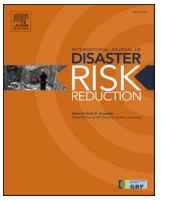
International Journal of Disaster Risk Reduction 125 (2025) 105574

Contents lists available at ScienceDirect

International Journal of Disaster Risk Reduction

journal homepage: www.elsevier.com/locate/ijdrr



Anatomy of a historic blackout: Decoding spatiotemporal dynamics of power outages and disparities during Hurricane Beryl

Xiangpeng Li *, Junwei Ma, Ali Mostafavi

Urban Resilience.AI Lab, Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX, United States

ARTICLE INFO

Keywords:
Power outages
Spatial analysis
Community recovery
Infrastructure resilience
Equity

ABSTRACT

Power outages are a significant consequence of natural hazards, severely disrupting communities' restoration and recovery processes. Despite the increasing frequency and impact of hazard-induced power outages, empirical studies examining their spatial and temporal characteristics across impacted regions remain limited. This dearth of empirical insights inhibits the ability to quantify impacts and examine vulnerability and equity issues for effective resilience investments. This study investigates the spatial patterns and temporal variations in outage duration, intensity, and restoration/recovery following the 2024 Hurricane Beryl in Houston, Texas. This historic blackout caused widespread power disruptions across the Houston metropolitan area, leaving more than 2 million customers without power over several days, resulting in more than 143 million total customer-out hours. By examining the dynamic interplay between outage impact, recovery features, and socioeconomic and infrastructural factors, the analysis identified key determinants contributing to disparities in power outage impacts and recovery efficiency delineated by ZIP Code across Houston. The findings reveal that areas with higher population density and proximity to the hurricane's path experienced more severe initial impacts. Regions with higher median income showed faster recovery, while lower-income areas exhibited prolonged restoration periods, even with favorable infrastructural conditions, suggesting disparities in restoration speed. The study also highlights how urban development features, such as road density and land elevation, explain spatial disparities in power outage impacts and recovery. This research advances the understanding of power outage dynamics in large metropolitan regions through four key contributions: (1) empirical characterization of outages from a historic hurricane, highlighting infrastructure vulnerabilities in a high-density urban context; (2) comprehensive analysis using multiple metrics to capture spatiotemporal dynamics of outages and restoration; (3) leveraging of high-resolution outage data at fine geographic scales and frequent intervals to quantify and reveal previously masked spatial disparities; and (4) systematic examination of socioeconomic, urban development, and environmental factors in shaping disparities in outage impacts and recovery timelines. These findings provide infrastructure managers, operators, utilities, and decision-makers with crucial empirical insights to quantify power outage impacts, justify resilience investments, and address vulnerability and equity issues in the power infrastructure during hazard events.

* Corresponding author.
E-mail address: xplli@tamu.edu (X. Li).

https://doi.org/10.1016/j.ijdrr.2025.105574
Received 18 January 2025; Received in revised form 2 May 2025; Accepted 13 May 2025
Available online 15 May 2025

1. Introduction

Power outages are a major consequence of natural disasters such as floods, particularly in densely populated urban areas where electricity is vital for daily life, emergency services, and critical infrastructure [1–3]. A widespread interruption in electric service, lasting for days or even weeks, can severely disrupt the operation and availability of critical infrastructure, emergency services, and essential daily activities, posing significant challenges to both communities and recovery efforts [4]. As severe weather events become more frequent and intense, assessing power system vulnerabilities is essential for developing effective mitigation and response strategies [5–7]. The evaluation of spatial and temporal dimensions is a crucial component of integrated disaster risk management, yet it has often been overlooked in academic research [8–11]. Identifying the spatial distribution of power outages, the geographic extent of their impacts, and the evolution of risk over time is essential for effective disaster management [12]. Numerous studies have focused on either outage duration or recovery patterns individually, often relying on a single metric for analysis, for example, recovery duration [52], and resilience curves (B. [14]). High-resolution outage data allows researchers to accurately identify several patterns, demonstrating that the spatial and temporal dynamics of power outages and recovery are crucial for scientific research (S. M. [15,16]). Moreover, by capturing variations at a much finer scale, high-resolution data reveals heterogeneity that is often overlooked in publicly reported metrics, providing deeper insights into outage and recovery processes [17]. The cone of uncertainty defines the region around the projected storm path where the storm's center is most likely to travel [18]. While this provides a clear indication of the hurricane's trajectory and directly affected areas, it does not inherently convey the extent of impact or the recovery process. Explorations in other socioeconomic conditions and infrastructure systems lead to differing levels of disruption and recovery across affected regions, highlighting the need for a more comprehensive assessment beyond the storm's path and cone. When combined with socioeconomic and infrastructure data, this understanding becomes particularly valuable for both risk reduction strategies before an event [19] and response and recovery efforts [20]. In this context, identifying disparities in outage durations across different socio demographic groups is essential, as power disruptions often disproportionately impact vulnerable communities [13,21]. Furthermore, research indicates that median income is a significant predictor of power restoration time, reinforcing the existence of socioeconomic disparities in outage recovery [21]. Extreme events can cause severe physical damage to transportation infrastructure, resulting in substantial social disruptions and economic losses [22]. Therefore, incorporating features such as road density is crucial for assessing vulnerability and resilience.

Despite the growing recognition of the importance of empirical studies on power outages, these studies are still rather limited. First, many studies lack access to high-resolution outage data and fail to develop comprehensive metrics for accurately assessing outage and recovery dynamics. Second, while spatial disparities are often a focal point, spatial and temporal dynamics—including recovery speed and its variability across different communities—are frequently overlooked [23]. The interplay between spatial and temporal factors remains underexplored. Finally, there is a lack of systematic analyses that examine the role of socioeconomic, urban development, and environmental factors in shaping disparities in outage impacts and recovery timelines. Accordingly, this study seeks to answer the following research questions: (1) To what extent do power outages, their impacts, and recovery patterns vary temporally across communities during Hurricane Beryl in the Houston metropolitan area? (2) To what extent do the impact and recovery dynamics of power outages vary across different neighborhoods? (3) What was the extent of disparity in the impact and recovery of the power outages across different sub-populations (e.g., income groups)? (4) To what extent do features of neighborhoods, such as development density, tree canopy, elevation, and proximity to the path of hurricane explain the variations in the impact and recovery speed across different neighborhoods? By answering these questions, this study seeks to advance our ability to quantify the impact and recovery of power outage events using multiple outages features and to further our understanding of how power outages manifest across different neighborhoods and socioeconomic groups, and how neighborhood features interact with extreme weather events to shape outage severity and recovery.

To bridge this gap, the present research investigates Hurricane Beyl, which struck the Houston region in July 2024. By examining the multi-metric features of power outage impact, recovery from the original data, along with the socioeconomic and infrastructural variables, the study reveals critical factors contributing to inequities in both the intensity of outages and the speed of restoration across Houston's ZIP codes. The results demonstrate that the most severe initial outages occurred in more densely populated areas and those located nearer to the hurricane's trajectory. Moreover, while higher-income communities experienced faster recovery, lower-income neighborhoods endured extended restoration periods—even when infrastructure conditions were similar—emphasizing disparities in recovery rates. The research also highlights the influence of urban development features, including road network density and terrain elevation, in shaping spatial distribution patterns of outage impacts and restoration efficiency. In addition, this study shows how socioeconomic disparities intersect with environmental vulnerabilities to create distinct recovery trajectories across different areas, revealing that the same physical conditions produce varying outcomes based on available resources and existing infrastructure quality.

2. Study area and data

2.1. Study area

Hurricane Beryl impacted the Houston area in early July 2024, making landfall near Matagorda, Texas, on July 8 as a Category 1 hurricane with sustained winds of approximately 80 mph. The storm brought significant rainfall to the region, with some areas receiving more than a foot of rain, leading to widespread flooding. Wind gusts were notably strong, with certain locations experiencing gusts up to 107 mph, causing major damage to power poles, distribution lines, and regulator banks[24]. The combination of heavy rainfall and high winds resulted in substantial power outages, affecting more than 2.7 million customers in the Houston area [25]. The

storm's impact and the resulting power outages were further exacerbated by subsequent heat advisories, complicating recovery efforts and posing additional risks to the affected population [26]. Fig. 1 illustrates the path and areas directly affected by Hurricane Beryl in the Houston area, Texas. The hurricane's path [27,28] is depicted by the solid red line visible on the western side of the Houston area, and the direction of this hurricane is moving from South to north (bottom to top in Fig. 1). The blue-shaded regions represent the "hurricane cone", indicating the areas that are directly impacted by the hurricane. Nearly half of the county's ZIP Code Tabulation Areas (ZCTAs) fall within the directly affected zone, highlighting the widespread impact of Hurricane Beryl.

2.2. Data

The data used in this study consists of 15-min interval records of the number of customers experiencing power outages at the ZCTA level from June to September 2024, obtained from the Environment for Analysis of Geo-Located Energy Information (EAGLE-I™) platform (U.S. Department of Energy). This platform, developed by Oak Ridge National Laboratory (ORNL), serves as a geographic information system and data visualization tool, providing high-resolution customer outage data.

Based on this dataset, we have developed several outage-related features, including daily total customer-out hours over time, spatial distribution of total customer-out hours, power outage and restore dynamics, outage rate, restore rate, average daily outage hours per customer, median daily customer-out hours, median customer out percentage, disruption duration to 40 % customers out, disruption duration to 20 % customers out, restoration quantile, restore efficiency, net cumulative customer restoration, and recovery duration. We will further introduce each calculation in Section 3. Methods.

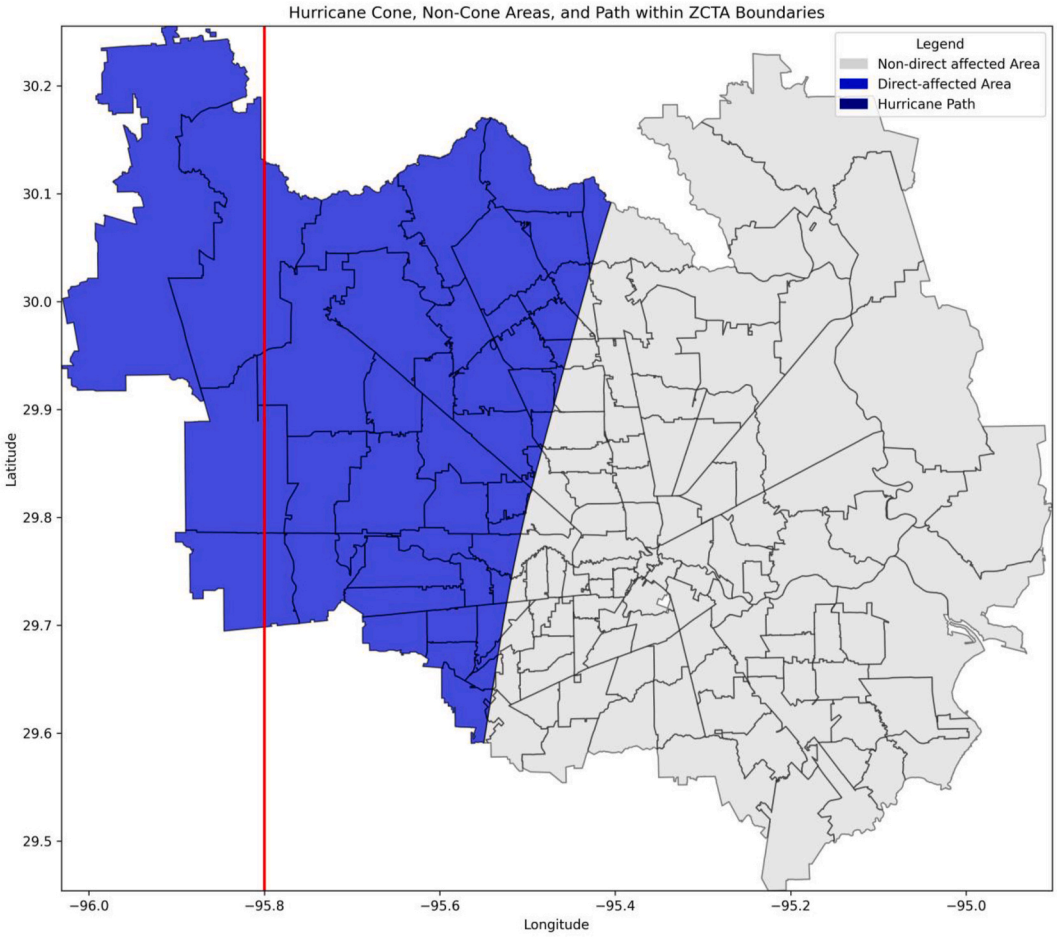


Fig. 1. Path and affected areas of Hurricane Beryl. The solid red line represents the trajectory of Hurricane Beryl as it moved through the region. The blue areas indicate directly affected regions, which are in the Hurricane cone. The gray areas are outside of the hurricane cone but within the Houston area.

3. Methods

3.1. Workflow of the study

The study followed the analyzing steps shown in Fig. 2. Our analysis began with the integration of the datasets, Outage hours were calculated by multiplying the number of affected customers by 0.25 h (15 min), providing an estimate of customer-out hours at each 15-min interval. We incorporated additional socioenvironmental variables, including population density, tree canopy, median income, road density, elevation, distance to the hurricane path, and cone data. Our workflow is structured into three key components: dimensions development, data analysis, and statistical analysis. During dimensions development, we categorized outage impacts and recovery dimensions each into three categories. For impacts, we analyzed severity, scale, and impact duration. For recovery, we developed metrics for restoration rate, resilience, and recovery duration. In the data analysis phase, we employed K-means clustering, box plots, regression analysis, and decision-tree modeling to examine the relationships between socioenvironmental predictors and impact and recovery metrics. Finally, in the statistical analysis phase, we conducted detailed impact and recovery analyses to identify significant predictors of disaster outcomes, as well as disparity analyses to understand inequities in disaster impact and recovery across clusters.

3.2. Spatiotemporal characteristics of power outages

To analyze the outage and recovery dynamics during the hurricane period, we utilized the EAGLE-I™ customer outage data aggregated by date to generate cumulative outage, cumulative restore, and resilience curves. We grouped the data by date and computed the total number of customers experiencing outages for each day, providing the foundation for tracking daily and cumulative trends.

