

# Extracting Resilience Events from Utility Outage Data Based on Overlapping Times and Locations

Arslan Ahmad, Ian Dobson  
Iowa State University

**Abstract**—To study resilience with real data, it is necessary to group the individual outages recorded by utilities into events in which the outages bunch up and overlap due to extreme weather. We show how to automatically group utility outage data into resilience events based on their time and location. The methods work with both detailed utility outage data and EAGLE-I data.

## I. INTRODUCTION

Resilience focuses on events in which a group of outages bunch up and overlap due to extreme weather. It is therefore foundational for methods driven by utility outage data to be able to group the individual outages into resilience events. For example, once the events are obtained, it is straightforward to compute the metrics and statistics describing the events [1], [2], [3], [4], [5]. Since the outage data is extensive, any systematic analysis requires the grouping of outages into events to be done automatically by an algorithm.

Using outage timing to group the outages into events has been shown to be quite effective. In particular, an automatic algorithm based on timing can identify large events that correspond well with extreme weather. However, the events obtained only by timing can include some outages that are geographically far enough away from the main group of outages that they are clearly unrelated to the main group, and it is very desirable to exclude these outliers. Therefore, in this letter, we improve and extend grouping based only on time to grouping based on both time and location. The principle of overlapping for grouping outages into events is also clarified. Discriminating events with both time and location is particularly useful when processing utility data over a wide region.

The utility outage data considered either includes the details of each outage (start time, restore time, location [1], [2]) or, in the case of data scraped from the web, such as EAGLE-I [6], the number of customers simultaneously out in each county at 15-minute intervals.

This letter develops methods that use only *utility outage data* to automatically group outages into events using *outage location as well as time*, so that further event-based quantitative resilience analysis can be done.

Previous work groups outages into events only based on time. Carrington [2] tracks the cumulative number of distribution outages over time and defines the start and end of an event when the cumulative number of outages passes and returns to a threshold number of outages. For a threshold value of zero, as assumed in [2], this accumulation of outages until

A. Ahmad and I. Dobson are with the Department of Electrical and Computer Engineering, Iowa State University, Ames Iowa USA; email: dobson@iastate.edu. Support from NSF grants 2153163 and 2429602, Argonne National Laboratory, and PSerc project S110 is gratefully acknowledged.

all are restored is equivalent to the events based on time in subsection II-A with infinite  $t_{\max}$ . Abdelmalak [4] uses a cumulative number of customers passing and returning to a threshold of 5% of a county's customers to extract events from EAGLE-I data. Papic [1] groups transmission outages into events by requiring the outages to overlap in time and for successive outages to start within 2 minutes. Ekisheva [3] groups transmission outages into events based on time according to subsection II-A with  $t_{\max}=1$  hour but also including successive outages starting within 5 minutes and excluding some repeats of momentary outages. NERC uses this outage grouping to find large transmission resilience events in its annual State of Reliability reports. All of these variations of grouping by time can be extended to also grouping by location using the cylinder concept of subsection II-C.

Another popular approach to finding the outages in events first uses weather data to find time intervals in which the weather is severe enough and then finds the outages that occurred in the affected area during that time interval [5], [7]. This has been applied to analyze specific hurricanes or other events with winds exceeding a given threshold. It is difficult to systematically define the edges of weather events, especially for the more moderate events.

Instead of defining events by first using the weather, our methods process the time and location of outages themselves to group them into events. By checking the weather during these events, our methods are confirmed to produce the larger events that are associated with a range of extreme weather events, such as hurricanes and storms<sup>1</sup>. This has several advantages: The events due to several causes and types of weather are systematically extracted, processing requires only the outage data set, and events of a range of sizes are extracted.

## II. EVENTS

### A. Events based on time

Each outage occurs over a time interval  $[o, r]$  where  $o$  is the outage start time and  $r$  is the outage restore time. Two outages overlap in time if their time intervals overlap; that is, if one of the outages starts before the other outage is restored. Then events are a maximal group of outages that are connected together by overlapping in time. That is, two outages are in the same event if they are connected by a series of overlapping outages<sup>2</sup>. For example, in Fig. 1A, outage 1 does not overlap

<sup>1</sup>We have not established that wildfires are addressed by our methods since they move in time and location differently.

