

Case Studies for Data-Oriented Emergency Management/Planning in Complex Urban Systems

Kun Xie¹, Kaan Ozbay^{1(✉)}, Yuan Zhu¹, and Hong Yang²

¹ Department of Civil and Urban Engineering, Urban Mobility and Intelligent Transportation Systems (UrbanMITS) Laboratory, Center for Urban Science and Progress (CUSP), New York University, Brooklyn, NY 11201, USA

{kun.xie, kaan.ozbay, yuan.zhu}@nyu.edu

² Department of Modeling, Simulation and Visualization Engineering, Old Dominion University, Norfolk, VA 23529, USA

hyang@odu.edu

Abstract. To reduce the losses caused by natural disasters such as hurricanes, it is necessary to build effective and efficient emergency management/planning systems for cities. With increases in volume, variety and acquisition rate of urban data, major opportunities exist to implement data-oriented emergency management/planning. New York/New Jersey metropolitan area is selected as the study area. Large datasets related to emergency management/planning including, traffic operations, incidents, geographical and socio-economic characteristics, and evacuee behavior are collected from various sources. Five related case studies conducted using these unique datasets are summarized to present a comprehensive overview on how to use big urban data to obtain innovative solutions for emergency management and planning, in the context of complex urban systems. Useful insights are obtained from data for essential tasks of emergency management and planning such as evacuation demand estimation, determination of evacuation zones, evacuation planning and resilience assessment.

Keywords: Emergency management/planning · Complex urban systems · Big data · Evacuation modeling · Hurricane

1 Introduction

Hurricanes can have devastating effects on coastal areas due to flooding, high wind, and rainfall, resulting in serious loss of life and property. To reduce the losses caused by hurricanes, it is necessary to build effective and efficient emergency management/planning systems. Essential tasks of emergency management/planning include determination of evacuation zones (identify evacuation zones in a way to indicate its inhabitants whether or not they are prone to hurricane-related risk in advance of disaster impacts), evacuation demand estimation (estimate origins, destinations and numbers of evacuees based on evacuation zones, demographic features and evacuation behavior), evacuation planning (determine the evacuation time, destinations and routes based on

the evacuation demand), and resilience assessment (evaluate the recovery ability of transportation systems in the post-hurricane periods).

The complexity of urban systems creates challenges for emergency management/planning. The urban transportation systems are multimodal, generally composed of the highway system, the pedestrian system and the public transit system. The urban transportation systems are further complicated by the random occurrences of incidents such as accidents, disabled vehicles, debris, downed trees and flooding. Therefore, it is challenging to evaluate the carrying capacities of urban transportation systems, especially during the hurricane-impacted periods when hurricane-related incidents such as downed trees and flooding are more likely to happen. On the other hand, it is difficult to precisely estimate the evacuation demands which are closely related to the evacuation zone divisions and evacuation behavior. The determination of evacuation zones is associated with a variety of factors such as ground elevation, evacuation mobility and demographic features. Moreover, different evacuation behavior (e.g. whether to evacuate or not, how to evacuate and where to evacuate) is present among inhabitants who are prone to hurricane-related risks.

In the era of “Big Data”, with increases in volume, variety and acquisition rate of urban data, there are a number of very exciting opportunities to implement data-driven emergency management/planning. Massive amounts of digitalized data such as evacuation zone maps, past incidents, geographical features, historical highway traffic volumes, public transit ridership can be available from multiple sources. Useful insights can be obtained from this big urban data for performing essential tasks of emergency management/planning. Therefore, this paper aims to present a comprehensive overview on how to use the big urban data to provide solutions and innovations for emergency management/planning in the context of complex urban systems.

New York City (NYC) is vulnerable to hurricanes. According to NYC Office of Emergency Management (OME), NYC has about 600 miles of coastline and almost 3 million people living in the areas at the risk of hurricanes [1]. In the morning of August 28th, 2011 hurricane Irene made landfall at Coney Island, NYC and in the evening of October 29th, 2012 hurricane Sandy landed in New Jersey. Hurricanes Irene and Sandy caused significant devastation to the east coast (especially to NYC), but also provide valuable data for the research on emergency management/planning. Moreover, New York City’s open data policy makes a variety of datasets from government agencies available to the public. NYC and its surrounding regions are selected as the study areas.

