Simulating Transmission Power Infrastructural System Performance during Hurricanes with a Physics-based Data Generation Engine V1: A User Manual

Fatemehalsadat Jafarishiadeh[1] Xiao Zhu[2],

Ge Ou[2], Mostafa Sahraei-Ardakani[1], and Zhaoxia Pu[3]

Executive Summary

Power systems are vulnerable to severe weather events like hurricanes. The physical infrastructure of the power system, such as the transmission towers, poles, and substations, are impacted by the hazard loadings (wind speed, flood level) that exceed the design limit. The physical damage to the power infrastructure leads to the topology and network change of the power system and, therefore, power disruption and outage. The research of power system resilience analysis, planning, proactive action and reaction, and retrofit is facing the critical challenge of lack of comprehensive and realistic study of the power system vulnerabilities due to the scarcity of real power infrastructural and system data and proper data integration scheme. A data generation that simulates the realistic transmission power infrastructure and system performance during hurricanes is developed to address the challenges. This manual summarizes the concept, structure, and implementation of the data generation engine v1, demonstrating the benchmark case study of Texas power system performance during Hurricane Harvey.

[1] Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, UT 84112
[2] Department of Civil & Coastal Engineering, University of Florida Gainesville, FL 32611
[3] Department of Atmospheric Sciences, University of Utah, Salt Lake City, UT 84112

Table of Content

1. Overview of the Physics-based Data Generation Engine

Power systems are vulnerable to severe weather events like hurricanes. The physical infrastructure of the power system, such as the transmission towers, poles, and substations, are impacted by the hazard loadings (wind speed, flood level) that exceed the design limit. The physical damage to the power infrastructure leads to the topology and network change of the power system, therefore, in power disruption and outage.

The research of power system resilience analysis, planning, proactive action and reaction, and retrofit is facing the critical challenge of the lack of comprehensive and realistic study of the power system vulnerabilities due to the scarcity of real power infrastructural and system data and proper data integration schemes. Therefore, a physics-law-based data generation engine is developed to fill these gaps. The data generation engine includes four elements, A. hazard model, B. infrastructural mapping, C. infrastructural damage simulation, and D. power network simulation. Figure 1 depicts the data flow of the data generation engine.

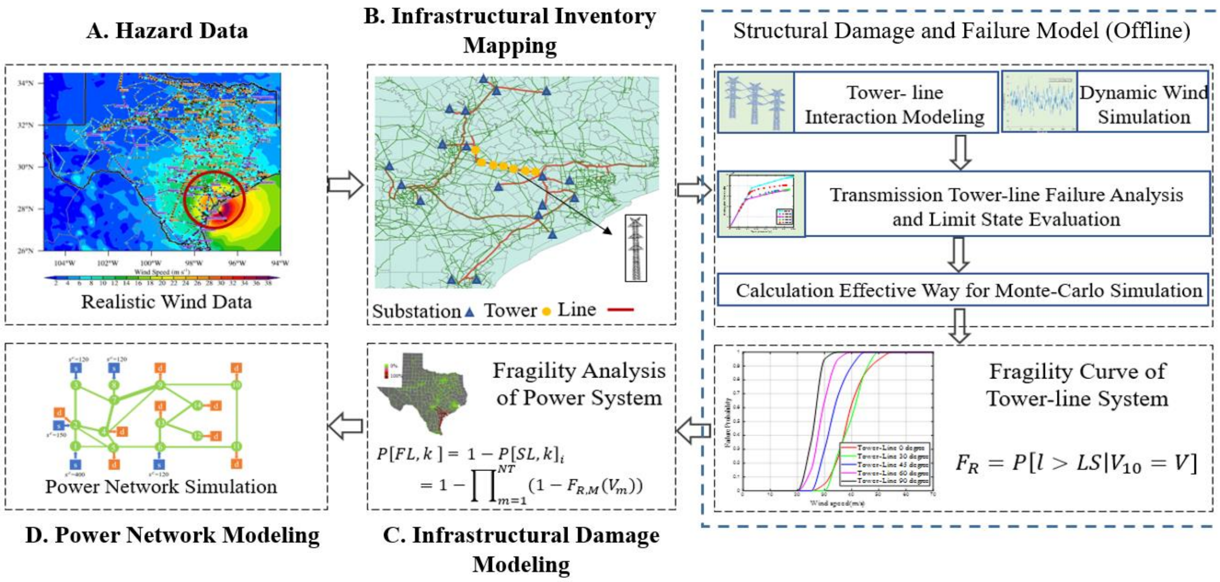

A. Hazard Data
Realistic Wind Data
B. Infrastructural Inventory Mapping
Substation ▲ Tower ● Line
Structural Damage and Failure Model (Offline)
Tower- line Interaction Modeling
Dynamic Wind Simulation
Transmission Tower-line Failure Analysis and Limit State Evaluation
Calculation Effective Way for Monte-Carlo Simulation
Fragility Curve of Tower-line System
$F_R = P[l > LS|V_{10} = V]$
Fragility Analysis of Power System
$P[FL,k] = 1 - P[SL,k]_i = 1 - \prod_{m=1}^{NT}(1 - F_{R,M}(V_m))$
Power Network Simulation
D. Power Network Modeling
C. Infrastructural Damage Modeling

Figure 1 - Realization of Physical Process of Power Infrastructural Damage and System Failure during Extreme Weather

Inspired by and benefiting from recently available infrastructural resilience software and cyberinfrastructure [1-5] developed by peer researchers and their teams, the unique features of the data generation engine version 1 are highlighted in the following:

- **Hazard Data:** High-resolution hazard information generated from the weather research forecasting (WRF) model is used as the hazard database.
- **Structural Inventory Data:** A complete structural inventory is designed following the industrial standard and is used to develop the structural performance and failure analysis model.

- **Power System Network Data:** Realistic large-scale synthetic power system models simulate the hazard impact on the power network. The synthetic power systems contain geographical information and network properties developed by Overbye et al. and have been validated with intensive real-world power system data [6-8].
- **Physical Infrastructure and Network Mapping:** The detailed structural inventory is mapped to the network following the geographical information of the network models. This mapping enables the physical world information (hazard intensities, infrastructural information, infrastructural damage, and failure) to transmit to the power system model.

As the first close-to-real-world power network performance simulation tool that captures the physics law of the transmission system during an extreme weather event, the first version release of the software focuses on the impact of the transmission tower-line damage on the power system. Many other realistic factors and uncertainties are currently excluded due to the complex nature of the modeling, and gaps in the state-of-the-art literature further open up frontier research opportunities. The appendix contains the terminologies of the power infrastructure and system resilience during a hurricane.

2. Data and File Structure of the Physics-based Data Generation Engine

The high-level data engine structure is illustrated in Figure 2. The hazard data is statically restored and provided in the hurricane database. The main executable codes of the data engine are divided into the infrastructure damage simulation (which covers the infrastructure and power network mapping module B in Figure and C in Figure 1) and the power system impact simulation (which covers module D in Figure).

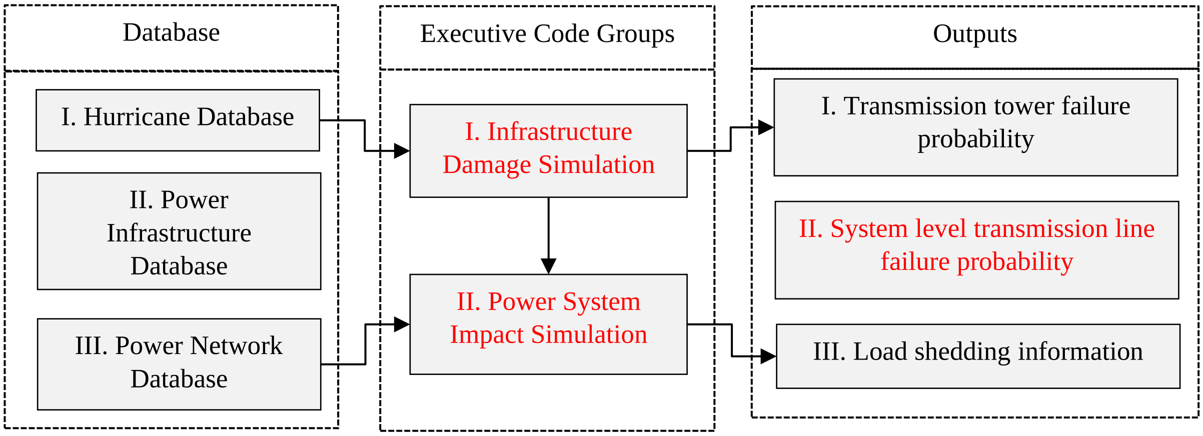

Database
Executive Code Groups
Outputs
I. Hurricane Database
II. Power Infrastructure Database
III. Power Network Database
I. Infrastructure Damage Simulation
II. Power System Impact Simulation
I. Transmission tower failure probability
II. System level transmission line failure probability
III. Load shedding information

Figure 2 - High-level diagram of data generation engine

The integrated data generation engine software user-defined inputs, and databases are tabulated in Table 1. This table shows that the user can investigate the power network's performance using different parameters (detailed described in Sections 2.1&2.2). The outputs of the developed software are listed in Table 2. This table explains the output file locations and types so that users can quickly locate and find those files. Moreover, after the software finishes running, a .txt file that summarizes the user-defined inputs and software outputs, including file name and location, will automatically be opened to help the users further understand what and where the files are.