Fig. 3 plots the trends of daily total customer-out hours from July 1, 2024, through September 1, 2024, with a red dashed baseline representing pre-event conditions. The baseline was calculated as the mean daily total customer-out hours between June 1, 2024, and July 1, 2024, capturing the normal, pre-disruption state. Red points mean the daily total custom-out hours is equal to or less than the pre-disruption state. On July 8, the total outage hours showed a sharp increase, coinciding with the impact of Hurricane Beryl. The outage hours peaked shortly after, reflecting the widespread power disruptions caused by the event. Following this peak, a steady decline in outage hours starting from July 10 is observed as recovery efforts progressed. The first instance where the daily outage hours returned to the baseline level occurred on July 27, 2024, indicating an overall recovery to pre-event conditions. The date is three days longer than the restoration date, which is July 24. Subsequent data points show continued stabilization, with some days remaining at or below the baseline, signifying that the recovery phase was nearing completion by the end of the observation period.

Fig. 4 shows the spatial distribution of total customer-out hours ranging from approximately 1 million hours (lightest shades) to more than 7 million hours (darkest shades) across the Houston area during Hurricane Beryl. We aggregated the data by summing the

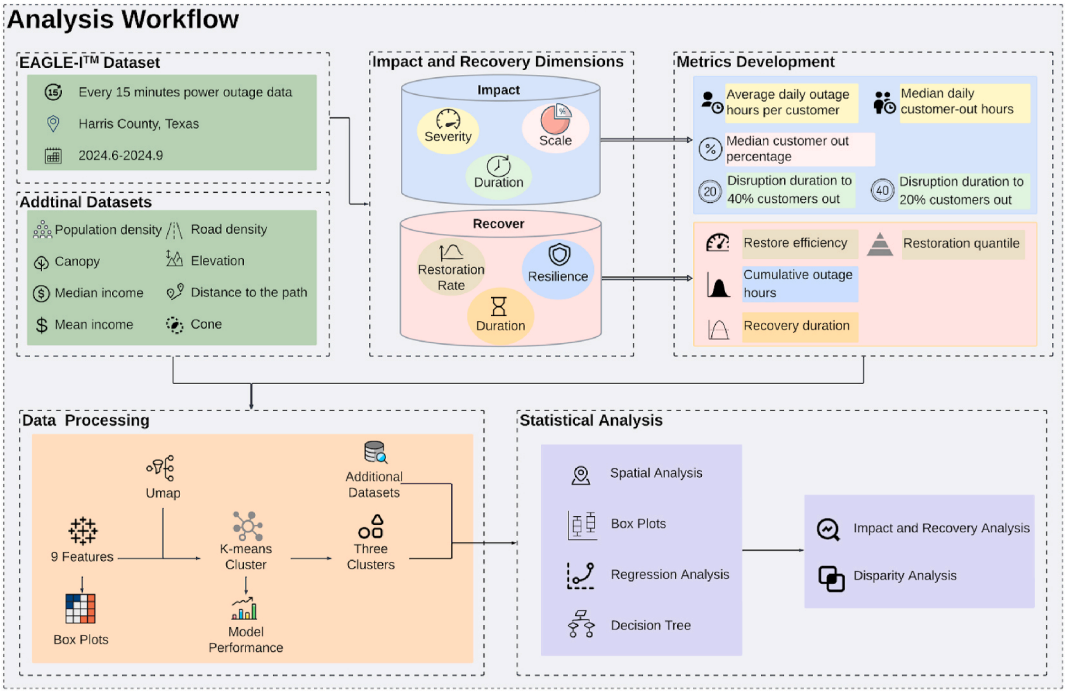


Fig. 2. Overview of the workflow and components.

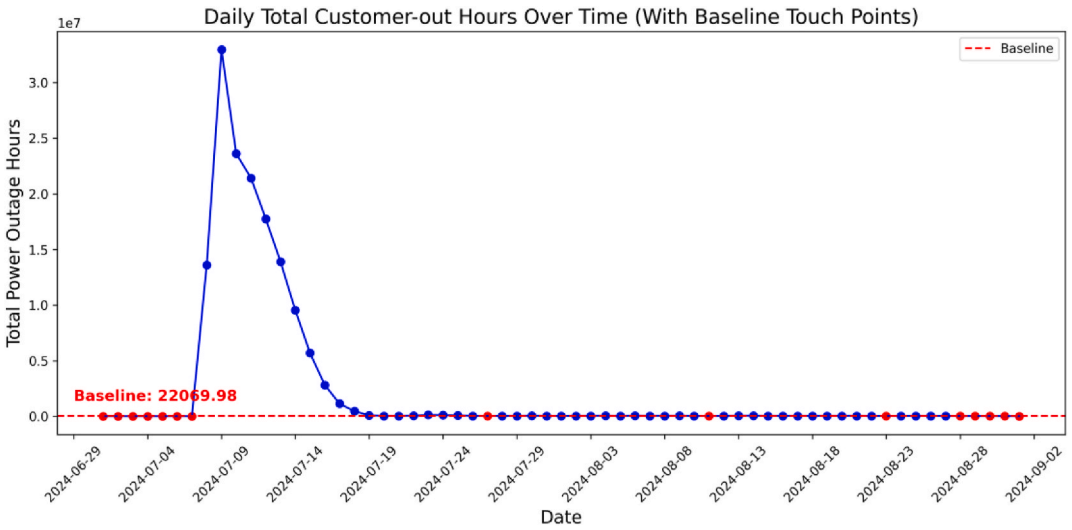


Fig. 3. Daily Total Customer-Out Hours Over Time. The continuous line illustrates the temporal trend of daily total customer-out hours, capturing fluctuations throughout the period. A dashed red line represents the baseline, serving as a reference for pre-event levels. Points at or below the baseline are highlighted in red, indicating recovery to normal levels, while points above the baseline are marked in blue, signifying periods of elevated outage levels.

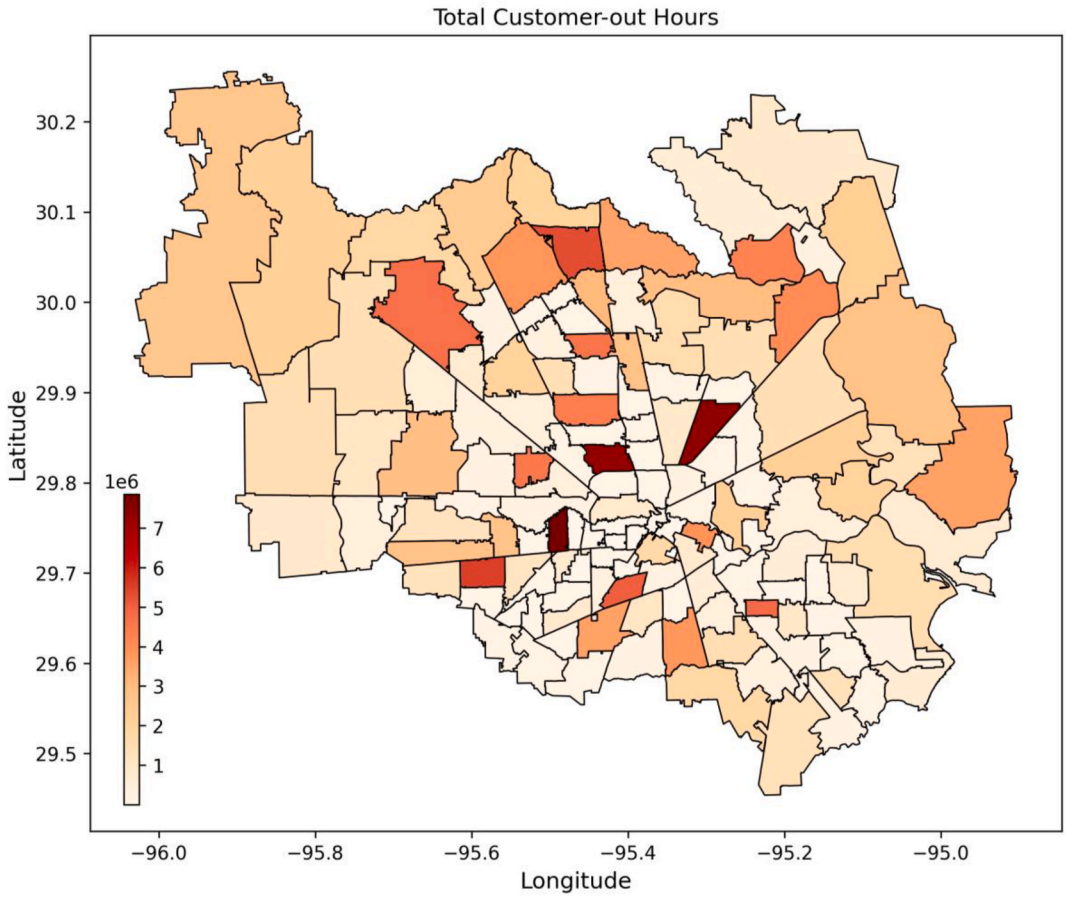


Fig. 4. Spatial Distribution of Total Customer-Out Hours. Darker red areas represent regions with higher total customer-out hours.

"outage hours." Between July 8, 2024, and July 27, 2024, the total customer-outage hours amounted to 143,455,003.25 h. In Fig. 4, areas with the highest total customer-out hours, represented in dark red, are concentrated in the central and northern parts of the region. In contrast, the peripheral regions, particularly in the eastern and western areas, show the lightest shades, indicating fewer customer-out hours due to shorter outage durations or lower population densities.

To reflect the persistent nature of outages, we calculated a modified cumulative outage curve. This curve was generated by determining, for each day, the maximum of the current day's outages and the previous day's cumulative outage value. This approach ensured that the cumulative outage accurately captured the highest sustained outage levels. Next, we calculated daily restoration values as the decrease in the number of outages from the previous day, and we constructed a cumulative restore curve by summing these daily values. To quantify the recovery progress relative to the outages experienced, we defined the resilience curve as the difference between the cumulative restore curve and the cumulative outage curve represented mathematically as

$$C(t) = R(t) - O(t) \quad \text{Eq.1}$$

where *C(t)* is the resilience curve, *R(t)* is the cumulative restore, *O(t)* is the cumulative outage, and *t* represents days.

We then defined the final date as the date when the restore catches the cumulative outage. The outage rate was calculated as the maximum cumulative customer-out divided by the number of days between the start of the outage and the final date. The restore rate was determined as the maximum cumulative customer-out of restore divided by the duration from the first restoration date to the final date. We defined the restore efficiency as the outage rate over the restoration rate.

From Fig. 5, we can see that the outage curve (blue) shows a sharp and rapid increase initially, indicating significant and sudden outages. This curve plateaus at approximately 1.3×10^8 cumulative customers, suggesting the peak outage was reached quickly and did not escalate further beyond this point. In contrast, the restore curve (red) progresses more steadily, indicating a gradual restoration process. The restore curve converges with the outage curve at the end, reaching approximately 1.3×10^8 cumulative restored cumulative customers, verifying that full restoration was achieved. The resilience curve (green) begins with negative values as the outages initially outweigh the restorations. We also calculate the area between the outage and restoration curve, which is also equal to the resilience curve to x-axis. This area is the cumulative customer-out, representing the net cumulative customer-out over time. The final date when the restoration catches the cumulative outage is July 24, 2024. We also plotted the distribution of outage rate and restoration rate (Fig. 6).

Fig. 6 (a) illustrates the spatial distribution of the outage rate, calculated as the cumulative number of customers affected per day during the study period. Areas with the highest outage rates, as indicated by the darkest red shades in Fig. 6(a), are primarily concentrated in the central and southwestern portions of the region, suggesting significant disruptions in power service. In contrast, lower outage rates are observed in peripheral areas, likely due to lower population density and fewer customers affected. On the other hand, the restoration rate distribution in Fig. 6(b), defined as the cumulative number of customers whose power was restored per day in each ZCTA, reveals that regions with the highest recovery rates, shown in dark blue, are predominantly located in the northern and northwestern areas. Some locations that experienced severe outages also exhibit high restoration rates.

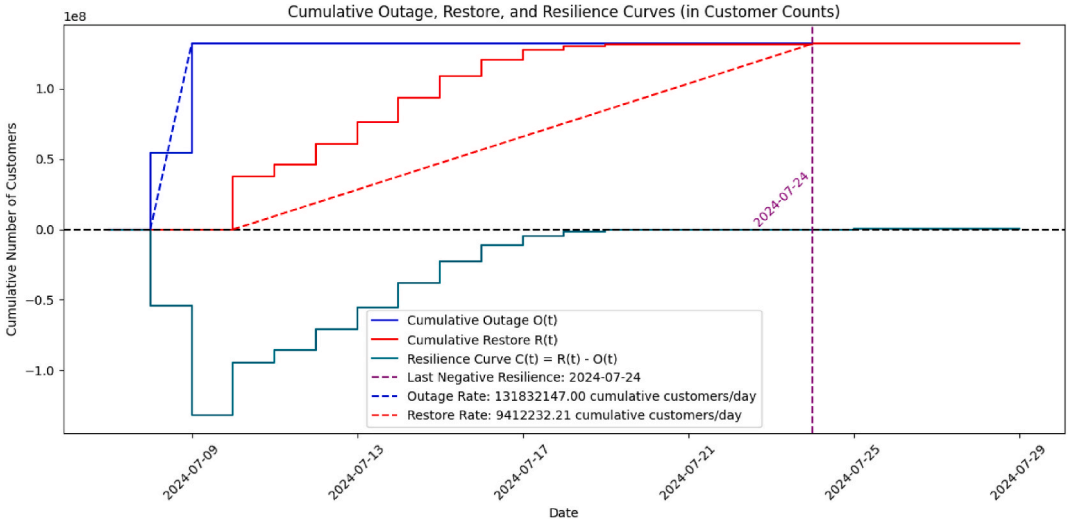


Fig. 5. Overview of the power outage and recovery dynamics. The cumulative outage curve (blue) depicts the total number of cumulative customers affected by outages over time, while the restore curve (red) shows the total cumulative number of customers restored within the same timeframe. The resilience curve (green) highlights net recovery progress, providing a visual representation of the restoration trajectory. Dashed lines indicate calculated outage and restore rates, serving as quantitative benchmarks to assess recovery performance. The vertical purple line marks the point when restoration efforts fully match the total outages, signaling the completion of recovery.