2 Big Urban Data

Massive amounts of data from multiple sources are collected to support data-oriented emergency management/planning. The major datasets are classified into eight groups including evacuation management data, traffic incident data, taxi and subway trip data, traffic volume and demand data, evacuation survey data, geographical data, building damage data and socio-economic data. The sources and practical usage of those datasets are summarized in Table 1, and more detailed descriptions are introduced in the following subsections.

Table 1. Summary of sources and usages for datasets collected

Dataset	Source	Usage
Evacuation management	NYC Office of Emergency Management (OEM)	Estimate the demand for evacuation and destination choices of evacuees
Traffic incident	Transportation Operations Coordination Committee (TRANSCOM)	Estimate incident-induced capacity losses
Taxi and subway trip	NYC Taxi & Limousine Commission (TLC) and Metropolitan Transportation Authority (MTA)	Calibrate and validate the evacuation models as well as assess the resilience of transportation systems
Evacuation survey	Northern New Jersey evacuation survey	Analyze the behavior of evacuees
Geographical	National Elevation Dataset (NED)	Determine the division of evacuation zones
Building damage	Environment Systems Research Institute (ESRI)	Additional indicator for risk evaluation
Socio-economic	U.S. Census Bureau	Estimate the evacuation demand and the division of evacuation zones

2.1 Evacuation Management Data

NYC Office of Emergency Management (OEM) provides Hurricane Evacuation Zones Map¹ (downloadable as GIS shapefiles) to help residents make decisions on evacuation. Evacuation zone division was updated in 2013 after Hurricane Sandy, adding 600,000 New Yorkers not included within the boundaries of the former 2001 evacuation zones. The zone division is updated according to the empirical data during Hurricane Sandy and storm surge simulations which are based on the current climate situation. The 2013 evacuations zones are listed from zone 1 to zone 6, from the highest risk to the lowest risk. Evacuation centers which offer shelters to evacuees during hurricanes are also presented in the Hurricane Evacuation Zones Map. Evacuation zones can be used to estimate the demand for evacuation and the locations of evacuation centers are related with the destination choices of evacuees.

2.2 Traffic Incident Data

Incident data of the interstate, US and New York State highways in New York City and its surrounding areas from Oct. 1st 2012 to Jan. 31st 2013 were obtained from Transportation Operations Coordination Committee (TRANSCOM). More detailed description of this dataset is given in [2]. A total of 354 incidents occurred during the evacuation period (12 AM, Oct. 26th, 2012–12 PM, Oct. 29th, 2012) before Sandy's

¹ Source: <http://maps.nyc.gov/hurricane/>.

landfall. Those incidents can be classified as six different types including accident, debris, disabled vehicle, downed tree, flooding and others. Accidents and downed trees are the major incident types during evacuation the evacuation period, and account for over 50 % of all the incidents. The incident durations were computed using the fields of *create time* and *close time* in the incident records. Each incident was located in the GIS map according to its coordinates and then was matched to the highway where it was detected. Incident data can provide information on the highway capacity losses which are attributed to the occurrence of incidents right before and during hurricanes.

2.3 Taxi and Subway Trip Data

Taxi trip data of NYC is made available to public by NYC Taxi & Limousine Commission (TLC) [3, 4]. The dataset includes taxi trips from years 2010 to 2013 and it contains pick-up and drop-off time and location information. The taxi trips generated is approximately 175 million per year. Subway ridership data were obtained from Metropolitan Transportation Authority (MTA) turnstile dataset, which includes subway turnstile information since May, 2010 and is updated every week. The data is stored in txt format and available through an official data feed [5]. The data is organized by weeks, remote units (stations) and control areas (turnstiles). Each station can have multiple control areas, and for each turnstile, there are two increment counters used to record numbers of entries and exits. Typically, counter readings of each turnstile is recorded every four hours. Taxi and subway trip data are used to calibrate and validate the evacuation models as well as to assess the resilience of transportation systems.

2.4 Traffic Volume and Demand Data

NY Best Practice Model (NYBPM) [6], which covers 28 counties in the Tristate area and involves more than 22 million population, provide well-calibrated background traffic demand trip tables. In addition, the traffic volumes on the main interstate highways, US highways, and NY highways in the NYC and surrounding regions were obtained from TRANSCOM. The traffic volumes obtained from traffic sensors were used to build evacuation response curves [7] for critical corridors during evacuation period of Hurricane Sandy.