Table 1 - Software Files and Codes Overview

	File type	**Category**
all_angleV10_R{i}km.mat † (i=1,2,4)	MATLAB data File	Input - Database I (see notes below)
all_speedV10_R{i}km.mat † (i=1,2,4)	MATLAB data File	Input - Database I (see notes below)
WindProfile_LatLon_R{i}km.mat † (i=1,2,4)	MATLAB data File	Input - Database I (see notes below)
TL_Fragility.xlsx	Excel File	Input - Database II
InfraS_GIS.xlsx	Excel File	Input - Database III
2K system_full	Excel File	Input - Database III
PS_loadfactor.txt	Text file	Input - Database III
Texas_Boundary.mat	MATLAB data File	Input - Database III
Texas_BaseMap.jpg	JPG File	Input - Database III
main.py*****	Python file	Executive Code User (see notes below)
main_offline.py*****	Python file	Executive Code User

		(see notes below)
InfraS_Failure_calc.py	Python file	Leading Code Group I
Hrcn_SCUC.py	Python file	Leading Code Group II

† Only i=4 (4km grid resolution wind filed) is provided on GitHub due to the file size limit. Other files can be downloaded here: https://www.dropbox.com/sh/vtthy31zp8qzlb9/AACrt8w5EW7xzC0cNyHTBGH7a?dl=0

* There are two ways to run the benchmark problem; main.py run a user-defined setup but requires an optimization software (default is IBM CPLEX, see the installation guide section in Section 3), where the main_offline.py runs a demonstrating scenario where the optimization is solved in advance and used default input parameters for software demonstration purpose. (Further explained in Section 4).

Table 2 - User Interactive Main Executive Code and Generated Outputs

File Name	**Property**	**Description (⁘) and Variable (⁕)**
main .py	Python File	⁘ The user executive file. The file integrates and runs all the functions defined to perform infrastructure damage simulation and power outage. After running, this file will create and open a .txt file that summarizes and shows the input and output files defined in the variables. The CPU consumed time is also shown. ⁕ resolution: Hurricane spatial resolution in Database I (select from 1, 2, 4) ⁕ time_interval_all: Hurricane temporal resolution in Database I (select any integer) ⁕ outputfd: Output folder name * Alfa: Line outage probability threshold (choose a value between 0 and 1, a smaller value is more conservative)
main_offline.py	Python File	⁘ The user executes file in an offline version, and the user-defined inputs are fixed for software demonstration purposes only. ⁘ resolution: Hurricane spatial grid resolution is fixed at 4km. ⁘ time_interval_all: Hurricane temporal resolution is fixed at 3 hours apart. ⁘ Alfa: 0.5 ⁕ outputfd: Output folder name
Tower_GIS.json	Output	⁘ The towers and line geographical coordinates.
LineFailure_P.json	Output	⁘ Transmission line failure probability at each time stamp
Ldamage_T1.png, Ldamage_T2.png, …	Output	⁘ The images of the transmission line failure probability visualization mapped into the concerned region
InfraS_Damage.avi	Output	⁘ The videos of the transmission line failure probability visualization mapped into the concerned region
PS_results.xlsx	Output	⁘ Summary, commitment, monitor (monitored lines), and load shedding sheets
LoadShedding.xlsx	Output	⁘ The load shedding information of buses, zones, and their geographical locations
LSH_1.png, LSH_2.png, …	Output	

2.1. Code Group I Infrastructure Damage Simulation – Code and File Structure

The structure of the developed code simulates the impact of the hurricane on the physical power infrastructure is detailed in Table 3. This table describes the functionality of the python file and the input-output it generates.

The python file '*TL_failure_calc.py*' contains the functions required to map the transmission tower infrastructure to the geographical location. It calculates their physical damage and failure with a given wind field and is imported into the '*main.py/main_offline.py*' code. Table 3 also contains inputs, outputs, and the purpose of each function. The dependency of various functions on each other is illustrated in **Error! Reference source not found.**.

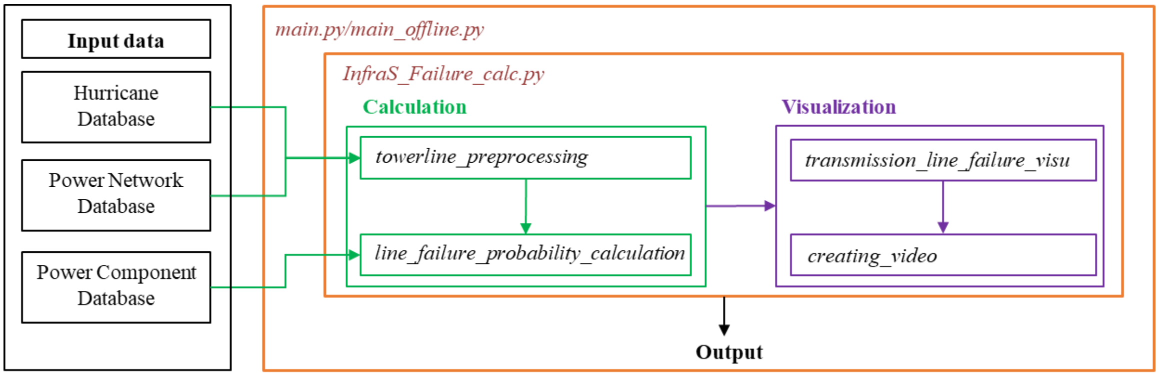


Figure 3 - Power system physical infrastructure failure and damage calculation flowchart

Table 1 - Code Structure of Executive Code Group I Infrastructure Damage Simulation

File Name	Property	Description and Variable (∗)
InfraS_Failure_calc.py	Python File	Python file defines the *RegionalHurricaneTLFailureCal* class and the necessary Python packages.
RegionalHurricaneTLFailureCal	Python Class	Python file contains all the functions needed to calculate and visualize the transmission tower-line failure probability and defines variables controlling the output. ∗ opfolder: Output folder name referred to outputfd in main.py/main_offline.py ∗ pre_process_fn: Transmission tower-line failure probability output file name referred to outputfn in main.py/main_offiline.py. ∗ pvideo: Optional parameter controls whether to play the generated video after finishing running the code. Default: True. ∗ dist: Optional parameter controls the distance between two towers along a transmission line. Default: 0.2 (km).
towerline_preprocessing(hurricane_model, powernetwork_model)	Python Function	This function calculates the number of towers in each line and each tower's geographic coordinate. ∗ hurricane_model: wind profile extraction/observation latitude and longitude coordinates referred to WindProfile_LatLon_R{i}km.mat (i=1,2,4) in Database I. *i* is referred to the variable resolution in main.py/main_offline.py. ∗ powernetwork_model: abstract synthetic power network model referred to InfraS_GIS.xlsx in Database III.
line_failure_probability_calculation(TLFragility,	Python Function	This function aims to calculate the time-dependent transmission line failure probability and the transmission tower failure probability for the hurricane event.

v10_angle, *v10_speed,* *time_interval)*		⁕ TLFragility: Transmission tower fragility curves referred to TL_Fragility.xlsx in Database II. ⁕ v10_angle: Mean wind angle at height 10m referred to all_angleV10_R{i}km.mat (i=1,2,4) in Database I. *i* is referred to as the variable resolution in main.py/main_offline.py. ⁕ v10_speed: Mean wind speed at height 10m referred to all_speedV10_R{i}km.mat (i=1,2,4) in Database I. *i* is referred to as the variable resolution in main.py/main_offline.py. ⁕ time_interval: hurricane temporal resolution referred to variable time_interval_all in main.py/main_offline.py.
transmission_line_failure_visu(Linefailure, *time_interval,* *boundary)*	Python Function	This function aims at visualizing the failure probability of the transmission line in the regional map, which is Texas_BaseMap.jpg in Database III. Cold to warm color maps represents the numerical probability from 0 percent to 100 percent. ⁕ Linefailure: Time-dependent transmission line failure probability file. This file is the output of the function *line_failure_probability_calculation*. The file is stored in the user-defined output folder outputfd with the file name outputfn, defined in main.py/main_offline.py. ⁕ time_interval: hurricane temporal resolution referred to variable time_interval_all in main.py/main_offline.py. ⁕ boundary: regional area boundary geographic coordinate referred to Texas_Boundary.mat in Database III. Image processing techniques will map this file into the Texas_BaseMap.jpg in Database III.
creating_video()	Python Function	This function aims at gathering the generated images into a single video. ⁕ time_interval: hurricane temporal resolution referred to variable time_interval_all in main.py/main_offline.py.
playing_video()	Python Function	The function determines whether the created video will be played after running the main.py/main_offline.py. The default is set to True, meaning the video will be played. This function is compatible with both MAC and Windows operational systems. ⁕ time_interval: hurricane temporal resolution referred to variable time_interval_all in main.py/main_offline.py.

