/kcopo.html> (accessed on 11 September 2017).
11. Miller, Rich How to Prevent Downtime Due to Human Error | Data Center Knowledge. August 2010. Available online: <https://www.datacenterknowledge.com/archives/2010/08/13/how-to-prevent-downtime-due-to-human-error> (accessed on 11 September 2017).
12. Freedman, A. Weather and Climate Disasters Cost the U.S. a Record \$306 Billion in 2017. 8 January 2018. Available online: [https://mashable.com/2018/01/08/2017-record-year-billion-dollar-disasters-third-warmest/#92t4\\_Zb.Hmq7](https://mashable.com/2018/01/08/2017-record-year-billion-dollar-disasters-third-warmest/#92t4_Zb.Hmq7) (accessed on 13 August 2019).
13. Electric Reliability Council of Texas. ERCOT Responds to Hurricane Harvey. 1 September 2017. Available online: <http://www.ercot.com/help/harvey> (accessed on 29 July 2018).
14. U.S. Energy Information Administration. Hurricane Irma Cut Power to Nearly Two-Thirds of Florida’s Electricity Customers. 20 September 2017. Available online: <https://www.eia.gov/todayinenergy/detail.php?id=32992#> (accessed on 29 July 2018).
15. Maloney, P. Last Year’s Weather, Wildfires Heighten Utilities’ Storm Hardening Efforts. 21 May 2018. Available online: <https://www.utilitydive.com/news/last-years-weather-wildfires-heighten-utilities-storm-hardening-efforts/523666/> (accessed on 13 August 2019).
16. Eaton’s Blackout Tracker. Power Outage Annual Report. 2017. Available online: <https://switchon.eaton.com/plug/blackout-tracker> (accessed on 29 July 2018).
17. Wisconsin Public Service. Why Power Outages Occur. 2017. Available online: [http://www.wisconsinpublicservice.com/home/power\\_occur.aspx](http://www.wisconsinpublicservice.com/home/power_occur.aspx) (accessed on 11 September 2017).

18. Davies Consulting. *State of Distribution Reliability Regulation in the United States*; Edison Electric Institute: Washington, DC, USA, 2005; Available online: <https://legalectric.org/f/2010/04/stateofdistributionreliability-2005.pdf> (accessed on 13 August 2019).
19. Kenward, A.; Raja, U. *Blackout: Extreme Weather, Climate Change and Power Outages*; Technical Report; Climate Central, Inc.: Princeton, NJ, USA, 2014; Available online: [https://www.eenews.net/assets/2014/04/14/document\\_ew\\_01.pdf](https://www.eenews.net/assets/2014/04/14/document_ew_01.pdf) (accessed on 15 September 2017).
20. Campbell, R.J. *Weather-Related Power Outages and Electric System Resiliency*; Congressional Research Service Report; Congressional Research Service: Washington, DC, USA, 2012; pp. 1–15. Available online: <https://fas.org/sgp/crs/misc/R42696.pdf> (accessed on 15 September 2017).
21. National Academies of Sciences, Engineering, and Medicine. *Enhancing the Resilience of the Nation's Electricity System*; The National Academies Press: Washington, DC, USA, 2017; Available online: <https://doi.org/10.17226/24836> (accessed on 6 October 2017).
22. Reed, D. Electric utility distribution analysis for extreme winds. *J. Wind Eng. Ind. Aerodyn.* **2008**, *96*, 123–140. [CrossRef]
23. Sun, H.; Wang, Z.; Wang, J.; Huang, Z.; Carrington, N.; Liao, J. Data-Driven Power Outage Detection by Social Sensors. *IEEE Trans. Smart Grid* **2016**, *7*, 2516–2524. [CrossRef]
24. Guven, N.; Ozay, N.; Tunah, E. GIS based outage analysis system for electric distribution networks. In Proceedings of the 8th Mediterranean Electrotechnical Conference on Industrial Applications in Power Systems, Computer Science and Telecommunications (MELECON 96), Bari, Italy, 16 May 1996; Volume 2, pp. 761–764. [CrossRef]
25. EPRI Electric Power Research Institute. EPRI Distribution Modernization Demonstration (DMD) Data Mining Initiative. (n.d.). Available online: <https://smartgrid.epri.com/DMD-DMI.aspx> (accessed on 12 August 2019).
26. Georgia Spatial Data Infrastructure. About | GaSDI. 2014. Available online: <https://www.georgiaspatial.org/gasdi/about> (accessed on 13 August 2019).
27. NOAA's National Centers for Environmental Information. Storm Events Database | National Centers for Environmental Information. 2018. Available online: <https://www.ncdc.noaa.gov/stormevents/> (accessed on 13 August 2019).



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).