2.5 Evacuation Survey Data

A random digit dial telephone survey was conducted between August and October of 2008 in northern New Jersey [7]. It covers a large urban region consisting of Passaic, Bergen, Hudson, Morris, Essex, Middlesex and Union Counties. The total population of the region is approximately 4.5 million. In total, 2,218 households were interviewed with a set of questions related to their evacuation experience, disaster preparedness (including hurricane, industrial accident and catastrophic nuclear explosion), evacuation decision choices, evacuation destinations, and evacuation mode choices. In addition, a series of questions regarding the characteristics of the household and

household members, such as income, vehicle ownership, family size etc. were asked. The evacuation survey data can be used to analyze the behavior of evacuees and thus more accurate evacuation demand can be obtained.

2.6 Geographical Data

Digital Elevation Model (DEM) data of NYC provides a representation of the terrain with elevations above the ground in a regular raster form. The DEM data of Manhattan was extracted from National Elevation Dataset (NED) developed by U.S. Geological Survey (USGS)². The resolution of the DEM data is 1 arc second (about 90 feet) and the pixel values are elevations in feet based on North American Vertical Datum of 1988 (NAD83). The average elevation which is associated with the flooding risk was aggregated for each grid cell. Another geographic feature collected for each cell is the distance to the coast, since areas closer to the coast are more likely to be affected by the storm surges. Geographical data can be used to infer the division of evacuation zones.

2.7 Building Damage Data

The building damage record during Hurricane Sandy was achieved from the Environment Systems Research Institute (ESRI) datasets³. Federal Emergency Management Agency (FEMA) inspectors conducted field inspections of damaged properties and recorded relevant information such as location and damage level, when households applied for individual assistance. The number of damaged building was obtained by summarizing households in the same location, assuming they are from a single multi-family building. Buildings damaged in historical hurricanes can be used as an additional indicator for risk evaluation.

2.8 Socio-economic Data

The socio-economic data based on 2011 census survey was retrieved from U.S. Census Bureau⁴. The socio-economic data is composed of demographic features (e.g. total population, population under 14 and population over 65), economic features (e.g. employment and median income), and housing features (e.g. median value and household average size). The demographic features can be used to estimate the evacuation demand. In addition, socio-economic data can affect the division of evacuation zones. For example, the zones with large number of elderlies and children tend to be more vulnerable and should be given higher priority of evacuation.

² Source: <http://ned.usgs.gov/>.

³ Source: <http://www.arcgis.com/home/item.html?id=307dd522499d4a44a33d7296a5da5ea0>.

⁴ Source: <http://factfinder.census.gov>.

3 Data-Oriented Emergency Management/Planning

This section presents five case studies on how to use big urban data to gain useful insights for decision-making in emergency management/planning. The main purposes and key datasets used for each case study are listed in Table 2. Those five cases studies are all data-oriented and related with each other. The evacuation behavior analysis and evacuation zone prediction can be used to estimate the evacuation demand; while the incident analysis provide information on the uncertainties of capacity supply of transportation systems. Evacuation simulation is used to evaluate whether the capacity supply could accommodate the evacuation demand under different evacuation scenarios. Resilience assessment is post-evaluation on the recovery ability of transportation systems.

Table 2. Summary of case studies in data-oriented emergency management/planning

Case study	Main purpose	Key datasets used
Evacuation behavior analysis	Estimate evacuation demand	Evacuation survey data
Evacuation zone prediction	Identify evacuation zones	Evacuation management data, geographical data, building damage data, and demographic data
Traffic Incident analysis	Predict capacity related uncertainties	Traffic incident data
Evacuation simulation	Evaluate whether the capacity supply can accommodate the evacuation demand	Evacuation management data, taxi and subway data, traffic volume and demand data and demographic data
Resilience assessment	Evaluate the recovery ability of transportation systems	Evacuation management data and taxi and subway data

3.1 Evacuation Behavior Analysis

A key issue in evacuation studies is to understand the evacuation behavior of residents. Questions related to whether to evacuate, when to evacuate, how to evacuate, where to evacuate, etc. are critical in developing reasonable evacuation plans. Thus it is necessary to examine the factors that affect the evacuees' decisions regarding these questions. Questionnaires have been designed to interview the residents and aim to identify the underlying factors affecting their decision makings (please see the subsection "Evacuation Survey Data" for more details). Based on the surveyed results, statistical models such as logistic regressions, multinomial logit models, etc. have been developed to examine the key factors affecting the decisions. Factors such as the socio-economic and demographic characteristics of the evacuees, locations, and type of the extreme events (i.e. hurricanes/explosions) are often considered in the modeling process. The advanced models usually help improve our predictions for evacuation planning. However, in practices, many models were developed independently.