2.2. Code Group II Power System Impact Simulation -– Code and File Structure

The structure of the developed code simulates the impact of the hurricane on the power system operation is detailed in Table 4. The python file '*Hrcn_SCUC.py*' contains the functions required to solve the optimization problem and is imported into the '*main.py/main_offline.py*' code. The following table also contains inputs, outputs, and the purpose of each function. The dependency of various functions on each other is illustrated in Figure 4.

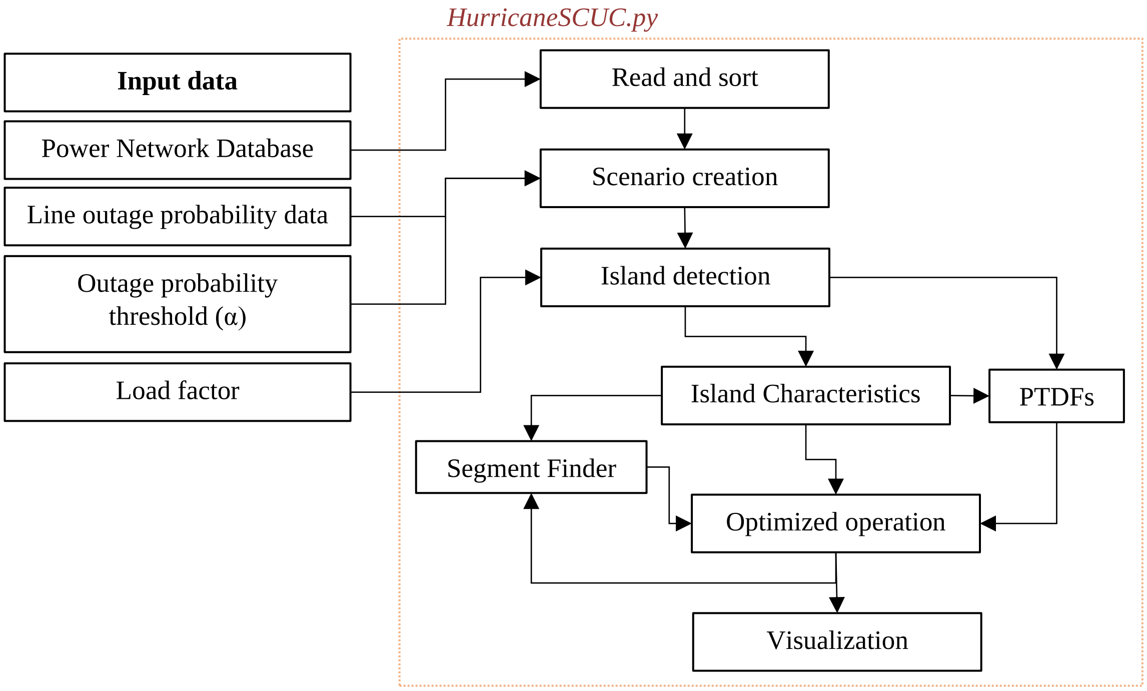


Figure 4 - The code structure for optimized operation of the power system

Table 4 - Code Structure of Executive Code Group II Power System Impact Simulation

File Name	Property	Description and Variable (⁎)
Hrcn_SCUC.py	Python file	This file is the main part of the executive code in group II. The required packages are imported, and related functions are defined in this file.
HurricaneSCUC	Python class	Classes provide a means of bundling data and functionality together. All the functions are defined under the class *HurricaneSCUC*.
clean_data(data0)	function	Select, check, and sort the data in the data0 excel file. The function returns line, generator, bus, and key NumPy arrays. ⁎ data0: line array.
ReadJSN(Datafile)	function	This file reads the line failure probability file (Datafile). The function returns outage probabilities and time stamps. ⁎ Datafile: Transmission line failure probability at each time stamp referred to as *LineFailure_P.json*.
scenario_gen(data0,data1 ,data2)	function	Based on the threshold (data2) defined by the user, lines with failure probabilities higher than the threshold are considered line failures. The function returns the line status array. ⁎ data0: line array, ⁎ data1: line failure probability JSON file referred to *LineFailure_P.json*. ⁎ data2: user defied line failure probability threshold.
Island(data0,data1,data2, data3)	function	The connections of the lines given the line and line status arrays are checked, and various segments and isolated nodes are identified for each time step of the load factor. The nodes in segments with more than one node are reviewed, and islands with the same nodes are considered the consistent structure. The formed islands, isolated

		nodes, and the time stamps of change in the structures are returned as outputs. ⁕ data0: line array. ⁕ data1: line failure probability JSON file referred to *LineFailure_P.json*. ⁕ data2: load factor. ⁕ data3: bus array.
Net_info(Bus, Line, Gen, LineStatus, Load_fact, new_dict, U, U_i):	function	This function reads the bus, line, generator, line status, load factor arrays, and island structures as dictionaries. As the structure of the power system changes, generators, buses, and lines will belong to different segments. This function returns the segments and the generators, lines, and buses in those segments.
Segment_finder(HH2,dict_L,M)	function	In this function, the lines are to be monitored, find what segment they belong to, and return the segment's name. ⁕ HH2: the hours in which the island structure does not change. ⁕ dict_L: a dictionary of lines for various segments. ⁕ M: monitored lines.
PTDF(Bus, Line, new_dict, U, U_i)	function	This function calculates power transfer distribution factors for various formed segments. The inputs are bus and line arrays and the structure of various islands.
Opt_operation(Power_system,Load_factor,LineFailureProbability,TIme0,Alfa)	function	This function finds the optimum solution to the power dispatch problem and returns the load shedding information, folder location for saving the results, and a summary of the optimization solution. ⁕ Power_system: power network database. ⁕ Load_factor: load factor. ⁕ LineFailureProbability: line failure probability JSON file referred to *LineFailure_P.json*. ⁕ TIme0: time duration of interest to run the optimization problem. ⁕ Alfa: user defied line failure probability threshold.
vis_img(data0)	function	This function generates a geographic bubble chart where latitude and longitude specify the geographic locations of load shedding. The areas of the bubbles are scaled according to the numeric value of load shedding. ⁕ data0: line array.

3. Software Environment and Installation Guide

To evaluate the performance of the power systems under hurricane events, the Data Generation V1 is developed using Python programming language. Using the developed Python scripts, some necessary packages and software are needed. Here, a brief description of the needed packages is introduced. The detailed installation instruction can be found in a separate file.