<sup>2</sup>In mathematical terms, outages overlapping is an equivalence relation, so that it partitions the outages into equivalence classes that are the events.

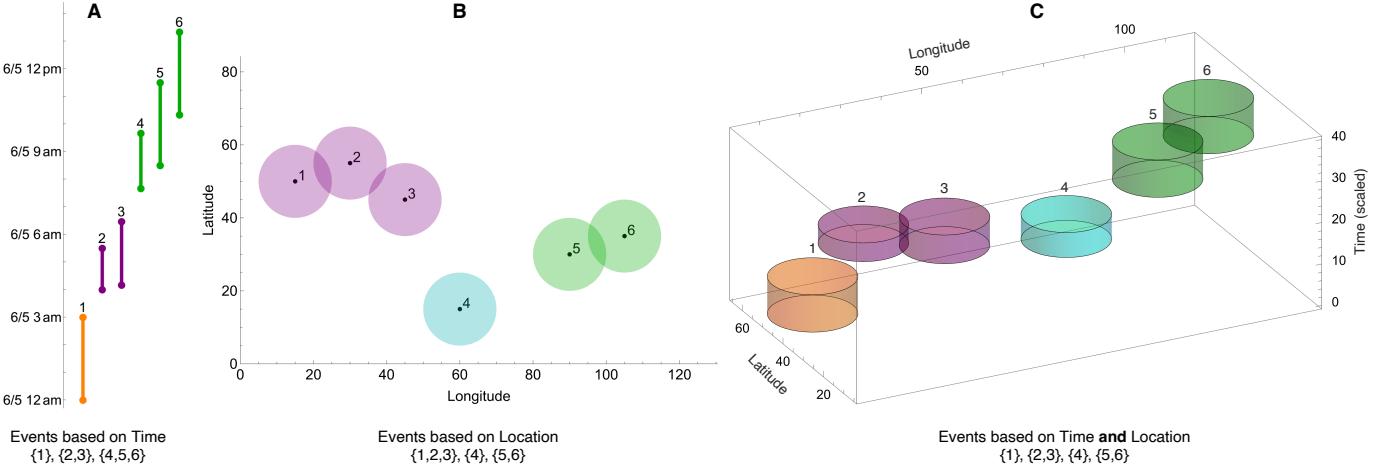


Fig. 1. Example of six outages resulting in different events in time, location, and time and location together. The distance threshold is  $d = 20$  km.

with any other outages, so it forms an event with only one outage. Outages 2 and 3 overlap but do not overlap with any other outages so they form an event with two outages. Outages 4 and 5 overlap, and outages 5 and 6 overlap, due to which outages 4,5,6 are in the same event (outages 4 and 6 need not overlap).

The definition of events in time is often slightly more elaborate than explained so far in order to limit the effect of outages with very long restoration times. Instead of associating each outage with the time interval  $[o, r]$ , each outage is associated with the time interval  $[o, \min\{r, o + t_{\max}\}]$ , limiting the time interval to a maximum duration of  $t_{\max}$  hours. These limited time intervals are then used to define the events. It is imperative to note that while the time limitation is used to define the overlapping outages in an event, the full outage duration is preserved when the event is analyzed.

Another way to characterize events uses the performance curve  $P(t)$  that tracks the cumulative number of unrestored outages [2]. An event begins with its first outage and then a series of outages with overlapping time intervals until all the outages are restored. Therefore,  $P(t)$  starts at zero outages, is a nonzero number of outages during the event, and first returns to zero outages at the end of the event. Indeed, the number of unrestored outages  $P(t)$  is the number of overlaps at time  $t$ .

The EAGLE-I utility data does not describe individual outages, but it does yield a performance curve  $P^{\text{cust}}(t)$  tracking the number of unrestored customers in a county at 15-minute samples of time  $t_1, t_2, \dots$ . To define county events based on time in the EAGLE-I data for Massachusetts, we set a threshold of 30 customers<sup>3</sup>. A county event starts when  $P^{\text{cust}}(t)$  exceeds 30 customers and ends just before  $P^{\text{cust}}(t)$  first drops below 30 customers. That is, an EAGLE-I county event occurs at a maximal series of consecutive samples  $t_m, t_{m+1}, \dots, t_{m+n}$  for which  $P^{\text{cust}}(t_k) > 30$  for  $k = m, m+1, \dots, m+n$ . An increment in outages occurs at  $t_k$  when  $P^{\text{cust}}(t_k) > P^{\text{cust}}(t_{k-1})$ . The county event outage duration is the time samples between and including the first outage increment at the start of the county event and the last outage increment in the county event.