They did not account for the potential interactions among different evacuation behavior. In the decision-making process, many evacuees are likely to make their choices on a question conditional on the decisions for other questions. Thus there is necessity to examine the issue considering possible interactions among different evacuation behavioral responses.

As a pilot study, we have applied the dataset from the telephone survey [8] to investigate the relationship between evacuation decision (the preference to evacuate) and evacuation destination choices under the hurricane scenario. For the responses of evacuation decision, the ordered probit regression model has been proposed as the responses are ordered in terms of multilevel preference:

$$y_i^* = X_i' \beta + \varepsilon_i$$

$$y_i = \begin{cases} 1 & \text{if } \tau_0 < y_i^* \leq \tau_1 \quad (\text{Response} = \text{very unlikely}) \\ 2 & \text{if } \tau_1 < y_i^* \leq \tau_2 \quad (\text{Response} = \text{not very likely}) \\ 3 & \text{if } \tau_2 < y_i^* \leq \tau_3 \quad (\text{Response} = \text{somewhat likely}) \\ 4 & \text{if } \tau_3 < y_i^* \leq \tau_4 \quad (\text{Response} = \text{very likely}) \end{cases} \quad (1)$$

where y_i^* denotes the latent variable measuring the evacuation decision of the i^{th} interviewed person; X_i is a vector of observed non-random explanatory variables; β is a vector of unknown parameters; and ε_i is the random error term. The latent variable y_i^* is mapped to the observed variable y_i , according to threshold parameters τ_j 's, with $\tau_{j-1} < \tau_j$, $\tau_0 = -\infty$, and $\tau_J = +\infty$.

In addition, the choices on the potential evacuation destinations were modeled by the multinomial logit model. Given one choice as a reference (i.e., public shelter), the probability of each choice π_{ij} is compared to the probability of the reference choice π_{iJ} . For choices $j = 1, 2, \dots, J - 1$, the log-odds of each choice is assumed to follows linear model:

$$\eta_{ij} = \log \left(\frac{\pi_{ij}}{\pi_{iJ}} \right) = Z_i' \alpha_j \quad (2)$$

where Z_i' is a vector of explanatory variables and α_j is a vector of regression coefficients for each choice $j = 1, 2, \dots, J - 1$. To identify the potential relationship between the evacuation decision and the choice of the evacuation destinations, we have proposed the use of the structural equation modeling, where the evacuation decision y_i is used as one of the explanatory variable in evacuation destination model (Eq. (2)). More detailed description of the proposed approach is reported in our recent work (Yang et al. [9]). An example of the structure equation modeling process is shown in Fig. 1. Though only two behavioral responses have been examined in the pilot study, the proposed method can be extend to examine more complicated interactions among multiple types of behavioral responses.

The key factors that affect the evacuation decision as well as the evacuation destination choices have been determined through a Bayesian estimation approach, which

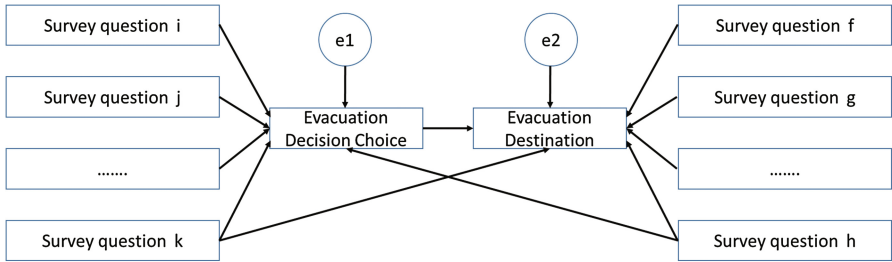


Fig. 1. Sample structural equation modeling process to explore multiple behavioral responses.

is not detailed here (See Yang et al. [9]). Other than the conventional factors such as age and distance to the shore, the modeling results suggest that there is only weak relationship between the evacuation decision choices and the evacuation destination choices. In other words, whether or not the individuals consider to evacuate, the decisions on choosing public shelters as well as other places as their evacuation destinations will not change notably based on the surveyed data.

3.2 Evacuation Zone Prediction

It is important for emergency planners to define evacuation zones which can indicate inhabitants whether or not they are prone to hurricane-related risk in advance of disaster impacts. The delineation of evacuation zones can be used to estimate the demand of evacuees, and thus it is helpful in developing effective evacuation management strategies. The evacuation zones defined currently cannot remain the same in the future, since the long-term climate change such as the rise of sea level would have major impacts on hurricane-related risks. One notable factor of climate change is global warming and the resulting rise of sea level. To manage emergency resources more efficiently, it is important to update the delineation of current evacuation zones to make it adaptable to the future hurricanes.