- Download and install Python, Anaconda, and Python IDEs (Spyder, Pycharm)
- *Download and install IBM CPLEX Optimization software* (for running power system dispatch optimization, this step can be skipped if only running ‘main_offline.py’)
- *Setup IBM CPLEX Python API*
- Create Python Local Virtual Environment
- Install the requirements.txt dependency file in the build Python Virtual Environment

4. Data Generation Engine Implementation on the Hurricane Harvey Benchmark Problem

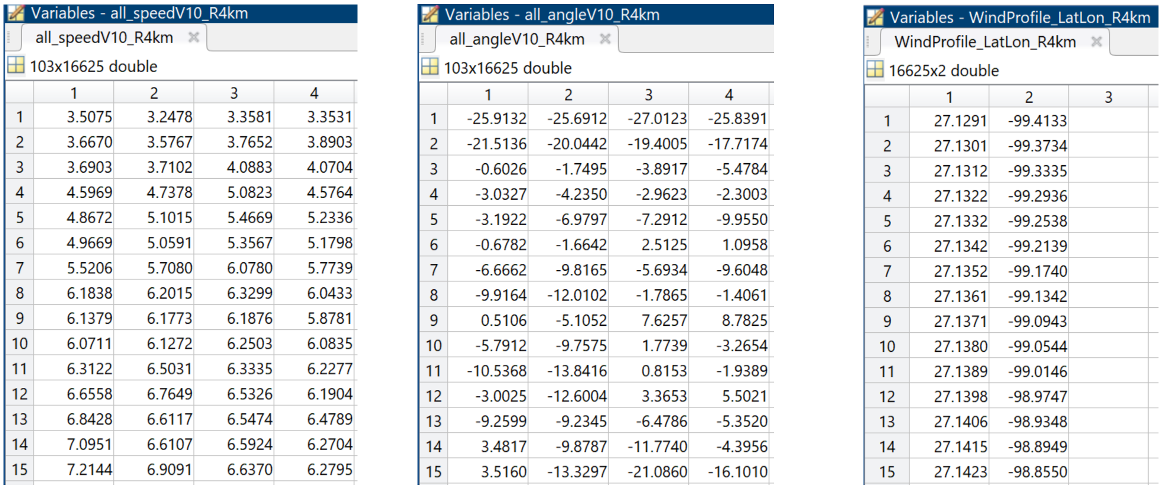
The impact of Hurricane Harvey on the power network database of the synthetic Texas 2000 bus system on the footprint of Texas is studied in this section. This system has 3206 transmission elements, 544 generators, and 2000 buses. As Hurricane Harvey hit Texas at 12 pm on the 25th of August (hour 16 of SCUC), it lasted until 3 pm on the 27th (hour 64 of SCUC). The impact on the infrastructure is assessed based on hazard data, and the probability of their failures is calculated. These failure probabilities are then used to generate scenarios of the compromised power network. The capability of the impaired power system in supplying electricity to consumers is explored, which might lead to load shedding. In that case, the location and amount of load shedding are calculated.

4.1. Hurricane Harvey Benchmark Scenarios and Inputs

The hurricane model used is the Hurricane Harvey numerical model, which utilized the data from the NCEP GFS FNL (National Centers for Environmental Prediction Global Forecast System final analysis) and the MODIS (Moderate Resolution Imaging Spectroradiometer).

4.1.1. *Hazard Data (Database I)*

The WRF model is a mesoscale numerical weather prediction system with 1 km of horizontal resolution and half-an-hour temporal resolution from the 25th to the 27th of August. The hazard data are processed into two categories: the wind speed and wind angle, to make them compatible with the database. The wind speed and angle are stored as a .mat file with the names ‘*all_angleV10_R{i}km.mat*’ and ‘*all_angleV10_R{i}km.mat*’. The wind speed and wind angle have the same data format, in which the rows represent the time the data has been collected, and the columns represent the values of the observations. The geographic locations of the hazard data are stored in .mat format with the name ‘*WindProfile_LatLon_R{i}km.mat*’, for which the rows are the index and the columns are latitude and longitude coordinates. The wind speed locations and wing angle are aligned with the ‘*WindProfile_LatLon_R{i}km.mat*’ file by the index. For example, for the wind speed in ‘*all_angleV10_R4km.mat*’, the column index can point to the row index in ‘*WindProfile_LatLon_R4km*’, at which the two columns at that row index are the location where the data has been observed. Figure 5 shows the screenshot of the hazard data in Database I. As shown in this figure, the column number in the wind speed and angle data is the same as the row number in the wind profile latitude and longitude data, which establishes the link between categorized hazard data files. Figure 6 shows the projection of the hazard data into the ‘*Texas_BaseMap.jpg*’ in database III.



Variables - all_speedV10_R4km

all_speedV10_R4km

103x16625 double

	1	2	3	4
1	3.5075	3.2478	3.3581	3.3531
2	3.6670	3.5767	3.7652	3.8903
3	3.6903	3.7102	4.0883	4.0704
4	4.5969	4.7378	5.0823	4.5764
5	4.8672	5.1015	5.4669	5.2336
6	4.9669	5.0591	5.3567	5.1798
7	5.5206	5.7080	6.0780	5.7739
8	6.1838	6.2015	6.3299	6.0433
9	6.1379	6.1773	6.1876	5.8781
10	6.0711	6.1272	6.2503	6.0835
11	6.3122	6.5031	6.3335	6.2277
12	6.6558	6.7649	6.5326	6.1904
13	6.8428	6.6117	6.5474	6.4789
14	7.0951	6.6107	6.5924	6.2704
15	7.2144	6.9091	6.6370	6.2795

(a)

Variables - all_angleV10_R4km

all_angleV10_R4km

103x16625 double

	1	2	3	4
1	-25.9132	-25.6912	-27.0123	-25.8391
2	-21.5136	-20.0442	-19.4005	-17.7174
3	-0.6026	-1.7495	-3.8917	-5.4784
4	-3.0327	-4.2350	-2.9623	-2.3003
5	-3.1922	-6.9797	-7.2912	-9.9550
6	-0.6782	-1.6642	2.5125	1.0958
7	-6.6662	-9.8165	-5.6934	-9.6048
8	-9.9164	-12.0102	-1.7865	-1.4061
9	0.5106	-5.1052	7.6257	8.7825
10	-5.7912	-9.7575	1.7739	-3.2654
11	-10.5368	-13.8416	0.8153	-1.9389
12	-3.0025	-12.6004	3.3653	5.5021
13	-9.2599	-9.2345	-6.4786	-5.3520
14	3.4817	-9.8787	-11.7740	-4.3956
15	3.5160	-13.3297	-21.0860	-16.1010

(b)

Variables - WindProfile_LatLon_R4km

WindProfile_LatLon_R4km

16625x2 double

	1	2	3
1	27.1291	-99.4133	
2	27.1301	-99.3734	
3	27.1312	-99.3335	
4	27.1322	-99.2936	
5	27.1332	-99.2538	
6	27.1342	-99.2139	
7	27.1352	-99.1740	
8	27.1361	-99.1342	
9	27.1371	-99.0943	
10	27.1380	-99.0544	
11	27.1389	-99.0146	
12	27.1398	-98.9747	
13	27.1406	-98.9348	
14	27.1415	-98.8949	
15	27.1423	-98.8550	

(c)

Figure 5 - Hazard data screenshot for 4km spatial resolution. (a) Wind speed; (b) Wind angle; (c) Wind profile latitude and longitude coordinate

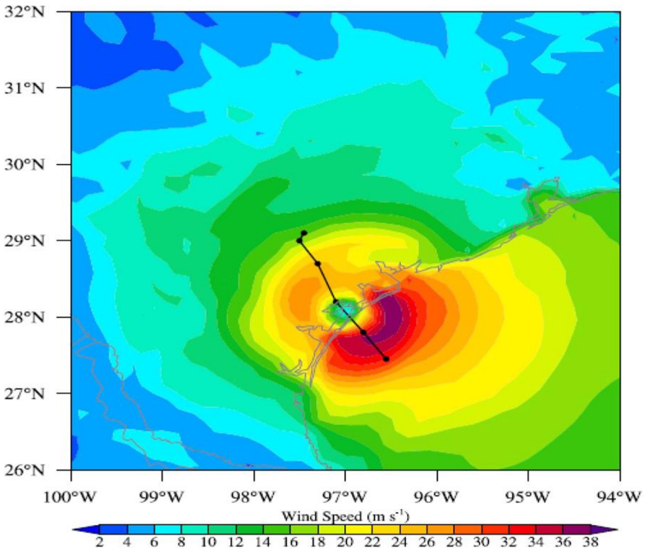
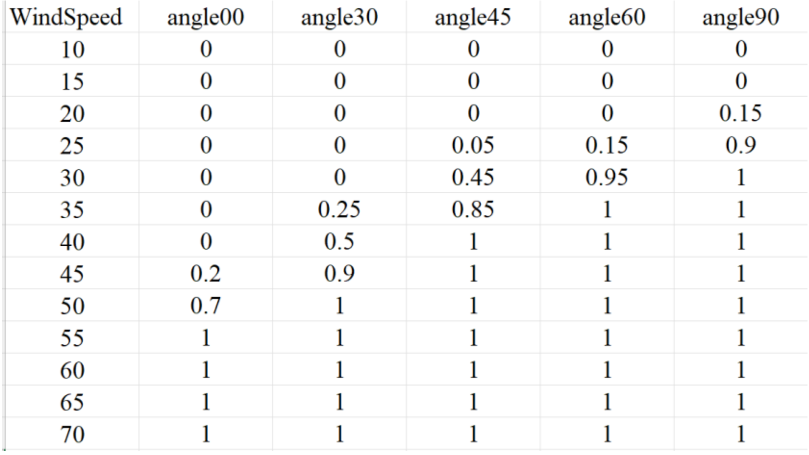