<sup>3</sup>The threshold is needed to distinguish the start and end of events while small number of customers are persistently outaged throughout the county, or while the recorded data is spuriously stuck at a small number of customers out.

Then two county events overlap in time when their county outage durations have time samples in common.

### B. Events based on location

Outages can be grouped together into events based on the distances between them. Two outages are close if their distance is less than or equal to a threshold distance  $d$ . Equivalently, associate each outage with a hypothetical disk of radius  $d/2$  centered at the outage location so that the outages are close if their disks overlap. Then events are a maximal group of outages that are connected together by their disks overlapping. That is, two outages are in the same event if they are connected by a series of overlapping outage disks. In Fig. 1B, the 4 events based on location are {outages 1,2,3}, {outage 4}, and {outages 5,6}.

The grouping based on location works with several specifications of distance. If location is given in the data as latitude and longitude, Euclidean or haversine distance can be used. For the EAGLE-I data, only the county location is given. We regard counties as nodes on a graph and join together with a graph edge any two counties with a sufficiently long common border. Then neighboring counties have a graph distance of 1, and the threshold  $d = 1$  regards a county and its neighbors as close.

### C. Events based on time and location

We account for both time and location by defining an event as a group of outages that both *overlap in time* and are *close in distance*. We consider the location and time information of outages as a 3-dimensional space, where the x and y axes represent location, and the vertical z-axis represents time. In the 3-dimensional space, each outage is associated with a cylinder with base radius  $d/2$  and a vertical time dimension given by the limited outage duration  $[o, \min\{r, o + t_{\max}\}]$  as shown in Fig. 1C. Two outages are close in time and location if their cylinders overlap in 3 dimensions; that is, both their limited durations and disks overlap. Then, as before, events are a maximal group of outages that are connected together by overlapping. In Fig. 1C, the 4 events based on

time and location are {outage 1}, {outages 2,3}, {outage 4}, and {outages 5,6}.

A straightforward way to implement event extraction from data is to create an undirected graph in which outages are represented as nodes, and outages that overlap in time and location are connected by edges. Then the events are the connected components of the graph<sup>4</sup>. The graph formulation lends itself to optimization for larger datasets using standard techniques such as interval trees, sweep line, k-d trees, and grid hashing. An efficient way to implement event extraction is to first group the outages sorted by time into initial events based on time, and then further partition each initial event according to both time and location.

#### D. Example Results

The time and location method is applied to automatically extract events from detailed publicly available outage data of investor-owned utilities in Massachusetts [8]. Thresholds of  $t_{max} = 3$  hours and  $d = 10$  miles are used. Data from NOAA's storm events database, DOE's OE-417 form, and local weather reports are used to verify whether the automatically extracted events correspond to actual weather events. Table I shows that the top 10 largest events extracted automatically from the outage data correspond nicely to major weather events.

TABLE I  
WEATHER ASSOCIATED WITH AUTOMATICALLY EXTRACTED EVENTS

# Outages	Start Date/Time (LST)	Duration (hr)	Weather Event
7739	2021-10-26 14:44:41	179.9	Oct. '21 Nor'easter
7392	2018-03-02 04:25:50	178.0	March '18 Nor'easter
5301	2018-03-12 05:22:42	178.2	March '18 Nor'easter
4622	2018-03-07 06:25:25	192.9	March '18 Nor'easter
3287	2013-02-08 14:48:00	149.1	Winter Storm Nemo
3248	2017-10-29 19:18:00	114.6	Oct. '17 Nor'easter
3178	2019-10-16 20:03:00	102.9	Coastal Storm
2775	2020-10-07 11:06:29	124.3	Derecho
2773	2020-08-04 13:00:00	129.3	Hurricane Isaias
2065	2020-04-13 09:15:48	67.7	Easter Tornado

Figure 2A shows actual events extracted from the Massachusetts detailed outage data. The 1194 outages shown are grouped into a single event if only time-based grouping is used. However, when both time and location are used, the outages are divided into multiple distinct events, the top three of which are shown. Figure 2B shows a different set of events extracted from the Massachusetts EAGLE-I data. The 8 county events shown are grouped into a single event if only time-based grouping is used, and are grouped into 4 events if both time and location are used.