To predict future evacuation zones, traditional methods rely on the estimation of surge flooding using models such as the SLOSH (sea, lake, and overland surges from hurricanes) model and the ADCIRC (a parallel advanced circulation model for oceanic, coastal, and estuarine waters) model [10]. However, the implementation of the SLOSH and ADCIRC models can be really time-consuming and costly. We aim to develop a novel data-driven method which can promptly predict future evacuation zones in the context of climate change. Machine learning algorithms are used to learn the relationship between current pre-determined evacuation zones and hurricane-related factors, and then to predict how those zones should be updated as those hurricane-related factors change in the future.

The map of Manhattan, which is the central area of NYC, was uniformly split into 150×150 feet² grid cells ($N = 25,440$) as the basic geographical units of analysis.

Evacuation zone category (E1, E2, E3 and S)⁵, geographical features (including average elevation above sea level and distance to coast), historical hurricane information (including building damage intensity), evacuation mobility (including distance to the nearest evacuation center, distance to the nearest subway station, distance to the nearest bus stop and distance to the nearest expressway), and demographic features (including total population, population over 65 and population under 14) in the current year were captured for each cell. A decision tree and random forest were trained to relate cell-specific features with current zone categories which could reflect the risk levels during storms. Ten-fold cross-validation was used to evaluate model performance and performance measures of the classification tree and the random forest are reported in.

Table 3. It was found that the random forest outperformed the decision tree in term of the accuracy and Kappa statistic [11]. Regarding the better performance, the prediction outcomes of the random forest are visualized in the GIS map and compared with actual evacuation zones as presented in Fig. 2. It is found that the estimated evacuation zone division is quite similar to the actual one (accuracy = 94.13 %). It implies that the random forest succeeds in learning the potential pattern of delineating zones with different risk levels. More details on description and specification of the proposed models are presented in our recent work (Xie et al. [12]).

Table 3. Performance measures of the classification tree and the random forest

	Classification tree	Random forest
Correctly classified instances	22965	23947
Incorrectly classified instances	2475	1493
Total number of instances	25440	25440
Accuracy	90.27 %	94.13 %
Kappa statistic	0.8420	0.9049

The sea level rises in the future were also estimated based on emission scenario Representative Concentration Pathway (RCP) 8.5 [13]. The RCP 8.5 scenario assumes that little coordinated actions are made among countries, so that the climate radiative forcing to the atmosphere from anthropogenic emissions is as high as 8.5 watts per square meter over the globe. The upper 95 % bounds of sea levels are estimated to be 36.3 inches for the 2050s and 45.1 inches for the 2090s. As a result of climate change, the terrain elevation above the sea level is expected to decrease. This will lead to a higher flooding risk and thus the evacuation zone categories need to be updated accordingly. The proposed random forest is used to predict the evacuation zones for the 2050s and 2090s, based on the expected decrease in average elevation above the sea level and assumption that other hurricane-related characteristics are kept the same the future.

⁵ “E1” corresponding to NYC 2013 evacuation zone 1, “E2” corresponding to NYC 2013 evacuation zone 2 and zone 3, and “E3” corresponding to NYC 2013 evacuation zone 4, zone 5 and zone 6, and “S” corresponding to the safe zone beyond the evacuation region.