Figure 6 - Wind velocity distribution in the Texas region during Hurricane Harvey

4.1.2. *Infrastructure Data (Database II)*

The infrastructure database contains the infrastructure inventory and structural model of the power infrastructure. In version 1 of the software, the transmission tower is uniformly designed. For rapid analysis, the fragility curves of the transmission tower will be generated using the tower finite element model offline. The transmission tower fragility curve expresses the failure probability as a function of the mean wind speed and relative wind angle at 10m height. The towers in the synthetic power grid are assumed to be identical. Therefore, a homogenous fragility curve is used for all the towers in the testbed. In database II, the infrastructure data are stored as '*TL_Fragility.xlsx*'. The spreadsheet's rows represent the wind speed in meter-per-second, and the columns represent the tower failure probability at each angle. Figure 7 shows the fragility calculated at discrete wind speed and angle. Figure 8 shows the interpolated curves for each column.



WindSpeed	angle00	angle30	angle45	angle60	angle90
10	0	0	0	0	0
15	0	0	0	0	0
20	0	0	0	0	0.15
25	0	0	0.05	0.15	0.9
30	0	0	0.45	0.95	1
35	0	0.25	0.85	1	1
40	0	0.5	1	1	1
45	0.2	0.9	1	1	1
50	0.7	1	1	1	1
55	1	1	1	1	1
60	1	1	1	1	1
65	1	1	1	1	1
70	1	1	1	1	1

Figure - 7 Infrastructure data represented by fragility curves screenshot

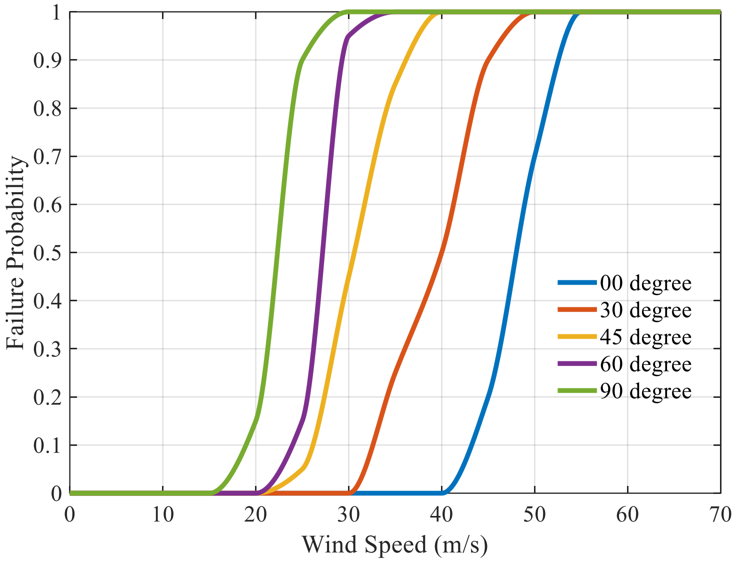
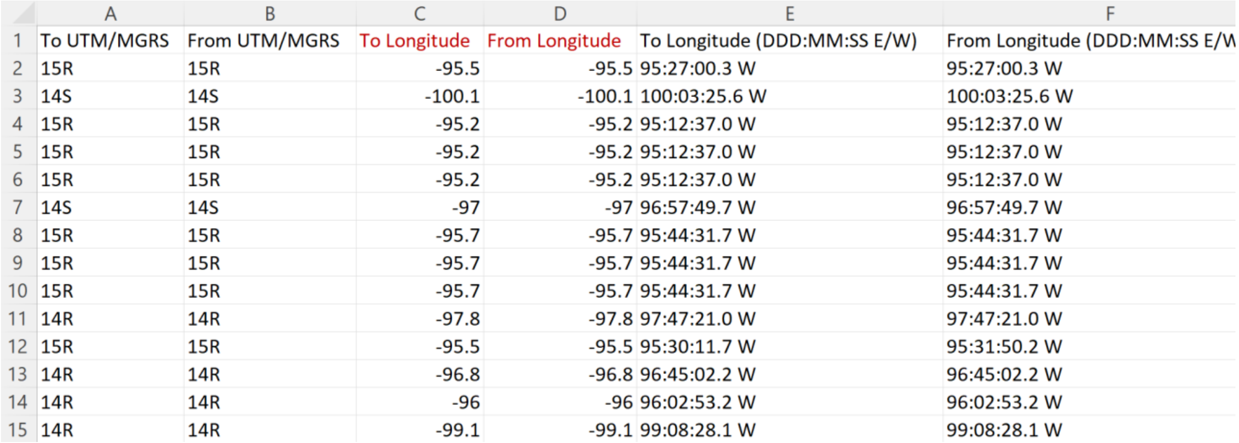


Figure 8 - Transmission tower fragility curves

4.1.3. *Power Network Data (Database III)*

The power network database of the synthetic 2000 bus system on the footprint of Texas is used as the test power system. This system has 3206 transmission elements, 544 generators, and 2000 buses. After the processing, there are 2668 transmission lines and 136913 transmission towers in this synthetic power network database. The power network data are stored in '*InfraS_GIS.xlsx*'. In this data file, the rows represent the transmission lines in the power network, whereas each column represents the transmission lines' geographic, end stations, and capacity information. Figure 9 shows the screenshot of the database excel file. The power network database contains information on the transmission lines, and the transmission tower information is missing. The distribution of the transmission line is calculated by assuming that 1) the transmission line is straight between two stations; 2) the distance between successive towers is 0.2km in a line. The tower geographic coordinate of each tower will be obtained and stored in the '*Tower_GIS.json*' output file using the '*towerline_preprocessing*' function defined in Table 3. Combined with the '*Tower_GIS.json*', ‘*Texas_BaseMap.jpg*’, and ‘*Texas_Boundary.mat*’, Figure 10 illustrates the distribution of the power network in the Texas region.



	A	B	C	D	E	F
1	To UTM/MGRS	From UTM/MGRS	To Longitude	From Longitude	To Longitude (DDD:MM:SS E/W)	From Longitude (DDD:MM:SS E/W
2	15R	15R	-95.5	-95.5	95:27:00.3 W	95:27:00.3 W
3	14S	14S	-100.1	-100.1	100:03:25.6 W	100:03:25.6 W
4	15R	15R	-95.2	-95.2	95:12:37.0 W	95:12:37.0 W
5	15R	15R	-95.2	-95.2	95:12:37.0 W	95:12:37.0 W
6	15R	15R	-95.2	-95.2	95:12:37.0 W	95:12:37.0 W
7	14S	14S	-97	-97	96:57:49.7 W	96:57:49.7 W
8	15R	15R	-95.7	-95.7	95:44:31.7 W	95:44:31.7 W
9	15R	15R	-95.7	-95.7	95:44:31.7 W	95:44:31.7 W
10	15R	15R	-95.7	-95.7	95:44:31.7 W	95:44:31.7 W
11	14R	14R	-97.8	-97.8	97:47:21.0 W	97:47:21.0 W
12	15R	15R	-95.5	-95.5	95:30:11.7 W	95:31:50.2 W
13	14R	14R	-96.8	-96.8	96:45:02.2 W	96:45:02.2 W
14	14R	14R	-96	-96	96:02:53.2 W	96:02:53.2 W
15	14R	14R	-99.1	-99.1	99:08:28.1 W	99:08:28.1 W

Figure - 9 Power network database screenshot (only part of the information is shown)

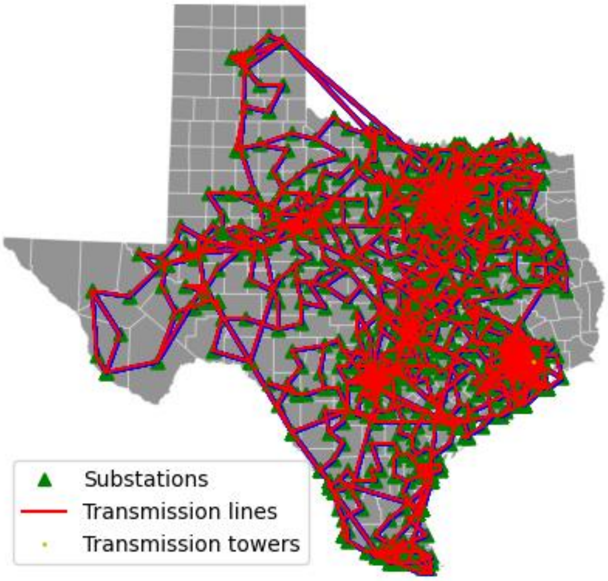


Figure 10 - The distribution of the power network components processed by the power infrastructure module

The line failure probabilities have values between 0 and 1. Assuming the line outage probability threshold is 0.5, one scenario is generated in which the status of 191 lines is changed to 0 throughout the hurricane, indicating line failure. By the 72nd hour, 3015 lines remain operational. After solving the optimal operation problem, the load shedding in various zones in Texas during the 72 hours is calculated, and the results are shown in **Error! Reference source not found.**Figure 11.