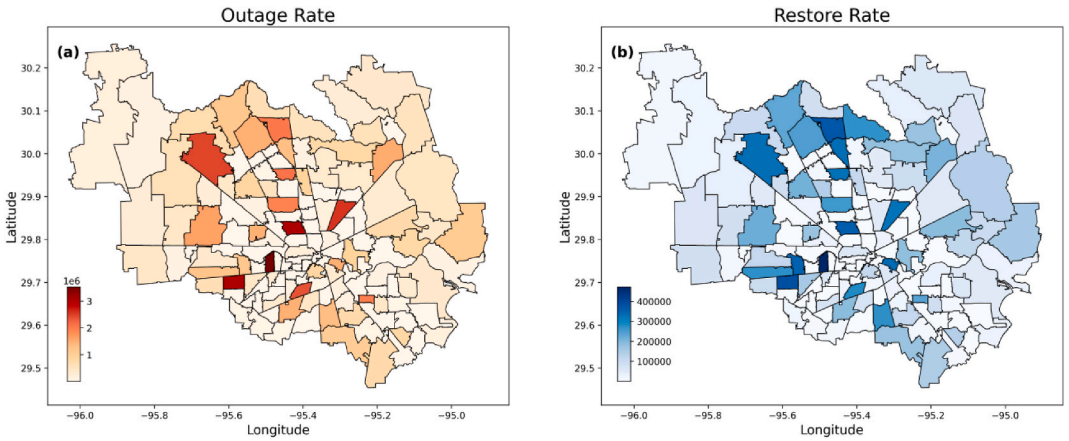


Fig. 6. The outage rate and restore rate in the Houston area, Texas, across ZCTAs. (a) Outage Rate; (b) Restore Rate. Darker red areas represent regions with higher outage rates, and the darker blue areas correspond to higher restore rates.

3.3. Power outage impact and recovery features

We developed nine features from the outage data and divided them into two key dimensions: impact and recovery (see Table 1). The impact dimension is developed from three categories: (1) severity: average daily outage hours per customer, median daily customer-out hours; (2) scale: median customer out percentage; (3) duration: disruption duration to 40 % customers out, and disruption duration to 20 % customers out. The Recovery dimension can also be divided into three categories: (1) restoration rate: restoration quantile; (2) restore efficiency; (3) resilience: net cumulative customer restoration; duration: recovery duration.

3.3.1. Power outage impact features

3.3.1.1. Average daily outage hours per customer. To calculate the average daily outage hours per customer for each ZIP Code (Fig. 7(a)) during the disruption period (July 8 through July 27, 2024), we first aggregated the total customer-out hours, which represents the cumulative duration of outages experienced by all customers in each area. Then, to estimate the number of customers, we identified the maximum number of customers-out in each ZIP Code for each day, capturing the highest count of affected individuals during daily peak outages. Using these metrics, we calculated the average daily outage hours per customer for each ZIP Code by dividing the daily total customer-out hours by the daily maximum number of customers out. Finally, to represent the overall impact during the event, we computed the average values of the feature over disruption days for each ZIP Code. The overall average daily outage per customer was 12.41 h, resulting in a total average outage per customer of 248.2 h over 20 days. Fig. 7(a) displays the spatial distribution of the average daily outage hours per customer. We observed that the average daily outage hours ranged from approximately 9 to 15 h. High

Table 1
Illustration of the power outage impact and recovery features.

Feature Category	Feature Name	Feature Description	Feature Interpretation
Impact	Average daily outage hours per customer	(Daily total customer-outage hours/daily maximum number of customers out per ZCTA)/disruption days per ZCTA	Reflects the severity of the disruption for each customer.
	Median daily customer-out hours	Daily total customer-out hours/disruption days	Captures the daily severity of customer-out hours
	Median customer out percentage	Median of (daily customers out/maximum customers out) over disruption days	Indicates the scale of the disruption, showing how widespread the impact is.
	Disruption duration to 40 % customers out	Number of days from July 8 to when 40 % of customers still power outage	Reflects the timeline of disruption and the subsequent impact duration.
	Disruption duration to 20 % customers out	Number of days from July 8 to when 20 % of customers still has power outage	Reflects the timeline of disruption and the subsequent impact duration.
Recovery	Restoration quantile	Quantile of restoration rates based on the percentage change in cumulative customer-out hours over a rolling seven-day period	Reflects the peak efficiency of restoration efforts.
	Restore efficiency	Restore rate/outage rate	Indicates the balance between recovery speed and outage speed.
	Net cumulative customer restoration	Area between the outage and restoration curve, representing the net cumulative customer-out over time per ZIP Code	Quantifies the gap between outages and recovery, reflecting system resilience.
	Recovery duration	The number of days between the first exceedance of total customer-out hours baseline and the first return to or below the baseline	Reflects the temporal extent of the impact, capturing how long the disruption lasts.

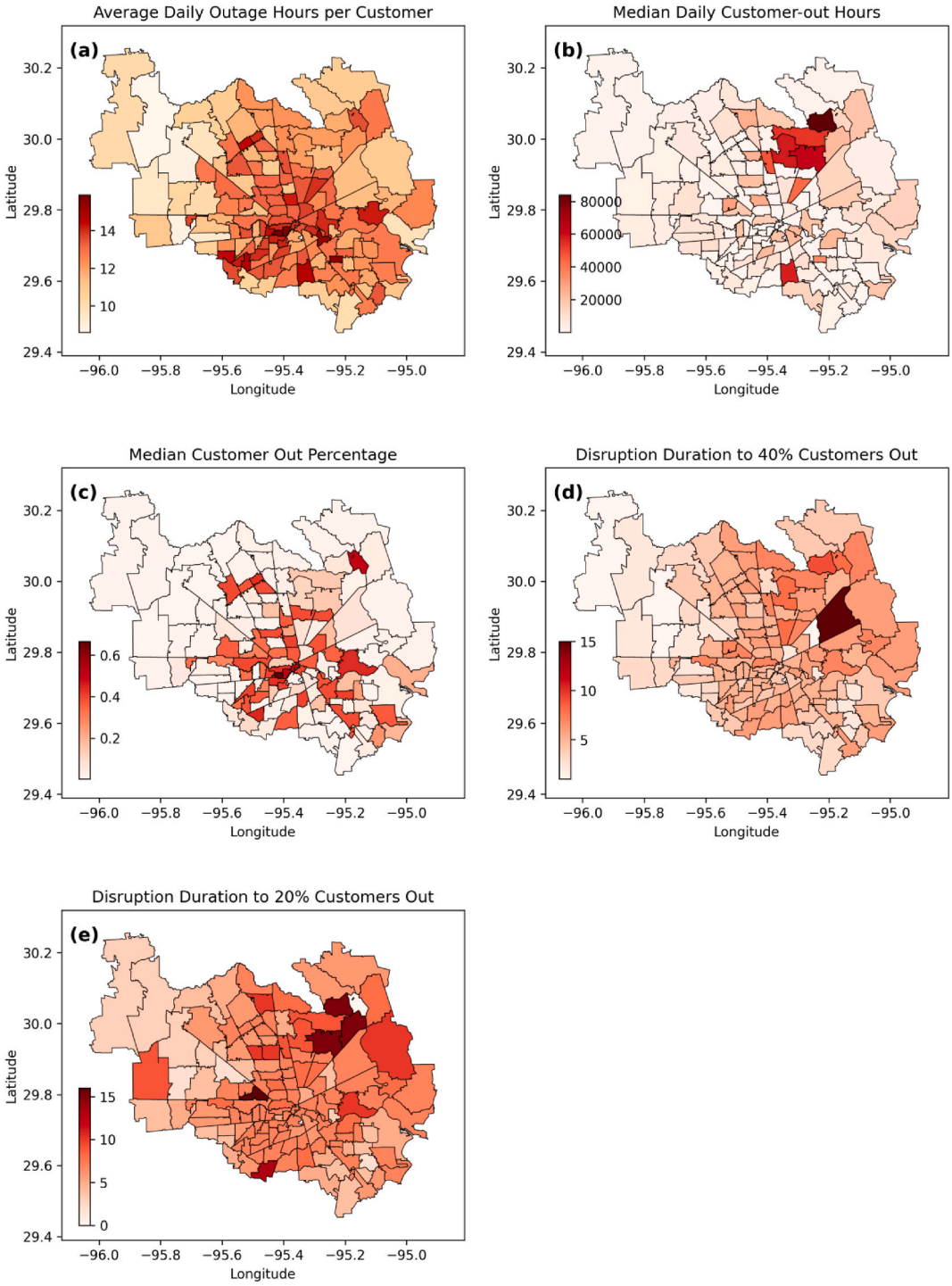


Fig. 7. Spatial distribution of the power outage impact features. (a) Average outage hours per customer; (b) Median daily customer-out hours; (c) Median customer out percentage; (d) Disruption duration to 40 % customers out; (e) Disruption duration to 20 % customers out. Darker red areas indicate ZIP Codes with higher values of the features, signaling areas where outages were more prolonged or severe. Conversely, lighter regions represent areas with less severe disruption.

average outage hours (red) are concentrated primarily in the central and southwestern areas of the region. Specifically, some ZIP Codes in the central areas exhibit values exceeding 14 h, indicating prolonged disruptions for customers in these areas. Conversely, the lighter shades, representing lower average outage hours (closer to 9–11 h), are predominantly distributed along the periphery, particularly in

the northern and eastern parts of the region.

3.3.1.2. Median daily customer-out hours. Median daily customer-out hours represent the median value of daily cumulative outage hours (Fig. 7(b)). First, we calculated the daily total outage hours by summing the outage hours, which represents the cumulative outage duration across all affected customers. Next, we calculated the median daily outage hours for each ZIP Code over the selected period (July 8 through July 27). Fig. 7(b) illustrates the median daily customer-out hours across the Houston area during the disruption period, with a gradient of red shades representing varying outage durations. Darker shades indicate areas with higher median outage hours, while lighter shades represent less severe impacts. The most affected regions are predominantly concentrated in the north-eastern part of the Houston area, where the median daily customer-out hours exceeded 60,000 h, indicating prolonged outages in these areas. In addition, a smaller cluster of high values is observed in the southern region.

3.3.1.3. Median customer-out percentage. For each ZIP Code, the maximum number of customers affected by outages during the disruption period was identified and recorded in the raw dataset. The customer outage percentage (Fig. 7(c)) was calculated for each day as the ratio of customer-out to customer-max. This percentage represents the proportion of customers in each ZIP Code which experienced outages on any specific day and time. To summarize the impact over time, the median customer out percentage for each ZIP Code was computed during the period July 8 through July 27. The maximum daily median percentage of customers out of power was 84.25 %. Fig. 7(c) displays the spatial map of the feature, with a red gradient indicating the severity of outages. Darker shades represent ZIP Codes with higher median outage percentages, while lighter shades indicate less severe impacts. The central part of the county experienced the highest median outage percentages, with some areas reaching as high as 60 %, indicating that a significant portion of customers in these regions was affected during the outage period. Additionally, elevated outage percentages are observed in parts of the southeastern and northeastern regions, reflecting pockets of concentrated disruption. In contrast, the western regions of the county experienced lower outage percentages, as indicated by the lighter red shades.

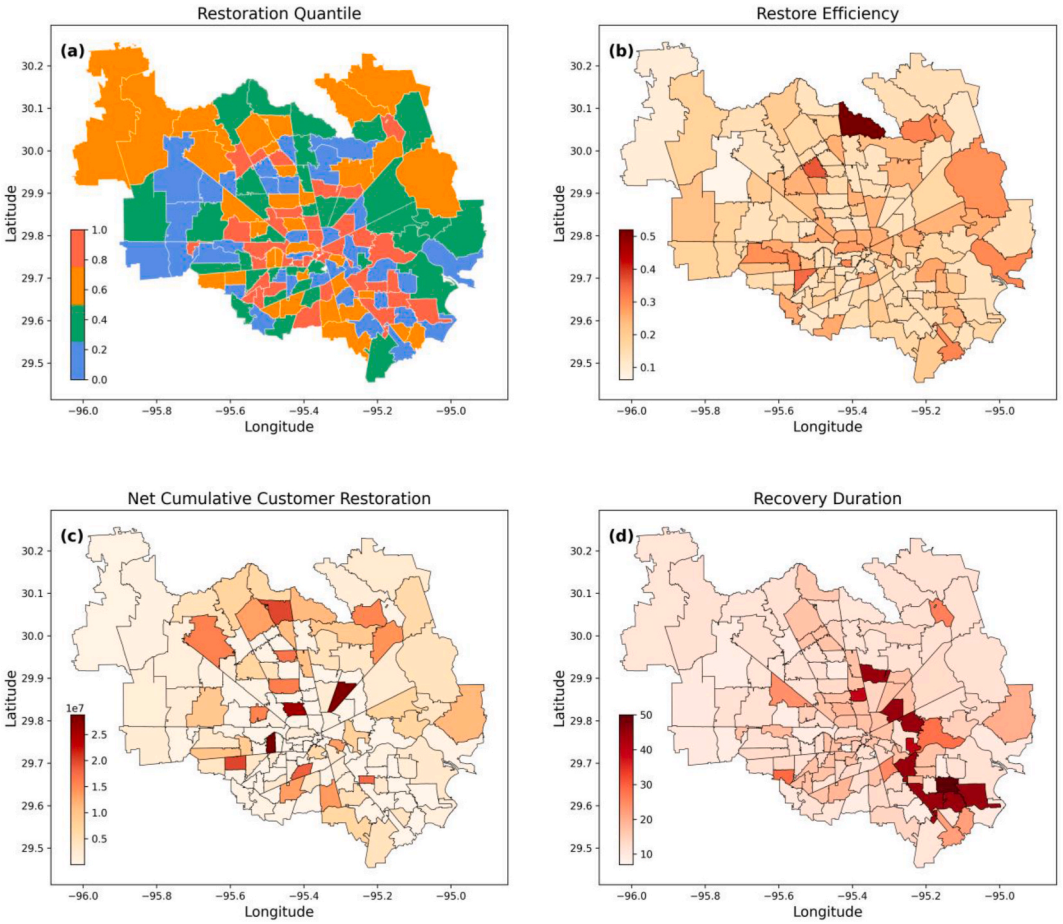


Fig. 8. Spatial distribution of the power outage recovery features. (a) Restoration quantile; (b) Restore efficiency; (c) Net cumulative customer restoration; (d) Recovery duration Darker red areas indicate ZIP Codes with higher values of the features, signaling areas where recovery was slower. Conversely, lighter regions represent areas with faster recovery.