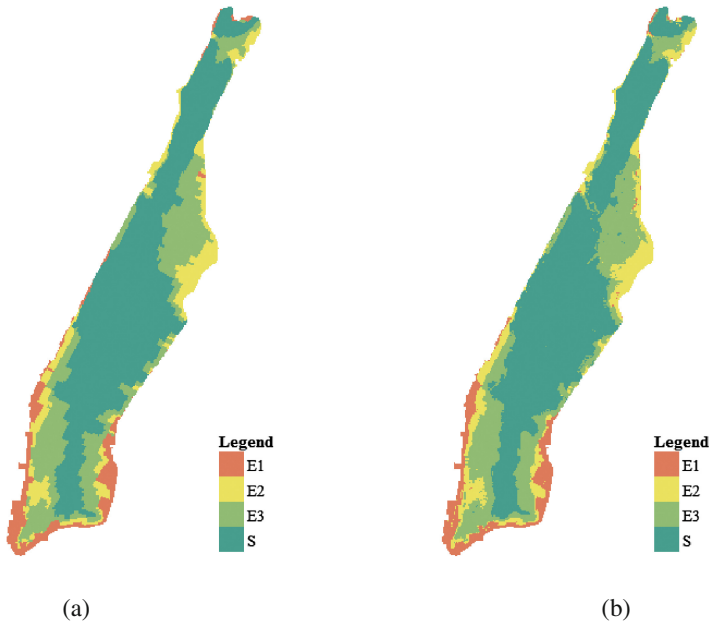


Fig. 2. Current evacuation zones (a) and predicted evacuation zones using the random forest (b).

The predicted future evacuation zones are presented in Fig. 3. Compared with the current zoning, the areas with need of evacuation are expected to expand in the future.

3.3 Traffic Incident Analysis

Incidents are defined here as any occurrence that temporarily reduce highway capacity such as accidents, disabled vehicles and downed trees. Capacity losses caused by incidents are closely related to the incident types, frequencies and durations. The section aims to investigate the characteristics of incidents in the context of hurricane Sandy, and to propose an approach to accommodate the uncertainty of roadway capacities due to incidents.

The incident data used is introduced in subsection “Incident Data” above. As shown in Fig. 4, the proportions of incident types vary greatly between the Sandy week (Oct. 26th, 2012 ~ Nov. 1st, 2012) and the regular time (time intervals before and after the Sandy week). In the Sandy week, the proportions of debris, downed trees, flooding and weather related incidents increased significantly. Meanwhile, there were fewer accidents and disabled vehicles compared with the regular time.

The relationship between incident frequency during the evacuation period of Hurricane Sandy (12 AM, Oct. 26th, 2012–12 PM, Oct. 29th, 2012) and highway characteristics such as road length and traffic volume was investigated. The incident frequency during evacuation for each highway section was obtained. Negative binomial (NB) models can accommodate the nonnegative, random and discrete features of

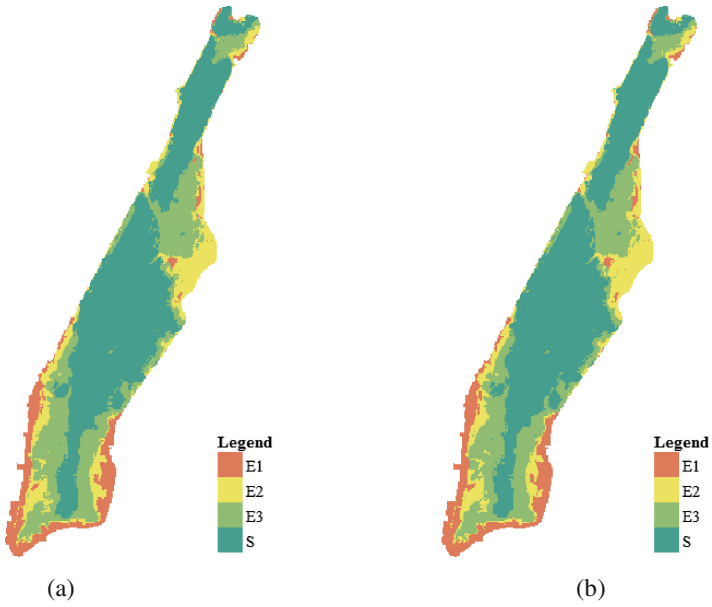


Fig. 3. Predicted evacuation zones for the 2050s (a) and the 2090s (b).

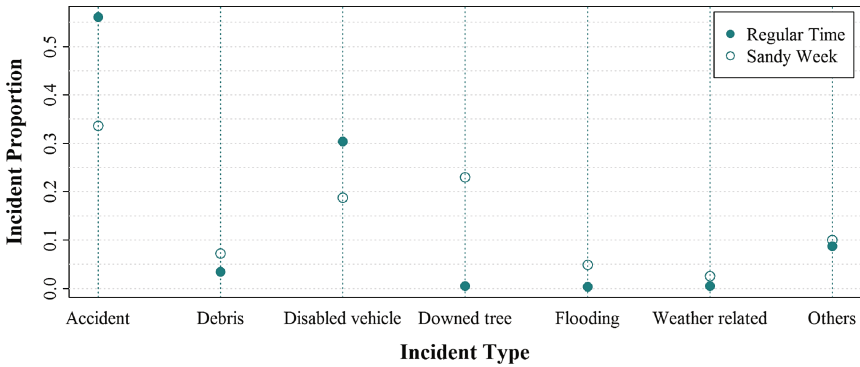


Fig. 4. Proportions of incident types in the regular time and the Sandy week. (Source: Xie et al. (2015) [2])