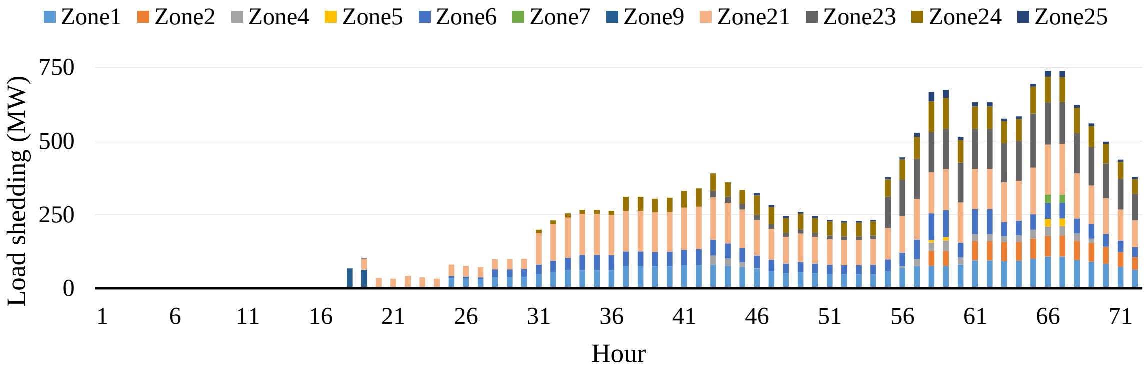


Figure 11 - Cumulative load shedding at impacted zones

4.2. Benchmark Problem and Demonstration Scenario

The user is provided with two scripts to run: main.py and main_offline.py. The demonstration scenario discusses the pre-determined parameters in the offline script main_offline.py. By executing the codes, the user can calculate, visualize the power network infrastructure component damage and optimize the power delivery, and investigate how different parameters include: hazard data spatial and temporal resolution and thresholds of the transmission line failure probability.

4.2.1. *Default user-defined parameters for running the demonstration scenario*

Executing the offline script will run the code with the spatial resolution of the data being 4km. As the latitude and longitude coordinates of the observed wind profiles are provided, the user can use other resolutions at 1 km and 2 km in the running main.py. The minimum wind profile temporal resolution in Hurricane Harvey is half an hour. In the main_offline.py file, a temporal resolution is determined as 3 hours. If the user runs main.py, the spatial resolution parameter can be chosen from $[1.0, 2.0, 3.0]\ km$. For each specified spatial resolution, the temporal resolution can choose from $[0.5, 1.0,1.5, 2.0, 2.5, 3.0]\ hours$.

The line failure probability threshold is the key parameter for calculating infrastructure-induced power outage. By choosing a lower value, more line outages are considered in the system, and the impact of the hurricane on the power system is more severe. Correspondingly, it is expected that the amount of load shedding calculated would increase. The default in the main_offline.py is set as 0.5, and based on that, the optimization solution is loaded from the 'Input_offline' folder. This folder contains the pre-calculated solution to the optimization problem. The script will then plot the load shedding on the geographical map of the area under study. User is asked to set their own threshold parameter for running the main.py version. The online script (main.py) allows the user to choose a value between 0 and 1. Based on the value chosen, the optimization problem will be solved (requiring a specific software such as IBM CPLEX). This part is time-consuming and, depending on the size of the power system chosen and the number of days of the hurricane might take a few minutes to days to solve.

After running the software, the transmission line failure and load shedding will automatically show as a video. In this user-defined output folder, all the calculated data will be stored and can be used for any other purposes.

4.3. Outputs from the Demonstration Scenario

The executive code group I targets calculating the failure probability of the transmission tower and lines under hazard. All the outputs are contained in the user-defined folder named ‘*outputfd*’. The calculating and mapping of the transmission tower and line into the region of interest is stored in the JSON file with the file name ‘*outputfn*’. The transmission line failure probability is stored as ‘*LineFailure_P.json’*. Visualization of the transmission line failure is stored as a cluster of PNG images and an *avi* format video in the folder ‘*outputfd*’. Table 5 summarizes the output files of code group I by choosing the input parameters: 4km spatial resolution; 3 hours temporal resolution.

Table 5 - Executive Code Group I outputs

Name	**Location**	**Containing**	**Description**
Tower_GIS_R4.0km.json	‘*outputfn*’ folder	The towers and line information. Each line contains: ‘Number of Towers’, ‘Tower coordinate’, ’Tower associated wind profile’, ’Stations’, ’Line end coordinate’.	• Key-value format; the first layer of keys is the line number; the value contains the tower information for each line • Number of Towers: Number of towers for each line • Tower coordinate: latitude and longitude for each tower • ’Tower associated wind profile: wind profile index in the hazard database (Database I) • Stations: the station number from the power network database (Database III) • Line end coordinate: the two end station coordinate in the hazard database (Database I)
LineFailure_P.json	‘*outputfn*’ folder	Transmission line failure probability at each time stamp	• Key-value format; the first layer of keys is the line number; the value contains the failure probability at each time stamp
Ldamage_T1.png, Ldamage_T2.png, …	‘*outputfn*’ folder	The images of the transmission line failure probability visualization mapped into the Texas region	Images of the failure probability of the transmission line at each time stamp were mapped into the Texas region. Cold to warm color maps represents the numerical probability from 0 percent to 100 percent.
InfraS_Damage.avi	‘*outputfn*’ folder	AVI formatted video	Video of the transmission line failure at each time stamp. It gathers all the images sequentially.

The results of the main.py script on the power system operation can be divided into two groups, as shown in Table 6:

1- Two excel files containing the solutions to the unit commitment optimization problem, and the load shedding details are saved in the ‘*Result_ps*’ folder. The ‘*PS_results.xlsx*’ file contains Summary, Commitment, Monitor (monitored lines), Power (total power produced), and Load_Shedding sheets. In the ‘Summary’ sheet, the objective value for the optimization problem, the total time required to solve the problem, and load shedding values are saved. The commitment status of various generators in the system is specified in the ‘Commitment’ sheet. The total power generated by each generator is stated in the ‘Power’ sheet. Monitored lines are listed in the ‘Monitor’ sheet. The load shedding bus number and values are detailed in the ‘Load_shedding’ sheet. The second excel file is ‘LoadShedding.xlsx’, which contains the load shedding information of buses, zones, and their geographical locations.

2- The images of the geographical area impacted by the hurricane and the location of load shedding are saved in the ‘Image_ps’ folder. The plotting tool accordingly incorporates the map's comparative values of load shedding and scale bubbles. These load shedding images are generated for all the hours in the simulation, and there might be several hours with no load shedding depending on the hour the hurricane impacts the power system.

The results of the ‘main_offline.py’ script only contain the images on load shedding location as described above.

Table 6 - Results of the main.py script

Name	**Location**	**Containing**	**Description**
PS_results.xlsx	‘Result_ps’ folder	summary, commitment, monitor (monitored lines), ptotal (total power produced), and load shedding sheets	• In the ‘Summary’ sheet, the objective value for the optimization problem, the total time required to solve the problem, and load shedding values are saved. • The commitment status of various generators in the system is specified in the ‘Commitment’ sheet. • The total power generated by each generator is stated in the ‘Power’ sheet. • Monitored lines are listed in the ‘Monitor’ sheet. • The load shedding bus number and values are detailed in the ‘Load_shedding’ sheet.
‘LoadShedding.xlsx’	‘Result_ps’ folder	Lshedding_loc	This sheet contains the load shedding information of buses, zones, and their geographical locations
Hour1.png, Hour2.png, …	‘Image_ps’ folder	-	Images of load shedding as scaled bubbles in the geographical map of the impacted area.