3.3.1.4. Disruption duration to 40 % customers out and disruption duration to 20 % customers out. To calculate the days required to reach specific outage percentages (40 % and 20 % of customers out, Fig. 7(d) and (e)), we focused on the maximum daily customer outage percentages for each ZIP Code during the disruption period. First, the percentage of customers out was calculated daily as the ratio of customers experiencing outages to the maximum number of customers affected. The data was then grouped by ZIP Code and date to determine the maximum outage percentage for each ZIP Code on each day. For each ZIP Code, thresholds of 40 % and 20 % customer outages were defined. To determine the days required to reach these thresholds, we compared the daily outage percentages against the thresholds to identify the day when the percentage was closest to the defined threshold. The time to reach these thresholds was calculated as the difference in days from the baseline date (July 8, 2024), which marks the beginning of the disruption period. Fig. 7(d) and (e) illustrate the number of days required for each ZIP Code to reduce outages to 40 % and 20 % of the customer-out percentage, respectively. The color gradient, ranging from light pink to dark red, represents the time to reach this threshold, with darker shades indicating longer durations. The northeastern region of the Houston area shows the most significant delays, with some areas requiring up to 14 days to reach the 40 % threshold, and 16 days to reach 20 %. In contrast, central and southwestern regions display shorter recovery times, typically under 6–8 days to 40 % and 20 %, as indicated by the lighter shades. While the northeastern region again exhibits the most extended recovery times, the southeastern and southern parts of Houston area also show delays, with several ZIP Codes taking between 8 and 12 days to reach these thresholds.

3.3.2. Power outage recovery features

3.3.2.1. Restoration quantile. To calculate the restoration quantile, we first calculated restoration rates based on the percentage change in cumulative customer-out hours over a rolling seven-day period (Eqs. (2) and (3)). Next, we divided the calculated restoration rates into four quantiles Q1 to Q4), with Q1 representing the fastest restoration rates and Q4 indicating the slowest.

$$S_t = \sum_{i=t-6}^{t} H_i \quad \text{Eq. 2}$$

where S_t represents seven-day rolling sum of customer-out hours on day t, and H_i represents daily customer-out hours on day i (from t-6 to t).

$$R_t = \frac{S_t - S_{t-1}}{S_{t-1}} \times 100\% \quad \text{Eq. 3}$$

where R_t represents restoration rate (percentage change) on day t, S_t represents seven-day rolling sum of outage hours on day t, and S_{t-1} represents a seven-day rolling sum of outage hours on day *t-1*.

The restoration rate quantile map (Fig. 8(a)) displays the spatial distribution of restoration rates across the Houston area: blue (Q1, fastest restoration), green (Q2), orange (Q3), and red (Q4, slowest restoration). Areas in the central-eastern and peripheral parts of the county generally fell into Q1 and Q2, indicating faster restoration rates. In contrast, several ZIP Codes in the northeastern and central regions are categorized as Q3 and Q4, reflecting slower recovery rates.

3.3.2.2. Restore efficiency. To compute the restore efficiency (restore rate/outage rate), we analyzed customer outage and restoration data over the period from July 8 through July 30, 2024, focusing on the interplay between outage severity and recovery efforts for each ZIP Code. The restore efficiency was derived as the ratio of the restore rate to the outage rate, which will mitigate the effect of the varying customer numbers. This metric quantifies the balance between the speed of recovery efforts and the severity of the outage, with higher values indicating faster recovery relative to the outage impact. Fig. 8(b) displays the restore efficiency for each ZIP Code in the Houston area. Darker red represents higher restore efficiency, signifying areas where restoration efforts outpaced outage severity. In contrast, lighter shades indicate lower restore efficiency, reflecting slower recovery relative to the magnitude of the outages. From the map, we observe that the northeastern region exhibits the highest restore efficiency, suggesting efficient recovery efforts despite significant outages. Conversely, several ZIP Codes in the western and southeastern regions display lower restore efficiency, indicating slower recovery relative to outage severity.

3.3.2.3. Net cumulative customer restoration. The net cumulative customer restoration was determined by calculating the area between the outage and restoration curve, which is also the area of the resilience curve. Higher cumulative customer-out restoration indicates slower recovery or less effective restoration efforts relative to the extent of the outage. The cumulative customer-out map (Fig. 8 (c)) provides a spatial overview of the feature; the darker red indicates higher cumulative customer-out, corresponding to larger gaps or delays in recovery. The central and southeastern regions exhibit the highest cumulative customer-out. In contrast, ZIP Codes in the western regions display smaller cumulative customer-out.

3.3.2.4. Recovery duration. We calculated recovery duration to measure the time required for fully recovery to the pre-event status, which is more realistic for high-resolution power outage data [29]. The first step involved establishing a baseline outage level for each ZIP Code. This baseline was defined as the average daily outage hours during a pre-event period from June 1 to July 1, 2024, representing normal operating conditions before the disruption. By setting this baseline, we could identify deviations in outage patterns caused by the event for each area. We examined the daily outage hours for each ZIP Code for the disruption period (July 8 through

September 1, 2024). The outage start date was identified as the first day when outage hours exceeded the baseline; the recovery date was determined as the first subsequent day when outage hours returned to or fell below the baseline. The recovery duration was then calculated as the number of days between the outage start date and the recovery date. Fig. 8(d) illustrates the spatial distribution of recovery duration across the study area. The lighter shades indicate shorter recovery durations where power restoration was achieved more quickly. These efficiently restored areas are predominantly located in the northeastern and northwestern parts of the study area. In contrast, darker red shades represent longer recovery durations, highlighting areas where power outages persisted for extended periods. These prolonged outage regions are concentrated in the central and southern portions, where recovery durations exceed 40 days.

A bar chart (Fig. 9) illustrates the distribution of recovery completion dates for each ZIP Code in 2024. The x-axis displays the recovery dates from July through August, while the y-axis shows the number of ZIP Codes that achieved recovery on each date. We can see from the blue bars, the majority of recoveries occurred between July 19 and July 25, 2024, with peak activity observed on July 20 (32 recoveries) and July 25 (33 recoveries). Recovery activity decreased substantially after July 25, and it continued to August 27.

3.4. K-means clustering

To analyze the spatial disparities in power outage impacts and recovery, we adopted K-means clustering to group ZIP Codes based on the features. Before conducting the clustering analysis, we performed a Pearson and Spearman correlation analysis to examine the relationships between features and ensure minimal multicollinearity among them. Correlation coefficients range from -1 to 1, with values approaching 1 indicating strong positive correlations, values near -1 indicating strong negative correlations, and values close to 0 indicating weak or negligible correlations [30]. Fig. 10(a) illustrates that all Pearson correlation coefficients are below 0.7, confirming that the features are suitable for clustering analysis. Fig. 10(b) presents the Spearman correlations, with most values remaining below 0.8, except for the correlation between Median daily customer-out hours and Net cumulative customer restoration ($r = 0.6$, $\rho = 0.9$). This suggests a strong positive monotonic relationship between the variables based on their ranks (as indicated by the high Spearman value), while the linear relationship between them is relatively weaker (reflected in the lower Pearson value) [31], suggesting that it still good for cluster analysis.

We then applied Uniform Manifold Approximation and Projection (UMAP) to reduce the high-dimensional feature space to two dimensions, while preserving the underlying structure and relationships within the data [32]. UMAP is a nonlinear dimensionality reduction technique commonly used for data visualization and as a preprocessing step for machine learning tasks such as clustering [33]. After transforming the data using UMAP, we applied the K-means algorithm to identify distinct clusters (Eq. (4)). K-means iteratively assigned each data point to the nearest cluster centroid while minimizing the variance within clusters and maximizing the variance between clusters [34]. The final cluster assignments were added as a categorical variable, enabling us to incorporate the clustering results into subsequent analyses.

$$J = \sum_{i=1}^{K} \sum_{\chi \epsilon C_i} ||x - \mu_i||^2 \quad \text{Eq. 4}$$

where K represents the number of clusters, C_i is the set of points in cluster, x is an individual data point, μ_i denotes the centroid of

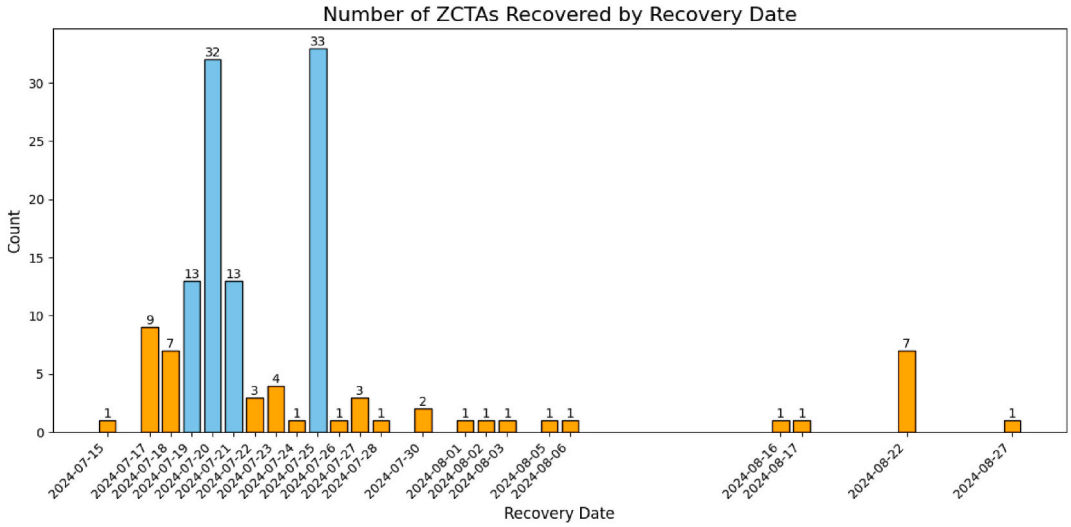


Fig. 9. Number of ZCTAs recovered by recovery date. Each bar corresponds to a specific date. The height of the bar reflects the number of recoveries on that date. The bars are color-coded to differentiate recovery levels: blue bars represent dates with more than ten recoveries, indicating significant recovery activity, while orange bars denote dates with ten or fewer recoveries.

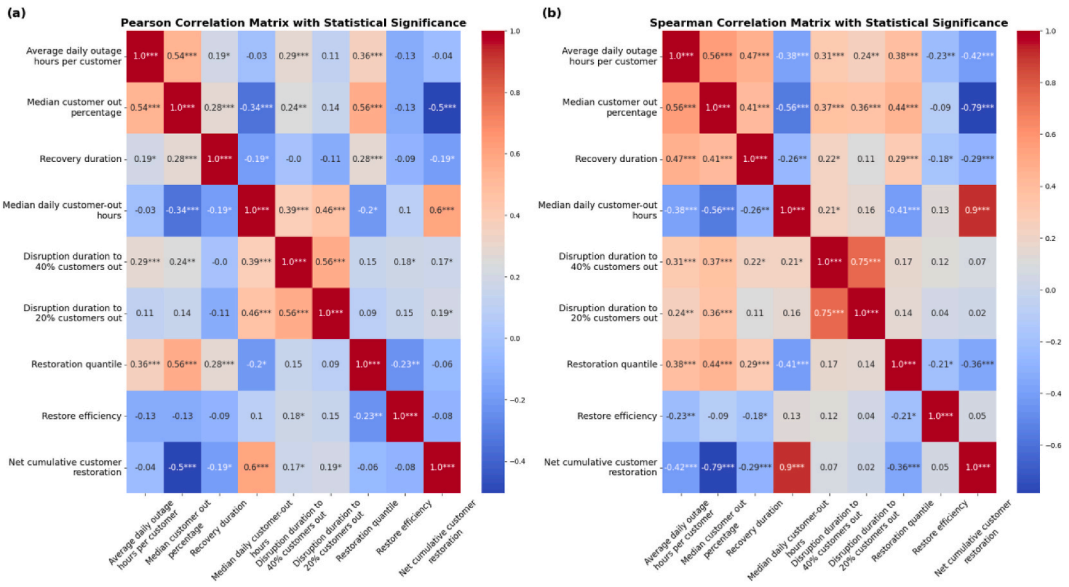


Fig. 10. Pearson correlation matrix between power outage and recovery features. The number within the cell indicates the correlation coefficient. $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$.

cluster, $||x - \mu_i||^2$ is the squared Euclidean distance between a data point and its cluster centroid.

3.5. Socioeconomic and infrastructure features

In order to reveal the underlying factors that shape spatial disparity in power outage impacts and recovery, we analyzed key factors that shape the extent of vulnerability to power outages. Based on the review of the existing literature, we selected the following factors: population density, canopy cover, median income, road density, elevation, proximity to the hurricane path, and whether the region lies within the hurricane cone (Table 2).

Population density, defined as the total population per unit area, was derived from census data (U.S. Census Bureau, 2020). Tree canopy coverage, represented as the mean tree canopy percentage within each ZCTA, was obtained from the Multi-Resolution Land Characteristics (MRLC) Consortium dataset (U.S. Geological Survey, 2025). Sociodemographic indicators, such as median household income, were also sourced from census data (U.S. Census Bureau, 2022), providing insights into the economic conditions of the regions analyzed. Road density, calculated as the total length of roads per ZCTA within each ZCTA, was similarly derived from census data (U.S. Census Bureau, 2024), reflecting infrastructure characteristics. Elevation data, specifically the median elevation within each ZCTA, was sourced from the SRTM 30m Global 1 arc second V003 dataset [39], offering a topographic perspective on the regions studied. The cone and distance-to-path was determined using the data from National Hurricane Center and KMZ viewer [27,28], which allows us to view the directly affected areas and the path of the hurricane.

These factors capture a distinct dimension of socioeconomic and infrastructural impact on the hurricane impact and recovery efforts, making them invaluable for explaining the spatial variation in both the extent of power outages and the speed of restoration during hurricanes. Higher population densities not only lead to more stress on aging infrastructure and exacerbate outage impacts but also may warrant prioritization in restoration [40]. Proximity to the hurricane's path is a key factor in determining exposure to damaging winds, rainfall, and storm surge—thus intensifying the likelihood of power loss [41]. Socioeconomic context, represented by

Table 2
Socioeconomic and infrastructural features used in this study.

Factor name	Definition	Sources
Population density	Total population over area	Census Data Reference [35]
Canopy	Mean tree canopy percentage in each ZCTA	Multi-resolution Land Characteristics (MRLC) Consortium [36]
Median income	Median household income	Census data Reference [37]
Road density	Road density in each ZCTA	Census data Reference [38]
Elevation	Median elevation in each ZCTA	SRTM 30m Global 1 arc second V003 [39]
Cone	The ZCTA under the direct affected area	KMZ viewer [27,28]
Distance to path	Distance to the hurricane path	KMZ viewer [27,28]

median income, further shapes the resources available to communities for infrastructure maintenance and repairs, as well as their ability to advocate for faster service restoration [42]. Meanwhile, mean canopy cover can increase the risk of trees and branches damaging power lines, and prolonging outage durations [43]. Road density is critical for enabling or hindering repair crews' access to damaged sites, influencing the speed at which power can be restored. Also, road network density can capture the complexity of the topology of power grid networks as distribution lines are mainly built along roads [44]. Finally, mean elevation not only captures susceptibility to flooding and storm surge in lower-lying areas but also reflects the potential for stronger wind forces at higher elevations [45,46].