event frequencies and have been proved better to deal with the over-dispersed data by introducing an error term [14]. A NB model was used to replicate incident frequencies of highway sections, and it can be expressed as follows:

$$\begin{aligned}
 f_i &\sim \text{Negbin}(\theta_i, r) \\
 \ln(\theta_i) &= \alpha X_i
 \end{aligned}
 \tag{3}$$

where f_i is the observed incident frequency for freeway section i , θ_i is the expectation of y_i , X_i is the explanatory variables, α is the vector of regression coefficients to be estimated, and r is the dispersion parameter. Results show that the logarithm of traffic volume and the logarithm of highway length are positively associated with the incident frequencies. In addition, more incidents are expected to happen in interstate highways compared with other highways. The developed incident frequency model can be used to predict the probability of incident occurrence for each highway section in the capacity-loss simulation.

Duration distributions vary for different incident types. The relationship between the incident type and duration can be explored using a lognormal model [2, 15]. A lognormal model assumes a linear relationship between the logarithm of incident durations and explanatory variables. It can be expressed as:

$$\begin{aligned} \ln(d_j) &\sim Normal(\mu_j, \sigma^2) \\ \mu_j &= \beta Z_j \end{aligned} \quad (4)$$

where d_j is the observed duration for incident j , μ_j and σ^2 are the mean and variance of the normal distribution, Z_j is the explanatory variables (dummy variables indicating the incident types), β is the vector of regression coefficients to be estimated. Accidents, debris and disable vehicles are expected to have shorter duration than other incidents; while duration of incidents such as downed tree and flooding tend to be shorter. These modeling results can be used to generate the duration for each incident in the capacity-loss simulation.

The incident type proportions, incident frequency and incident duration models developed are used as inputs for simulating incident-induced capacity losses for the whole study network (40442 links) during the evacuation period. Monte Carlo simulation method is used to generate observations randomly from specified distributions [16]. A detailed simulation procedure to generate capacity losses is introduced in our recent paper [17]. The main steps of this novel approach are summarized as:

- Step 1:* Use the incident frequency model estimate the expectation of incident frequency for each link
- Step 2:* For each incident, generate incident type according to the type proportions during evacuation period
- Step 3:* Use the incident duration model to estimate the duration for each incident

The results of the incident simulation can tell us the likely locations of incidents as well as their types and durations. Based on the incident simulation results, the capacity loss of each link can be estimated and used as inputs in the network-wide evacuation simulation.

3.4 Evacuation Simulation

Simulation of hurricane evacuation is an important task in emergency management/planning. However, this process has to face two challenges: (1) how to estimate

evacuation demand based on socio-economic characteristics and evacuation zone division; and (2) how to deal with the uncertainty due to the roadway capacity losses because of highway incidents. The evacuation simulation model built in this study incorporates most recent hurricane experiences in the New York metropolitan area.

We propose an hour-by-hour evacuation simulation based on a large-scale macroscopic network model of the New York metropolitan area developed in the TransCAD Software [18]. This model reflects the latest traffic analysis zones (TAZs), road network configuration, and socio-economic data. The procedure for the network-wide evacuation simulation is shown on Fig. 5. Prior to traffic assignment, it is crucial to estimate evacuation demand and generate capacity losses for road network. For demand estimation, the first step is to identify the evacuation zones, then estimate the number of people that need to be evacuated based on the socio-economic data. Generated evacuation demand is distributed to each hour according to the empirical evacuation curve obtained from the traffic volumes observed. Unlike most of the previous studies that assume static highway capacities, we attempt to treat the highway capacities to be stochastic, based on the outcomes from incident-induced capacity loss simulation (as described in the previous subsection). The hour-by-hour capacity losses are simulated for the whole network. Three scenarios are developed, including one base and two evacuation scenarios (one considers incident-induced capacity losses and the other doesn't). Under the base scenario, the trip tables are constructed from the background traffic in the regular time, while under the two evacuation scenarios, the trip tables consist of both assumed background traffic and additional evacuation demand.