4.4. Result Visualizations

One distinguishing feature of the proposed Data Generation Engine V1 is visualizing the transmission line failure probability under hazard. For the Hurricane Harvey benchmark case, the visualization of the transmission line helps the user get an intuitive idea of the transmission line state as the hurricane formulates, lands, and fades. Figures 12 to Figure 15 shows the failure progression of the transmission line during Hurricane Harvey from the 25th of August to the 27th of August. The input parameters are 4km spatial resolution; 3 hours temporal resolution.

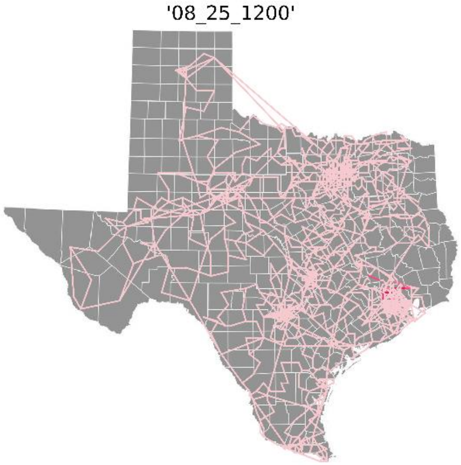


Figure 12 - The failure of the transmission line in the 2000-bus Texas power grid at 12:00 am on the 25th of August

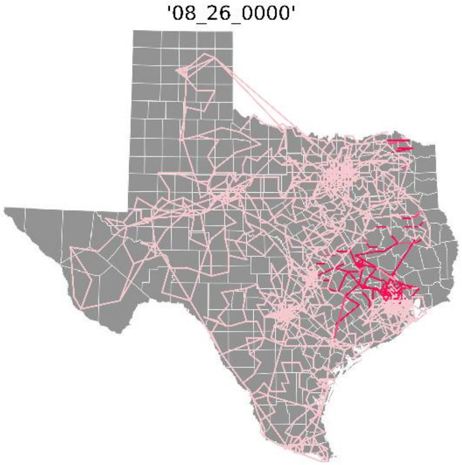


Figure 13 - The failure of the transmission line in the 2000-bus Texas power grid at 00:00 am on the 26th of August

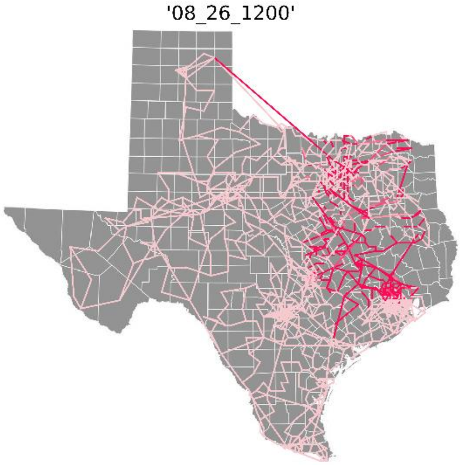


Figure 14 - The failure of the transmission line in the 2000-bus Texas power grid at 12:00 am on the 26th of August

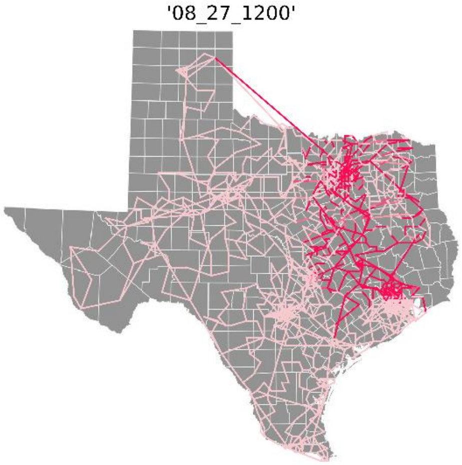


Figure 15 - The failure of the transmission line in the 2000-bus Texas power grid at 12:00 am on the 27th of August

The line failure probability data is one input data used for the optimal power system operation. Also, the power network database of the synthetic 2000 bus system on the footprint of Texas is used as the test power system. This system has 3206 transmission elements, 544 generators, and 2000 buses. As Hurricane Harvey hits Texas at 12 pm on the 25th of August (hour 16 of SCUC) and lasts until 15 pm 27th (hour 64 of SCUC), the time window of optimal power system operation should contain this period. Therefore, the time window of 3 days (72 hours) is chosen to be studied for the operation schedule. Assuming the line outage probability threshold is 0.5, the status of 191 lines is changed to 0 throughout the hurricane, indicating line failure. By the 72nd hour, 3015 lines remain operational. The increase in line outages during the three days of study is shown in Figure 16. After running the optimal operation code, the load shedding in various zones in Texas during the 72 hours is calculated, and the results are shown in Figure 17.

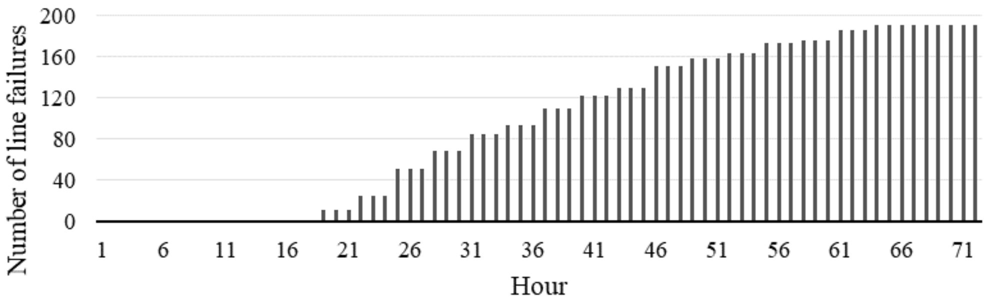


Figure 16 - Number of line outages during the hurricane event

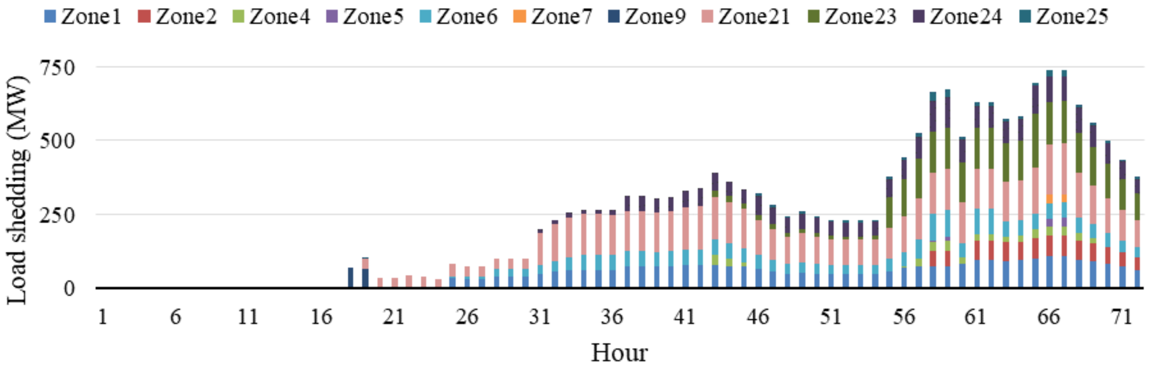


Figure 17 - Cumulative load shedding at impacted zones

The load shedding increases as the number of line outages increases, as shown in Figure 18 and Figure 19. The load shedding at hours 46 and 72 at different buses in their respective zones is depicted with the same colors and size relative to the amount of load shedding.