4. Results

Three clusters were finally generated with a silhouette score of approximately 0.574. To validate the robustness of the clustering, we conducted ANOVA (analysis of variance) tests across the three clusters. ANOVA results demonstrated statistically significant differences (p-value <0.05) across clusters, which confirmed that the clusters captured meaningful and distinct groupings in the data. Fig. 11 illustrates the spatial distribution of three distinct clusters identified through our analysis, with each ZIP Code color-coded according to its cluster assignment. We also created boxplots in Fig. 12 to compare key impact and recovery features across the three identified clusters. Each cluster exhibits distinct characteristics that provide insights into the varying patterns of power outage impact and recovery processes.

Cluster 0: high-impact outages with slow and long recovery

Cluster 0 (blue) is predominantly concentrated in the central and southern portions of the study area and demonstrates the most severe outage impacts. While the median daily customer-out hours are lower than those in other clusters, which may be because of less number of affected customers, this cluster records the highest average daily outage hours per customer, approaching 13.5 h, and exhibits the greatest median customer-out percentage, indicating a high proportion of affected customers. Additionally, the disruption duration is prolonged, as evidenced by the extended time required for outages to drop below both 40 % and 20 % thresholds. The recovery process shows moderate delays, reflected in elevated restoration quantiles and lowest restore-to-outage restore efficiency compared to other clusters. Despite having a relatively small net cumulative customer restoration, which may indicate small affected customers, the median value of recovery duration exceeds that of Cluster 1 and Cluster 2.

Cluster 1: moderate impact with quick and short recovery

Cluster 1 (orange), primarily occupying the eastern section of the Houston area, experiences moderate outage impacts with

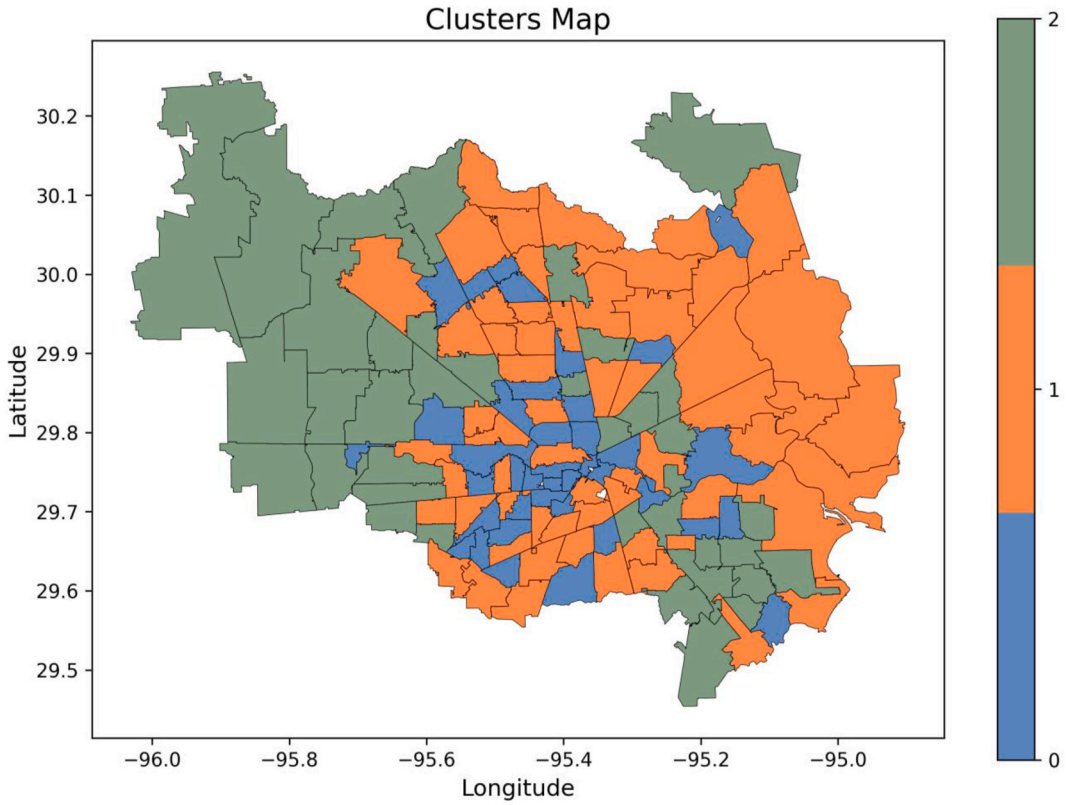


Fig. 11. Spatial distribution of the three clusters regarding power outage impact and recovery in the Houston area. The three clusters are represented by different colors.

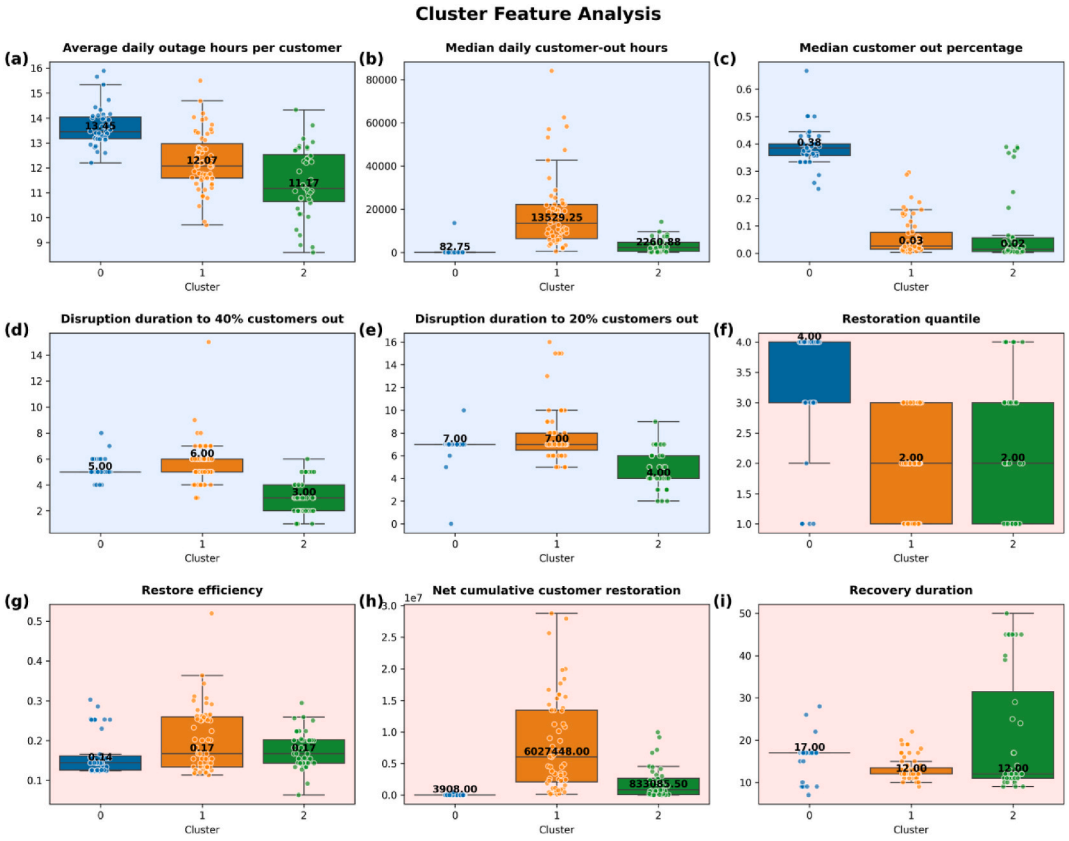


Fig. 12. Box plots of the outage impact and recovery features Values across Clusters. (a) Average daily outage hours per customer; (b) Median daily customer-out hours; (c) Median customer out percentage; (d) Disruption duration to 40 % customers-out; (e) Disruption duration to 20 % customers-out. (d) Restoration quantile; (e) Restore efficiency; (f) Net cumulative customer restoration; (g) Recovery duration.

superior recovery duration. The cluster records an average of approximately 12 daily outage hours per customer, slightly lower than Cluster 0, but exhibits the highest median daily customer-out hours, suggesting widespread concurrent service disruptions. The median customer-out percentage falls between the higher level in Cluster 0 and the lower level in Cluster 2, while the timeframes for reaching the 40 % and 20 % restoration thresholds remain high. In terms of recovery, Cluster 1 demonstrates strong performance, with lower restoration quantiles, indicating faster restoration, and higher restore-to-outage efficiency, reflecting more effective operations compared to Cluster 0. Although this cluster has a larger resilience area, which suggests greater separation between outage and restoration curves, which may be because of the widespread affected customers, it ultimately achieves the shortest overall outage duration among all clusters.

Cluster 2: low initial impact with quick but uneven recovery

Cluster 2 (green) is most prevalent in the southeastern and western regions of the study area. This cluster experiences the lowest outage severity, with average daily outage hours per customer remaining below 12 h, the moderate median daily customer-out hours, and the lowest median customer out percentage, indicating relatively minor disruptions. The cluster also achieves shortest timeframes for reducing outages to both 40 % and 20 % thresholds, suggesting its relatively low initial outage impact. However, despite this efficient recovery, Cluster 2 experiences challenges in uneven recovery. Its moderate resilience area, positioned between Clusters 0 and 1, indicates a typical separation between cumulative outage and restoration trends. Although it has small restoration quantile and large restore efficiency, and the small median value of recovery duration, Cluster 2 exhibits a widely dispersed recovery range, spanning from approximately 10 to 30 days. This highlights significant variability in the recovery process across different locations, potentially due to differences in infrastructure, resource allocation, or geographic constraints.

In the next step, we created the spatial distribution of quantitative differences across clusters for the impact and recovery features in (Figs. 13 and 14). For average daily outage hours per customer, cluster 0 exhibited the highest values, reflecting prolonged outages in the areas. Cluster 1 showed intermediate values, while cluster 2 displayed the lowest values. For median daily customer-out hours, cluster 1, especially the northern part in this cluster, exhibited significantly higher values, whereas clusters 0 and 2 showed comparatively lower values. The median customer outage percentage highlights differences in the extent of customer impact. Cluster 0 demonstrated the highest values of the feature, especially in the central region, indicating widespread impacts within these areas. In contrast, clusters 1 and 2 showed the lowest values, suggesting that outages in these areas do not affect a large portion of customers. The disruption days to 40 % customer-out and days to 20 % customer-out are similarly distributed in the map and they reveal

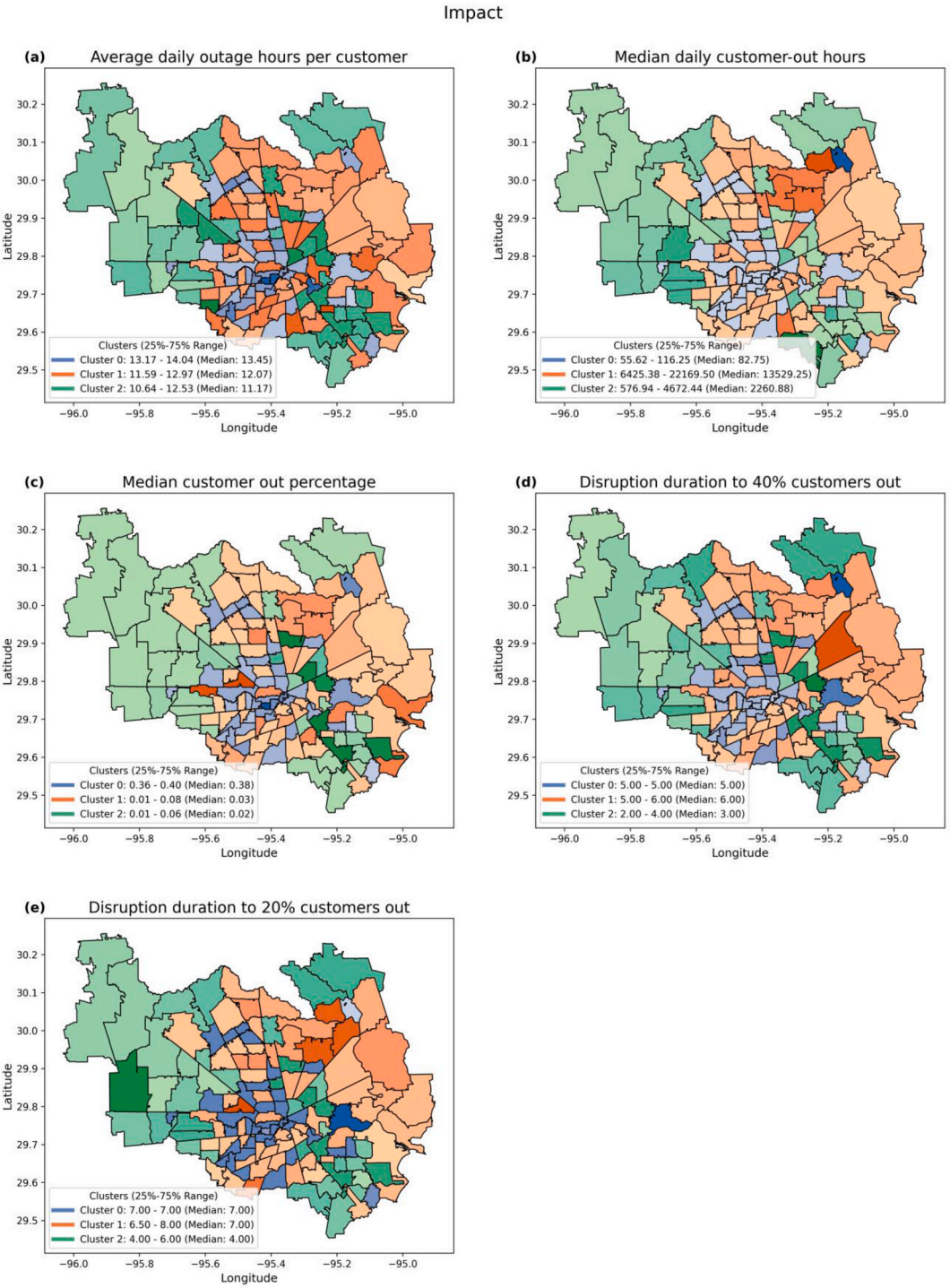


Fig. 13. Spatial distribution of impact features. (a) Average daily outage hours per customer; (b) Median daily customer-out hours; (c) Median customer-out percentage; (d) Disruption duration to 40 % customers-out; (e) Disruption duration to 20 % customers-out. Each feature was normalized within its respective cluster to reflect relative values. We used a color gradient to represent low, medium, and high values for each cluster, with shades assigned dynamically based on ranked cluster medians and interquartile ranges.

significant disparities across clusters in recovery timelines. Cluster 1, mainly in the northeastern regions, consistently showed longer durations, suggesting delays in impact efforts. Cluster 0 demonstrated the shortest restoration times, reflecting efficient recovery processes. Cluster 2 showed moderate restoration times.