### III. CONCLUSION

We explain an overlapping outage principle used to group outages into events based on time, and generalize it to group outages into events based on both time and location. This new way to systematically and automatically extract events from utility data is an improved foundation for the quantitative

<sup>4</sup>There is also a concept of connected or path-connected components in topology giving an elegant definition of events: Events are the outages in the connected components of the union of all the outage cylinders.

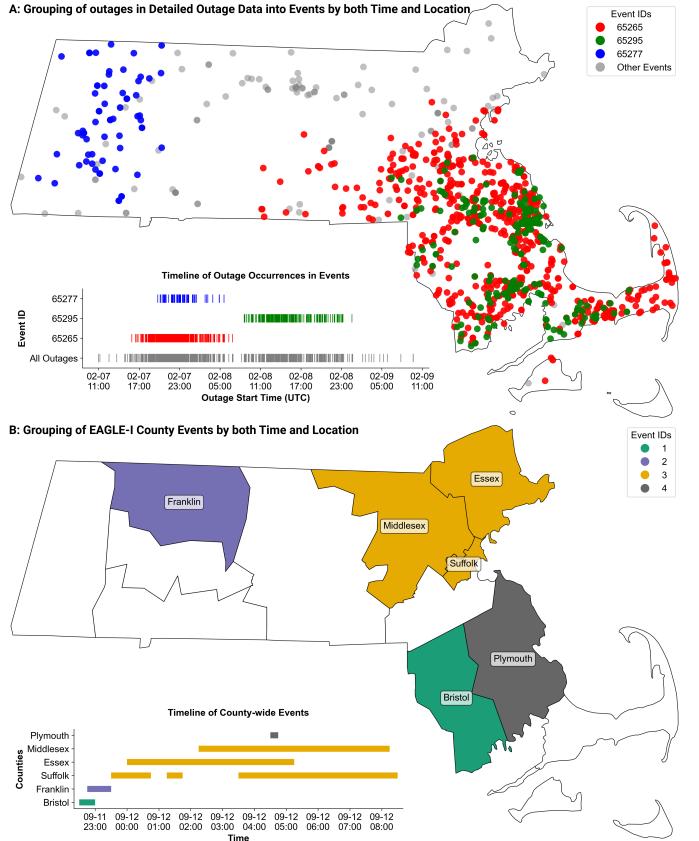


Fig. 2. Two different examples of events extracted by time and location grouping in A detailed outage data, and B EAGLE-I data.

study of widespread resilience events driven by real data. In particular, it can remove unrelated outages far away from weather events in detailed outage data and extend EAGLE-I analyses to widespread events beyond the county level.

### REFERENCES

- [1] M. Papic, S. Ekisheva, and E. Cotilla-Sanchez, "A risk-based approach to assess the operational resilience of transmission grids," *Applied Sciences*, vol. 10, no. 14, p. 4761, 2020.
- [2] N. K. Carrington, I. Dobson, and Z. Wang, "Extracting resilience metrics from distribution utility data using outage and restore process statistics," *IEEE Trans. Power Systems*, vol. 36, no. 6, pp. 5814–5823, 2021.
- [3] S. Ekisheva et al., "Assessing transmission resilience during extreme weather with outage and restore processes," in *Prob. Methods Applied to Power Sys.*, Manchester UK, Jun. 2022.
- [4] M. Abdelmalak et al., "Quantitative resilience-based assessment framework using EAGLE-I power outage data," *IEEE Access*, vol. 11, pp. 7682–7697, 2023.
- [5] S. M. Lee et al., "Quantifying the power system resilience of the US power grid through weather and power outage data mapping," *IEEE Access*, vol. 12, pp. 5237–5255, 2024.
- [6] DOE, *EAGLE-I*, Available: <https://eagle-i.doe.gov/>, Accessed: Jul. 12, 2025.
- [7] S. Dunn et al., "Fragility curves for assessing the resilience of electricity networks constructed from an extensive fault database," *Natural Hazards Rev.*, vol. 19, p. 04017019, 2018.
- [8] Commonwealth of Massachusetts, *Power outages*, Available: <https://www.mass.gov/info-details/power-outages>, Accessed: Feb. 10, 2025.