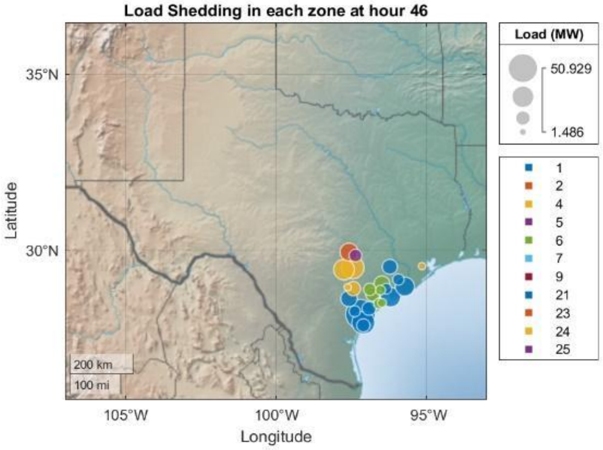


Figure 18 - Load shedding at hour 46

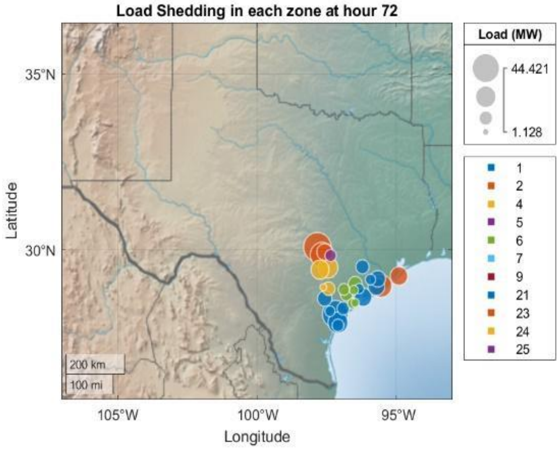
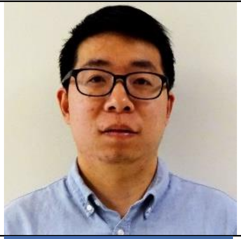
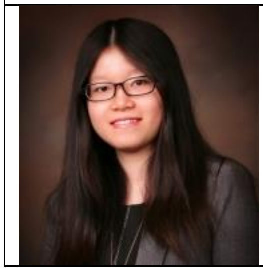


Figure 19 - Load shedding at hour 72

5. Development Team and Engagement Statement

Development Team	
	Fatemehalsadat Jafarishiadeh Ph.D. Candidate, Electrical and Computer Engineering University of Utah Software development and power network simulation
	**Xin Li** Ph.D. Student, Atmospheric Sciences University of Utah Wind field data generation
	Xiao Zhu Ph.D. Candidate, Civil and Coastal Engineering, University of Florida Software development, infrastructural data generation, infrastructural damage simulation
	Dr. Mostafa Ardakani Leading electrical engineering team that employs the failure probabilities and includes them within the power system operation and planning models. Electrical and Computer Engineering, University of Florida
	Dr. Zhaoxia Pu Leading the atmospheric sciences team that develops high-resolution weather forecast models to provide appropriate input data to the rest of the team. Atmospheric Sciences, University of Utah
	**Dr. Gaby Ou** Leading the civil engineering team estimates the failure probability of power infrastructure elements. Civil and Coastal Engineering, University of Florida

The software is developed by a multi-displinary team with expertise in meteology, civil engineering, and electrical engineering. The software and dataset generated in the project are complied with the Creative Commons Atribution 4.0 International License. The development team welcome the reuse and reproduction of the data generation engine and the datasets. Projects developed with the reuse of the data generation engine shall cite the data DOI: 10.5281/zenodo.7139787.

Questions, comments, and collaboration interests are welcome to contact the PIs: Gaby Ou[1] and Mostafa Ardakani[2].

Acknowledgment

The development of this software and user manual was supported by the Cyberinfrastructure for Sustained Scientific Innovation (CSSI) program in National Science Foundation, under award 2004658, Elements: Open Access Data Generation Engine for Bulk Power System under Extreme Windstorms. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

[1] mailto:Gaby.Ou@essie.ufl.edu
[2] mailto: mostafa.ardakani@utah.edu

Reference

[1] H. Beckmann and D. M. Simpson, "Risk assessment and GIS in natural hazards: issues in the application of HAZUS," *International Journal of Risk Assessment and Management,* vol. 6, no. 4-6, pp. 408-422, 2006/01/01 2006, doi: 10.1504/IJRAM.2006.009544.

[2] P. Bocchini, B. Davison, A. M. Esnard, A. Lamadrid, D. Mitsova-Boneva, and A. Sapat, "Probabilistic resilience assessment of interdependent systems (PRAISys)," ed: Retrieved December, 2019.

[3] G. G. Deierlein *et al.*, "A Cloud-Enabled Application Framework for Simulating Regional-Scale Impacts of Natural Hazards on the Built Environment," (in English), *Frontiers in Built Environment,* vol. 6, the 25th of November 2020, doi: ARTN 55870610.3389/fbuil.2020.558706.

[4] W. Elhaddad, F. McKenna, M. Rynge, J. Lowe, C. Wang, and A. Zsarnoczay, "NHERI-SimCenter/WorkflowRegionalEarthquake: rWHALE (Version v1.1.0). Zenodo. http://doi.org/10.5281/zenodo.2554610," ed, 2019.

[5] N. SimCenter, *An application framework for regional earthquake simulations*. 2019.

[6] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, "Grid Structural Characteristics as Validation Criteria for Synthetic Networks," (in English), *Ieee Transactions on Power Systems,* vol. 32, no. 4, pp. 3258-3265, Jul 2017, doi: 10.1109/Tpwrs.2016.2616385.

[7] H. Y. Li, J. H. Yeo, A. L. Bornsheuer, and T. J. Overbye, "The Creation and Validation of Load Time Series for Synthetic Electric Power Systems," (in English), *Ieee Transactions on Power Systems,* vol. 36, no. 2, pp. 961-969, Mar 2021, doi: 10.1109/Tpwrs.2020.3018936.

[8] T. Xu, A. B. Birchfield, K. S. Shetye, and T. J. Overbye, "Creation of synthetic electric grid models for transient stability studies," 2017, pp. 1-6.

Appendix

Table A.2 - Nomenclature, Definition, and Simplifications in the Data Engine V1.

Nomenclature	Definition	Application in Data Engine V1
Hazard intensity	A measure of hazard severity often aligns with the design capacity of physical infrastructure.	The hazard intensity of the Data Engine V1 is the 10-m high wind speed, angle at each geographical location, and temporal time stamp during the hurricane evolution. • Precipitation information is not considered. • Dynamic wind loading is not considered. • Wind profile variations are not considered.
Power infrastructure	Physical infrastructure in a power system includes steel transmission towers, concrete and wood poles, physical lines between transmission towers and distribution poles, generators, and substations.	The transmission lines are responsive to the wind speed during hurricanes and are found to cause power outages to the network. Therefore, the power infrastructure primarily designed and mapped in the Data Engine V1 is the transmission tower-line system. • Only lattice steel transmission tower-line infrastructure is considered.
Structural limit state model	A model simulates the physical infrastructure's limit state given different static loading inputs.	The structural model is implicitly embedded in the Data Engine V1 fragility model. • Input to the structural damage model is the hazard intensity along the structural height (wind profile).
Limit State	A condition that the designed structure no longer fulfills the relevant design criteria.	The structural limit state includes the ultimate limit state and serviceability limit state in general. • Only the collapse state of the transmission tower is considered to generate the fragility model.
Fragility model	A fragility model predicts the probability that the infrastructure exceeds a certain performance limit state.	The fragility model in Data Engine V1 predicts the collapse probability of each transmission tower with a given hazard intensity. • The top displacement of the transmission tower determines the failure criterion. • The transmission tower-line coupling phenomenon is considered. • Only collapse condition is considered.
Transmission tower failure probability	The failure probability is calculated as the number of failed transmission towers over the total number of cases at each hazard intensity.	The total number of transmission tower failure probability calculation cases is 20 in the Data Engine V1. The failure probability resolution is 5%.
Power system model	A power system model is a mathematical representation of the electrical network.	The transmission system is the power system that connects generators to consumers. The hazard impacts the transmission system and based on the damage inflicted on the system, the system's load shedding is determined.
Bus	Also referred to as a “node,” a “station,” or a “substation.” A common connection point for two or more electrical components, such as a transformer or a generator.	Elements in the power system model
Generator	A machine transforms mechanical power into electrical power.	Elements in the power system model
Power-Line	Also referred to as “transmission lines,” that carry large amounts of	Elements in the power system model

	electricity from centralized generation plants to lower voltage distribution lines and substations that supply local areas.	
Islanding	The process whereby a Microgrid separates itself electrically from the main power grid and operates independently, using its internal power source(s), may later rejoin the main grid.	Due to the impact of the hazard on the power system, transmission lines can get damaged. The failure of transmission lines can separate the power system into two or more segments. These segments are referred to as islands.
Load	The amount of electricity used at any given moment by a single customer or a group of customers.	In Data Engine V1, the load is all the customers connected to one bus.
Load Shedding	Intentionally turning off power to a customer or group of customers, usually for reliability reasons such as to avoid a blackout or equipment damage.	As a result of the hazard impacting power system and transmission line failures, load shedding occurs. Load shedding is an output of Data Engine V1.