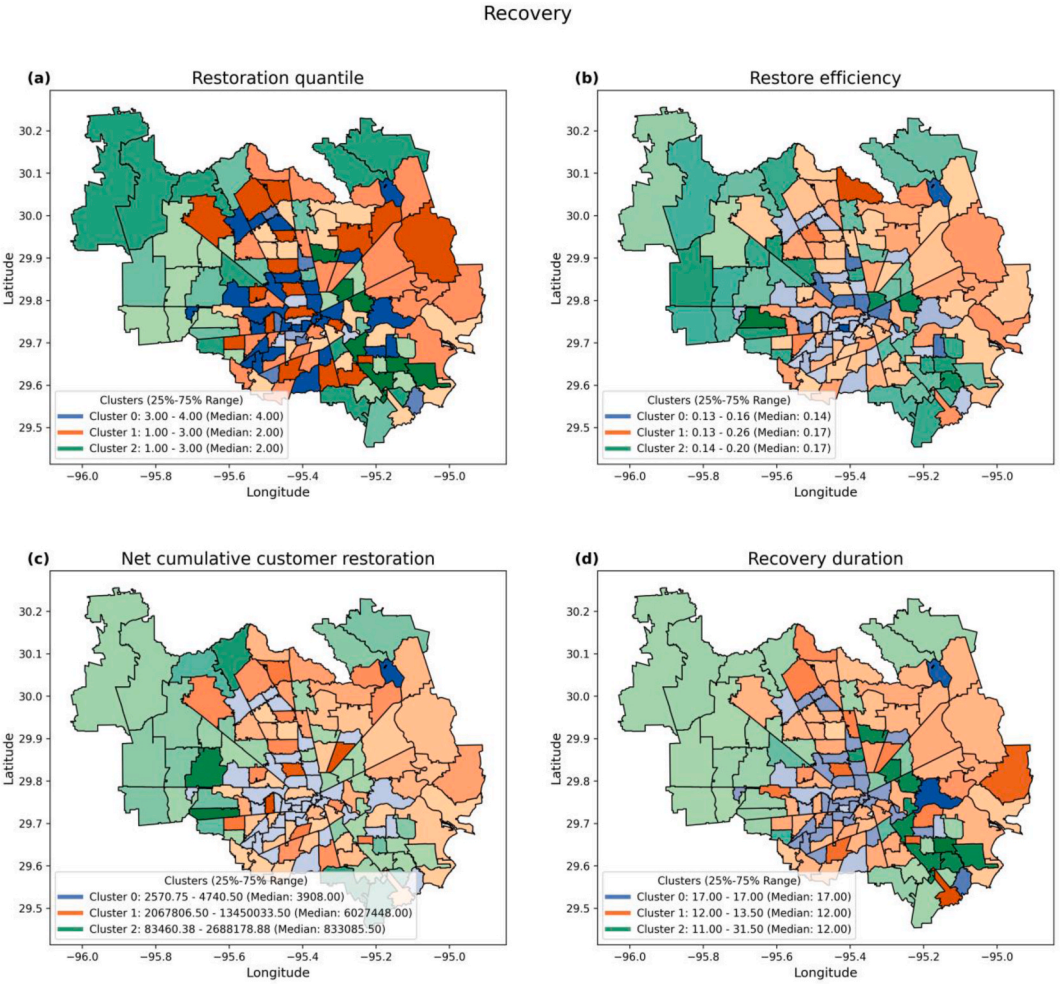


Fig. 14. Spatial distribution of recovery features. (a) Restoration quantile; (b) Restore efficiency; (c) Net cumulative customer restoration; (d) Recovery duration. We used a color gradient to represent low, medium, and high values for each cluster, with shades assigned dynamically based on ranked cluster medians and interquartile ranges.

The restoration quantile reveals stark differences in recovery speed across clusters. Cluster 0, especially in central and eastern regions, consistently exhibited the highest quantile values in central regions, indicating the lowest recovery rate. Conversely, clusters 1 and 2 demonstrated lower quantile values in western and eastern regions, reflecting a quicker recovery rate. For the restore efficiency, Clusters 0 and 2, concentrated in the western regions, displayed higher recovery rates, while Clusters 1 showed slower recovery rates. The net cumulative customer restoration metric highlights cumulative delays in recovery. Cluster 1 exhibited the largest net cumulative customer restorations, located in the northern and eastern regions, indicating significant lags between outage progression and restoration. Cluster 2 displayed moderate net cumulative customer restoration, with delays concentrated in the western areas. Cluster 0 had the smallest resilience areas, reflecting effective restoration efforts and minimal delays. For the recovery duration, cluster 0 has the longest recovery times, with a consistent median of 17.00 days. This cluster is primarily located in the central and southwestern parts of the region. The uniform shading across the areas within this cluster indicates a consistently delayed recovery process across these regions. Cluster 1 demonstrates shorter recovery durations ranging from 12.00 to 13.50 days, with a median of 12.00 days. This cluster is mainly distributed across the eastern and central regions. The consistent recovery durations across the areas in this cluster suggest an overall efficient and uniform recovery process. Cluster 2 also has a median recovery duration of 12.00 days, identical to cluster 1. The recovery durations in Cluster 2, however, vary more widely, ranging from 11.00 to 31.50 days. Notably, the areas with longer recovery times are predominantly located in the eastern parts of this cluster, indicating localized delays in recovery efforts.

The boxplots provide a comprehensive visual comparison of key factors across three distinct clusters (Fig. 15). Cluster 0 exhibits the highest median population density, suggesting urban areas with concentrated infrastructure, suggestions that this area might have great power outage impact and a likely greater demand for rapid restoration [40]. In contrast, clusters 1 and 2, with lower population densities, may represent suburban or rural contexts. Canopy cover varies significantly, with Cluster 2 exhibiting the highest median levels, suggesting denser vegetation, which may be susceptible to power outages and experience longer recovery times [47]. Cluster 0,

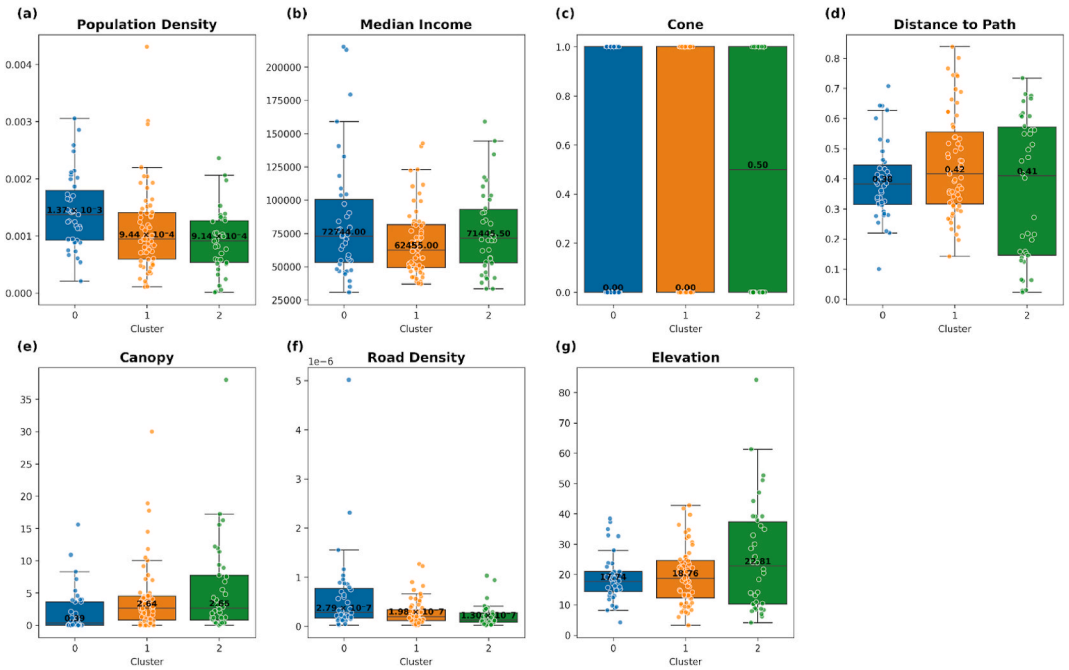


Fig. 15. Box plots of the urban form and structure features across the three clusters. (a) Population Density; (b) Median Income; (c) Cone; (d) Distance to Path; (e) Canopy; (f) Road Density; (g) Elevation.

characterized by lower vegetation, aligns with its urban nature. Income patterns reveal a clear gradient, with Cluster 0 and cluster 2 encompassing higher median income potentially benefiting the recovery process (C.-C. [48]), and cluster 1 reflecting the lowest incomes, potentially limiting resources for preventive measures and restoration. Road density is highest in Cluster 0 compared to Clusters 1 and 2, suggesting that power outages may also disrupt traffic signalization, further hindering the transportation network and exacerbating overall impacts[49]. Elevation also varies, with cluster 0 occupying the lowest median elevations, possibly increasing flood risks with power outage impact [46], while clusters 1 and 2 feature intermediate and highest elevations, respectively. Lastly, the clusters differ in their distance to the hurricane path and cone. Cluster 2 has almost half of areas in the affected cone, but Cluster 0 and 1 has more less areas in the cone. Cluster 0 exhibits the closest proximity, increasing storm exposure, while Cluster 2 includes areas both far from and directly within the storm's path, reflecting bifurcated risk levels. Synthesizing these insights, Cluster 0 represents

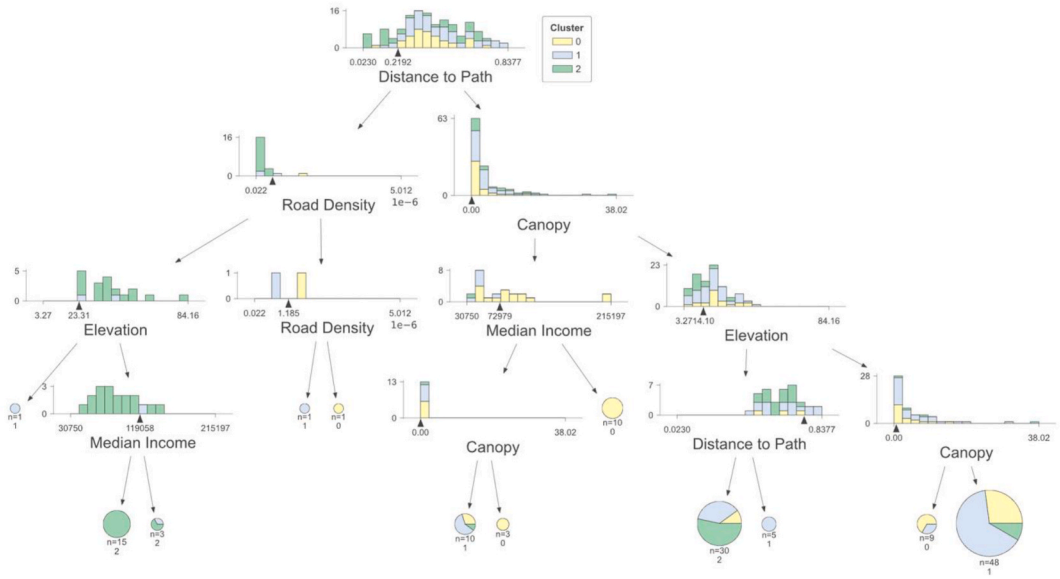


Fig. 16. Decision tree of the socioeconomic and environmental features.

higher-income, densely populated, low-elevation areas with well-developed roads and substantial storm exposure; cluster 1 includes lower-income ZIP Codes farther from the storm path but with moderate road density and elevation; and Cluster 2 contains highest canopy, elevations, and varied storm proximity, along with high income.

To further analyze how the combination of these factors shapes the pathways contributing to different levels of power outage impact and recovery (based on cluster classifications) with socioeconomic and environmental features, we employed a decision tree model (Fig. 16). This method was selected to elucidate the hierarchical structure and thresholds of feature values that influence cluster assignments, emphasizing the interpretability of feature interactions. The decision tree provides a detailed visualization of the pathways that lead to cluster assignments, and it consists of 12 distinct pathways, with each pathway representing a unique sequence of feature splits and thresholds that collectively determine the cluster classifications. This decision tree visualization illustrates how various features—such as distance to the hurricane path, road density, canopy coverage, elevation, and median income—contribute to the classification of clusters (Cluster 0, Cluster 1, and Cluster 2). Note that when a split value is very close to the minimum or maximum value on the axis, it may not be displayed. For example, in the split from "distance to path" to "canopy," the split node for canopy appears as 0. This indicates that the split point is very close to the minimum value of canopy, which is 0. In Fig. 16, the primary split occurs at distance to path, indicating that proximity to the hurricane's trajectory is the most significant factor in determining cluster membership. The left side of the tree, representing areas with lower distances to the path, exhibits a mix of clusters, while the right side, corresponding to areas farther from the path, is predominantly associated with Cluster 2. The left side of the plot contains $15 + 3$ counts of Cluster 2, while the right side has 30 counts, indicating a clear division within Cluster 2 based on this feature. Following this initial division, road density and canopy coverage emerge as key secondary splitting factors. For regions closer to the hurricane path, road density plays a crucial role in further distinguishing clusters, while for regions farther away, canopy coverage becomes a dominant factor in classification. Further refinements in the tree structure introduce elevation and median income, which contribute additional insights into the disparities in power outage impacts and recovery. Higher elevations appear to be more associated with Cluster 2, and median income helps differentiate cluster membership, with higher-income areas tending to align more with Cluster 0. Cluster 1 is more spread across different nodes in the decision tree, which aligns with its moderate impact and quick recovery characteristics. This spread suggests that Cluster 1 is influenced by multiple factors rather than being strongly defined by just one dominant characteristic.

At the leaf nodes, pie charts display the distribution of clusters, showing that some terminal nodes are heavily dominated by a single cluster, while others contain a mix, indicating that certain features do not strongly separate the groups. The decision tree highlights that distance to the hurricane path is the most influential determinant of cluster classification, reinforcing that areas closer to the storm experience more severe outage impacts and distinct recovery patterns. Additionally, infrastructure and environmental factors, such as canopy cover and road density, contribute to variations in power outages and recovery dynamics, likely due to their role in determining infrastructure resilience. Socio-economic characteristics, including elevation and income levels, further refine the classification, suggesting that both geographic and financial resources influence how communities experience and recover from power disruptions.

We also conducted a cluster-wise regression analysis to investigate the relationships between socioeconomic and environmental features with the impact and recovery features. Our aim was to understand the dynamics of power outage impacts and recovery across clusters with distinct socioeconomic and infrastructural characteristics. For each cluster, we ran separate linear regressions to evaluate

Table 3
Linear regression results between predictors (socioeconomic and environmental features) and target variables (power outage impact and recovery features).

Cluster	Target	Predictor	Coefficient	P-Value
0	Average daily outage hours per customer	Population density	0.501	0.005
0	Median customer out percentage	Population density	0.046	0.006
0	Median daily customer-out hours	Distance to path	1396.820	0.012
0	Net cumulative customer restoration	Median income	−0.585	0.023
0	Net cumulative customer restoration	Distance to path	43378.392	0.012
1	Disruption duration to 40 % customers out	Canopy	0.656	0.024
1	Restoration quantile	Population density	0.337	0.018
1	Restoration quantile	Road density	−0.258	0.025
1	Restoration quantile	Median Income	0.434	0.014
2	Average daily outage hours per customer	Distance to path	1.046	0.019
2	Recovery duration	Population density	−6.689	0.031
2	Median daily customer-out hours	Population density	−0.092	0.001
2	Median daily customer-out hours	Distance to path	−2163.479	0.034
2	Disruption duration to 40 % customers out	Road density	0.818	0.001
2	Disruption duration to 20 % customers out	Population density	−1.086	0.004
2	Disruption duration to 20 % customers out	Road density	0.880	0.004
2	Disruption duration to 20 % customers out	Elevation	−1.427	0.012
2	Disruption duration to 20 % customers out	Distance to path	−1.241	0.029
2	Disruption duration to 20 % customers out	Median Income	−1.144	0.012
2	Restoration quantile	Population density	−0.627	0.022
2	Restore efficiency	Elevation	−0.045	0.015
2	Restore efficiency	Distance to path	−0.040	0.035

the effect of predictors on target variables. The use of standardized predictors allowed us to control for variability in units across features. Significant predictors (p-value <0.05) were identified and interpreted for each cluster, emphasizing their unique roles in shaping outage impacts and recovery outcomes (Table 3).

Our analysis highlighted distinct patterns in clusters 0 and 2, which are characterized by relatively high median incomes. Cluster 0, situated closer to the hurricane path based on its median distance, showed a notable relationship between distance to the path and outage severity. In this cluster, a shorter distance to the path significantly contributed to an increase in impact and recovery metrics, including median daily customer-out hours and net cumulative customer restoration. Moreover, the high population density in cluster 0 is positively associated with the average daily outage hours per customer and median customer out percentage. This finding aligns with the expectation that densely populated areas experience greater challenges during power outages due to the higher demand for resources and complex restoration logistics. A high median income, however, improves their recovery process. Median income emerged as a significant predictor for recovery outcomes in high-income Cluster 0, underscoring the importance of income in enhancing recovery efforts. Specifically, in cluster 0, higher median income was associated with a reduced net cumulative customer restoration, suggesting income inequality in restoration speed across different ZIP Codes.

In cluster 2, the distance to the hurricane path exhibited varied effects on both impact and recovery metrics. This cluster includes areas that are either very close to or far from the hurricane's path, resulting in diverse outcomes. For impact metrics, distance to path shows a positive relationship with average daily outage hours per customer (coefficient = 1.046), indicating that areas farther from the hurricane path experienced slightly longer outages on average. The median daily customer-out hours, however, is negatively associated with distance to the path, suggesting that areas closer to the hurricane path experienced significantly more daily outage hours, reflecting the severe nature of the storm's immediate impact on these regions. Population density also has a notable effect, as it is negatively associated with the median customer-out hours (−0.092), implying that higher population density is linked to fewer custom-out hours. Disruption duration to 40 % customers shows a positive association with road density (0.82), suggesting that areas with more roads experience longer impact durations, potentially capturing the complexity of the topology of the power grid's downed power lines along the roads. Similarly, days to 20 % customer outages is influenced by several factors: it is negatively associated with population density (−1.08), elevation (−1.43), and distance to path (−1.24), indicating that higher-density areas, higher-elevation regions, and areas closer to the path recover faster. However, it is positively associated with road density (0.88), further supporting the notion that denser road networks are linked with the complexity of the topology of power grid networks. Median income also plays a role, with higher-income areas receiving a shorter impact to 20 % outages (−1.14), likely due to implicit inequality in restoration prioritization efforts. The restoration quantile, which reflects restoration rates in a seven-day window, shows a negative relationship with population density (−0.63), indicating that higher-density areas exhibit faster restoration rates (suggesting restoration efforts prioritize areas of greater population density). Similarly, the restore efficiency (restore/outage) is negatively associated with elevation (−0.04), suggesting slower restoration in higher-elevation areas. Overall, the findings for cluster 2 highlight the interplay between geographic and socioeconomic factors in shaping both the impact of and recovery from the hurricane. Areas closer to the path bear the brunt of the storm's impact but recover faster, potentially due to prioritized recovery efforts. In contrast, areas farther from the path experience less severe impacts but longer recovery times. The results also suggest the presence of implicit income inequality in prioritization of restoration efforts as areas with higher population had faster restoration.

Cluster 1, with the lowest income levels, faces distinct challenges in mitigating and recovering from power outages. Canopy cover significantly affects recovery times, as tree-related disruptions like fallen branches and obstructed roads slow the restoration process. In Cluster 1, tree canopy is positively associated with disruption duration to 40 % customers-out (0.66), suggesting that areas with more tree cover tend to experience more complex damages to power lines and have slightly longer recovery times to reduce outages to 40 %. Restoration quantile, a measure of restoration speed in a seven-day window, is positively associated with population density (0.34), indicating that more densely populated areas in this cluster experience slower restoration rates perhaps due to the complexity of damages caused by a higher canopy cover. Population density is associated with slower restoration rate, while higher road density facilitates faster recovery, reflecting a balance between infrastructure strain and repair accessibility. However, low-income levels limit these communities' ability to invest in resilient infrastructure or to mobilize resources, leaving them reliant on external aid. This cluster highlights the vulnerability of resource-constrained areas to prolonged outages, even when direct storm exposure is moderate.

These cluster-specific analyses underscore that hurricane-induced power disruptions result from the interplay of socioeconomic (e.g., income, population density) and infrastructural factors (e.g., elevation, canopy cover, proximity to the path). High-income, densely populated areas (cluster 0) can mitigate some impacts through robust infrastructure, yet direct exposure to storms remains a critical challenge. Mixed-vulnerability areas (cluster 2) require strategies addressing both immediate impacts in high-risk zones and prolonged recovery in outlying areas. Resource-limited communities (cluster 1) would benefit from targeted investments in power grid infrastructure and vegetation management.

5. Discussion and conclusion

Our study provides an extensive spatiotemporal analysis of power outage impacts and recovery patterns following Hurricane Beryl in Houston, Texas. Despite a growing body of research on storm-induced power outages, existing empirical studies remain constrained by their reliance on limited outage metrics and coarse spatial and temporal scales [1,3,4,50]. By focusing on only a narrow range of outage features—often aggregated to large geographic units such as counties—prior work has lacked the granularity to detect nuanced patterns of infrastructure vulnerability and recovery, leaving important intra-community disparities concealed. Also, while many studies highlight spatial inequalities, they frequently overlook the temporal dimension, preventing a robust understanding of how recovery speed varies across communities over time[13,51]. This study addresses these gaps by employing a high-resolution dataset

that captures multiple dimensions of power outages—severity, scale, and duration—at the more refined ZIP Code level. In doing so, it not only advances the empirical characterization of widespread outages but also contributes critical insights into the spatiotemporal interplay of sociodemographic, urban development, and environmental factors, thereby offering a comprehensive framework that can inform more equitable and resilient planning and recovery strategies.

The findings of this study offer four important contributions to the state of knowledge and practice. First, the study provides a detailed empirical analysis of power disruptions during a significant hurricane event within a major metropolitan area, yielding valuable insights into urban infrastructure vulnerabilities. Second, the study develops a comprehensive framework for analyzing outage patterns by incorporating multiple metrics, including duration and restoration rates, resulting in a more nuanced understanding of service disruption impacts. The analysis reveals key insights into the dynamics of impact and recovery across the Houston area during Hurricane Beryl. The cumulative outage curve peaked rapidly at approximately 1.3×10^8 customer power outages (based on 15-min intervals), reflecting a sudden and widespread disruption. Restoration followed a gradual trajectory, with cumulative restore values converging with the outage curve by July 27, 2024m signaling a return to baseline conditions. The total customer-outage hours amounted to more than 143 million during this period, with the northeastern and central regions experiencing the most severe disruptions, including prolonged average daily outages exceeding 14 h. Median outage percentages were highest in the northeastern area of the county, reaching up to 84.25 %, while central and southeastern areas faced significant disruptions. Recovery metrics highlighted stark disparities: the northeastern areas required up to 15 days to reduce outages to 20 %, whereas central and southwestern areas recovered within 5 days. Restoration quantiles and restore efficiency illustrated uneven recovery efforts, with central-eastern areas demonstrating faster restoration rates, while northeastern areas lagged behind. Spatial trends in recovery duration revealed prolonged outages exceeding 40 days in southern and central areas, contrasting with shorter power-outage durations in northern areas. Collectively, these findings emphasize the complex interplay of outage impact, recovery and socioenvironmental factors.

Third, the research harnesses granular outage data at both detailed geographic levels and frequent temporal intervals, revealing previously undetected patterns of spatial inequality. The analysis also identified three distinct clusters with varying vulnerability and recovery profiles. Areas closest to the hurricane path experienced the most severe initial outages but often saw faster restoration, likely due to prioritized response efforts. Cluster 0, characterized by high density and higher income levels, suffered significant initial impacts from direct storm exposure but showed efficient partial recovery rates despite longer total recovery duration. Median income helped reduce impact and recovery Cluster 1, comprising lower-income areas with moderate distance from the storm path and having canopy cover, exhibited slower daily restoration rates, highlighting how limited resources and pre-disaster investments amplify vulnerability. Cluster 2, featuring moderate income levels, variable storm proximity, and higher elevation, showed milder initial impacts but faced extended recovery periods, suggesting potential under-prioritization of peripheral areas. The results highlight significant spatial disparities based on regional characteristics. Areas with high population density and proximity to the hurricane path exhibited the highest impacts, including prolonged outage hours per customer and a higher percentage of customers affected. However, in areas with higher median income, recovery processes were more efficient, as evidenced by lower restoration quantiles, faster recovery durations, and smaller total customer-outage hours, which reflect implicit income inequality in prioritizing restoration efforts.

Fourth, the study conducts a thorough investigation into how various factors—including socioeconomic conditions, urban development patterns, and environmental characteristics—influence the disparities in outage impacts and recovery trajectories across different communities. Several key factors emerged as significant determinants of spatial disparity in outage impacts and recovery rates. In regions characterized by a mix of proximity to and distance from the hurricane path, recovery outcomes varied. Areas closer to the path experienced more severe initial impacts due to direct exposure but showed faster recoveries, likely due to prioritized recovery efforts. Conversely, areas farther from the path experienced less severe initial impacts but faced prolonged recovery durations. Population density in these regions helped reduce the proportion of affected customers and enhanced restoration rates. Real-world disaster recovery scenarios involve intricate interactions between environmental exposure and community characteristics, moving beyond simple cause-and-effect relationships. This study reveals nuanced relationships between environmental and socioeconomic factors in disaster recovery. Densely populated areas, typically wealthier, demonstrate resilience through income-driven resources that accelerate recovery despite proximity to storm paths. Areas with low median income consistently demonstrated weaker recovery outcomes, regardless of other geographic or environmental advantages. Areas with substantial tree canopy face prolonged disruptions, with this effect magnified in lower-income communities where vegetation management resources are scarce, and road infrastructure improvements could enhance recovery outcomes in these regions. High-elevation areas present a complex recovery pattern where lower population density extends disruption periods and slows recovery, often relegating these communities to lower priority status in restoration efforts. Limited road networks here paradoxically reduce initial disruption severity, likely because there is less infrastructure susceptible to damage, resulting in a less severe initial impact. These findings highlight how geographical advantages can simultaneously create recovery challenges, revealing counterintuitive interactions between physical terrain, infrastructure density, and socioeconomic resources that shape community resilience. It shows the opportunity to explore the complex situation of geographic vulnerability trade-offs, where the same features that provide protection during initial disaster phases may compromise recovery efficiency—a critical insight for developing more equitable resource allocation frameworks that account for both immediate impact severity and long-term restoration challenges across diverse community types. For example, communities on higher ground may avoid floodwaters but would benefit from backup microgrids to offset delayed access during recovery or improve community connections. Low-income, tree-dense neighborhoods could be supported through subsidized vegetation management and improved road connectivity to accelerate restoration.

Future research would benefit from incorporating higher-resolution grid characteristics data, conducting longitudinal studies of

repeated storms, and integrating community engagement to capture local insights into the restoration process. Additionally, incorporating other infrastructure-related variables, such as the median year structure built, along with other socioeconomic factors—such as ethnics, employment types—could provide deeper insights into the relationship between structural age, social vulnerability, and power resilience. Our findings reinforce the critical importance of equity and targeted investments in strengthening both immediate response capabilities and long-term power infrastructure resilience. Expanding these dimensions in future studies would provide a more comprehensive understanding of outage severity and recovery patterns, which is essential for developing equitable and effective resilience strategies.

CRediT authorship contribution statement

Xiangpeng Li: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Junwei Ma:** Writing – review & editing, Writing – original draft, Formal analysis. **Ali Mostafavi:** Writing – review & editing, Supervision, Project administration, Conceptualization.

Data availability

The power outage data that support the findings of this study are available pre request from EAGLE-I™. The other datasets used in this paper are publicly accessible and cited in this paper.

Code availability

The code that supports the findings of this study is available from the corresponding author upon request.

Declaration of competing interest

The authors declare that they have no known competing interests.

Acknowledgement

This work was supported by the National Science Foundation under Grant CMMI-1846069 (CAREER). We would also like to thank the Oak Ridge National Laboratory for providing the original power outage data. Any opinions, findings, conclusions, or recommendations expressed in this research are those of the authors and do not necessarily reflect the view of the National Science Foundation and the Oak Ridge National Laboratory.

Data availability

Data will be made available on request.

References

[1] N.M. Flores, A.J. Northrop, V. Do, M. Gordon, Y. Jiang, K.E. Rudolph, D. Hernández, J.A. Casey, Powerless in the storm: severe weather-driven power outages in New York State, 2017–2020, PLOS Climate 3 (5) (2024) e0000364, https://doi.org/10.1371/journal.pclm.0000364.
[2] L. Xu, K. Feng, N. Lin, A.T.D. Perera, H.V. Poor, L. Xie, C. Ji, X.A. Sun, Q. Guo, M. O'Malley, Resilience of renewable power systems under climate risks, Nat. Rev. Electr. Eng. 1 (1) (2024) 53–66, https://doi.org/10.1038/s44287-023-00003-8.
[3] R.Z. Zhou, Y. Hu, L. Zou, H. Cai, B. Zhou, Understanding the disparate impacts of the 2021 Texas winter storm and power outages through mobile phone location data and nighttime light images, Int. J. Disaster Risk Reduct. 103 (2024) 104339, https://doi.org/10.1016/j.ijdrr.2024.104339.
[4] J. Dugan, D. Byles, S. Mohagheghi, Social vulnerability to long-duration power outages, Int. J. Disaster Risk Reduct. 85 (2023) 103501, https://doi.org/10.1016/j.ijdrr.2022.103501.
[5] S. Baik, A.L. Davis, J.W. Park, S. Sirinterlikci, M.G. Morgan, Estimating what US residential customers are willing to pay for resilience to large electricity outages of long duration, Nat. Energy 5 (3) (2020) 250–258, https://doi.org/10.1038/s41560-020-0581-1.
[6] V. Do, H. McBrien, N.M. Flores, A.J. Northrop, J. Schlegelmilch, M.V. Kiang, J.A. Casey, Spatiotemporal distribution of power outages with climate events and social vulnerability in the USA, Nat. Commun. 14 (1) (2023) 2470, https://doi.org/10.1038/s41467-023-38084-6.
[7] K. Feng, M. Ouyang, N. Lin, Tropical cyclone-blackout-heatwave compound hazard resilience in a changing climate, Nat. Commun. 13 (1) (2022) 4421, https://doi.org/10.1038/s41467-022-32018-4.
[8] H. Egner, A. Pott (Eds.), Geographische Risikoforschung: Zur Konstruktion Verräumlichter Risiken Und Sicherheiten, Franz Steiner Verlag, 2010, https://doi.org/10.25162/9783515099868.
[9] S. Fuchs, M. Keiler, S. Sokratov, A. Shnyparkov, Spatiotemporal dynamics: the need for an innovative approach in Mountain hazard risk management, Nat. Hazards 68 (3) (2013) 1217–1241, https://doi.org/10.1007/s11069-012-0508-7.
[10] H.-D. Müller-Mahn, The Spatial Dimension of Risk: How Geography Shapes the Emergence of Riskscapes, Routledge, New York, 2013.
[11] J. Tang, Y. Li, S. Cui, L. Xu, Y. Hu, S. Ding, V. Nitivattananon, Analyzing the spatiotemporal dynamics of flood risk and its driving factors in a coastal watershed of southeastern China, Ecol. Indic. 121 (2021) 107134, https://doi.org/10.1016/j.ecolind.2020.107134.
[12] C. Aubrecht, S. Fuchs, C. Neuhold, Spatio-temporal aspects and dimensions in integrated disaster risk management, Nat. Hazards 68 (3) (2013) 1205–1216, https://doi.org/10.1007/s11069-013-0619-9.
[13] N. Coleman, A. Esmalian, C.-C. Lee, E. Gonzales, P. Koirala, A. Mostafavi, Energy inequality in climate hazards: empirical evidence of social and spatial disparities in managed and hazard-induced power outages, Sustain. Cities Soc. 92 (2023) 104491, https://doi.org/10.1016/j.scs.2023.104491.
[14] B. Li, A. Mostafavi, Unraveling fundamental properties of power system resilience curves using unsupervised machine learning, Energy and AI 16 (2024) 100351, https://doi.org/10.1016/j.egyai.2024.100351.

[15] S.M. Lee, S. Chinthavali, N. Bhusal, N. Stenvig, A. Tabassum, T. Kuruganti, Quantifying the power system resilience of the US power grid through weather and power outage data mapping, IEEE Access 12 (2024) 5237–5255, https://doi.org/10.1109/ACCESS.2023.3347129.
[16] Bell Vaiman, Chowdhury Chen, Hines Dobson, Miller Papic, Zhang, Risk assessment of cascading outages: methodologies and challenges, IEEE Trans. Power Syst. 27 (2) (2012) 631–641, https://doi.org/10.1109/TPWRS.2011.2177868.
[17] L.N. Dunn, M.D. Sohn, K.H. LaCommare, J.H. Eto, Exploratory analysis of high-resolution power interruption data reveals spatial and temporal heterogeneity in electric grid reliability, Energy Policy 129 (2019) 206–214, https://doi.org/10.1016/j.enpol.2019.01.042.
[18] B. Austgen, E. Kutanoglu, J.J. Hasenbein, A two-stage stochastic programming model for electric substation flood mitigation prior to an imminent hurricane, IISE Transactions (2024) 1–18, https://doi.org/10.1080/24725854.2024.2393654.
[19] M. Holub, J. Suda, S. Fuchs, Mountain hazards: reducing vulnerability by adapted building design, Environ. Earth Sci. 66 (7) (2012) 1853–1870, https://doi.org/10.1007/s12665-011-1410-4.
[20] H. Kienholz, Early warning systems related to Mountain hazards, in: J. Zschau, A. Küppers (Eds.), Early Warning Systems for Natural Disaster Reduction, Springer Berlin Heidelberg, 2003, pp. 555–564, https://doi.org/10.1007/978-3-642-55903-7_75.
[21] K. Best, S. Kerr, A. Reilly, A. Patwardhan, D. Niemeier, S. Guikema, Spatial regression identifies socioeconomic inequality in multi-stage power outage recovery after Hurricane Isaac, Nat. Hazards 117 (1) (2023) 851–873, https://doi.org/10.1007/s11069-023-05886-2.
[22] W. Sun, P. Bocchini, B.D. Davison, Resilience metrics and measurement methods for transportation infrastructure: the state of the art, Sustain. Resilient Infrastructure 5 (3) (2020) 168–199, https://doi.org/10.1080/23789689.2018.1448663.
[23] X. Li, Y. Jiang, A. Mostafavi, Hazard exposure heterophily in socio-spatial networks contributes to post-disaster recovery in low-income populations, Int. J. Disaster Risk Reduct. 108 (2024) 104537, https://doi.org/10.1016/j.ijdrr.2024.104537.
[24] M. Santini, F. Armas, A. Carioli, P. Ceccato, S. Coecke, L. De Girolamo, A.-M. Duta, B.M. Gawlik, L. Giustolisi, M. Kalas, A. Maddalon, I. Marí Rivero, S. Marzi, D. Masante, M. Mastronunzio, D. Moalli, G. Panzarella, K. Poljansek, C. Proietti (Eds.), Hurricane Beryl: JRC Emergency Report: Situation as at 09 July 2024 and Comparison with past Hurricanes, Publications Office, 2024, https://doi.org/10.2760/934959. European Commission.
[25] M. Zhang, These Houston areas received the strongest wind gusts during Hurricane Beryl. https://houstonlanding.org/these-houston-areas-received-the-strongest-wind-gusts-during-hurricane-beryl/, 2024, March 11.
[26] P. Arbaje, Hurricane Beryl's towering Texas grid impacts and price tag highlight massive injustices. https://blog.ucs.org/paul-arbaje/hurricane-beryls-towering-texas-grid-impacts-and-price-tag-highlight-massive-injustices/, 2024, August 2.
[27] KML, KMZ Viewer, *KML, KMZ viewer* [computer software]. https://kmlviewer.nsspot.net/, 2024.
[28] NHC NOAA, NHC GIS Archive—tropical cyclone advisory forecast. https://www.nhc.noaa.gov/gis/archive_forecast_results.php?id=al02&year=2024&name=Hurricane%20BERYL, 2024.
[29] J. Yu, H. Baroud, Quantifying community resilience using hierarchical Bayesian Kernel methods: a case study on recovery from power outages, Risk Anal. 39 (9) (2019) 1930–1948, https://doi.org/10.1111/risa.13343.
[30] P. Sedgwick, Pearson's correlation coefficient, BMJ 345 (jul04 1) (2012), https://doi.org/10.1136/bmj.e4483 e4483–e4483.
[31] J.C.F. de Winter, S.D. Gosling, J. Potter, Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: a tutorial using simulations and empirical data, Psychol. Methods 21 (3) (2016) 273–290, https://doi.org/10.1037/met0000079.
[32] L. McInnes, J. Healy, J. Melville, UMAP: uniform manifold approximation and projection for dimension reduction, (arXiv:1802.03426). arXiv (2020). http://arxiv.org/abs/1802.03426.
[33] J. Healy, L. McInnes, Uniform manifold approximation and projection, Nature Reviews Methods Primers 4 (1) (2024) 82, https://doi.org/10.1038/s43586-024-00363-x.
[34] G. Hamerly, C. Elkan, Learning the k in k-means, in: S. Thrun, L. Saul, B. Schölkopf (Eds.), Advances in Neural Information Processing Systems, vol. 16, MIT Press, 2003, in: https://proceedings.neurips.cc/paper_files/paper/2003/file/234833147b97bb6aed53a8f4f1c7a7d8-Paper.pdf.
[35] U.S. Census Bureau, 2020 census demographic and housing characteristics file (DHC). https://www.census.gov/programs-surveys/decennial-census/technical-documentation/complete-technical-documents.2020.html#dhc-and-dp, 2020.
[36] U.S. Geological Survey, Multi-resolution land characteristics consortium. https://www.mrlc.gov/data/type/tree-canopy, 2025.
[37] U.S. Census Bureau, 2018-2022 American community survey 5-Year estimates. https://www.census.gov/programs-surveys/acs/technical-documentation.html, 2022.
[38] U.S. Census Bureau, TIGER/line shapefiles. https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html, 2024.
[39] NASA JPL, NASA shuttle radar topography mission global 1 arc second number, NASA EOSDIS Land Processes Distributed Active Archive Center (2013). https://doi.org/10.5067/MEASURES/SRTM/SRTMGL1N.003.
[40] A. Bhattacharyya, M. Hastak, A data-driven approach to quantify disparities in power outages, Sci. Rep. 13 (1) (2023) 7247, https://doi.org/10.1038/s41598-023-34186-9.
[41] R. Snaiki, T. Wu, A.S. Whittaker, J.F. Atkinson, Hurricane wind and storm surge effects on coastal bridges under a changing climate, Transp. Res. Rec.: J. Transport. Res. Board 2674 (6) (2020) 23–32, https://doi.org/10.1177/0361198120917671.
[42] W.G. Peacock, S. Van Zandt, Y. Zhang, W.E. Highfield, Inequities in long-term housing recovery after disasters, J. Am. Plann. Assoc. 80 (4) (2014) 356–371, https://doi.org/10.1080/01944363.2014.980440.
[43] B. Kar, J. Brewer, O. Omitaomu, B. Turner, S. Levinson, Y. Chen, N. Roberts, M. Prica, A. Iyengar, RePOWERD: Restoration of Power Outage from wide-area Severe Weather Disruptions, 2022, https://doi.org/10.2172/1887668. ORNL/TM-2022/2621, 1887668; p. ORNL/TM-2022/2621, 1887668.
[44] J. Ma, B. Li, A. Mostafavi, Characterizing urban lifestyle signatures using motif properties in network of places, Environ. Plan. B Urban Anal. City Sci. 51 (4) (2024) 889–903, https://doi.org/10.1177/23998083231206171.
[45] Y. Ho, L. Li, A. Mostafavi, Integrated vision language and foundation model for automated estimation of building lowest floor elevation, Comput. Aided Civ. Infrastruct. Eng. 40 (1) (2025) 75–90, https://doi.org/10.1111/mice.13310.
[46] S.M. Quiring, L. Zhu, S.D. Guikema, Importance of soil and elevation characteristics for modeling hurricane-induced power outages, Nat. Hazards 58 (1) (2011) 365–390, https://doi.org/10.1007/s11069-010-9672-9.
[47] W.O. Taylor, P.L. Watson, D. Cerrai, E. Anagnostou, A statistical framework for evaluating the effectiveness of vegetation management in reducing power outages caused during storms in distribution networks, Sustainability 14 (2) (2022) 904, https://doi.org/10.3390/su14020904.
[48] C.-C. Lee, M. Maron, A. Mostafavi, Community-scale big data reveals disparate impacts of the Texas winter storm of 2021 and its managed power outage, Humanit. Soc. Sci. Commun. 9 (1) (2022) 335, https://doi.org/10.1057/s41599-022-01353-8.
[49] M.B. Ulak, A. Kocatepe, L.M. Konila Sriram, E.E. Ozguven, R. Arghandeh, Assessment of the hurricane-induced power outages from a demographic, socioeconomic, and transportation perspective, Nat. Hazards 92 (3) (2018) 1489–1508, https://doi.org/10.1007/s11069-018-3260-9.
[50] H. Hou, S. Zhu, H. Geng, M. Li, Y. Xie, L. Zhu, Y. Huang, Spatial distribution assessment of power outage under typhoon disasters, Int. J. Electr. Power Energy Syst. 132 (2021) 107169, https://doi.org/10.1016/j.ijepes.2021.107169.
[51] C.-W. Hsu, A. Mostafavi, Untangling the relationship between power outage and population activity recovery in disasters, Resilient Cities and Structures 3 (3) (2024) 53–64, https://doi.org/10.1016/j.rcns.2024.06.003.
[52] N. Coleman, C. Liu, Y. Zhao, A. Mostafavi, Lifestyle pattern analysis unveils recovery trajectories of communities impacted by disasters, Humanit. Soc. Sci. Commun. 10 (1) (2023) 803, https://doi.org/10.1057/s41599-023-02312-7.