We run network assignment model using the quasi-dynamic traffic assignment method described in Ozbay *et al.* [19] for each hour based on different scenarios and obtain results including the performance of network links and evacuation times between each O-D pairs of the study network. At last assignment results are analyzed to determine evacuation times from evacuation zones to safe zones and the performance of the network with and without consideration of capacity losses. Figure 6 shows the zonal travel times for two evacuation scenarios and observed taxi trips. It can be seen that travel times for Harlem and downtown areas are lower than Midtown, and travel times for east side of Manhattan is shorter than the east side for all scenarios. Compared

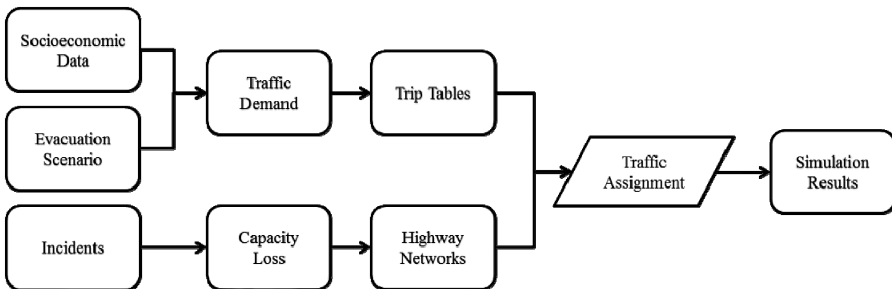


Fig. 5. Network-wide modeling methodology for hurricane evacuation combined with capacity losses due to incidents.

with the scenario with full capacity, the evacuation travel times for capacity loss scenario are significantly higher, and closer to the ones observed from the empirical taxi trip data.

3.5 Resilience Assessment

This subsection evaluates the resilience of roadway and transit systems in the aftermath of hurricanes using large-scale taxi and transit ridership datasets during Hurricanes Irene and Sandy. Recovery curves of subway and taxi trips are estimated for each zone category (evacuation zones 1~6 and safe zone).

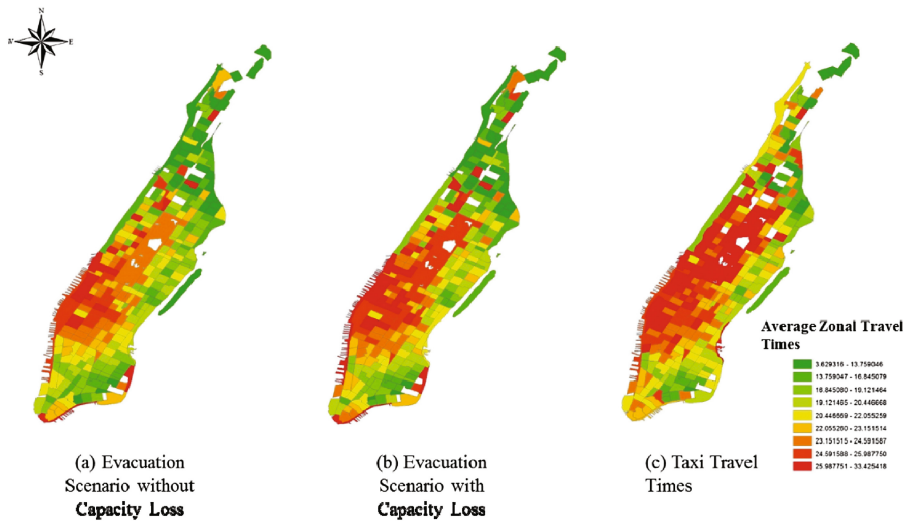


Fig. 6. Zonal travel times of Manhattan for (a) evacuation scenario without capacity losses, (b) evacuation scenario with capacity loss and (c) observed taxi trips.

The logistic function is used in modeling process, since characteristics of logistic model resembles evacuation and recovery activities, which are shown to follow an S-shape. Basic logistic function is shown in Eq. (5):

$$P_t = \frac{1}{1 + e^{-\alpha(t-H)}} \tag{5}$$

where P_t represents zonal recovery rate by time t , α is the factor affecting slope of the recovery rate, and H is half recovery time (the time when half of the lost service capacity is restored). According to Yazici and Ozbay [20], α can be regarded as the parameter that controls behavior of evacuees whereas H controls total clearance time ($2H$). So α and H together can be used to determine two factors of resilience, namely, severity of outcome and time for recovery.

Empirical and model estimated recovery curves are visualized in Fig. 7. For more detailed parameter estimates, please refer to a recently study by Yuan et al. [21]. X axis of each subplot range from 0 to 11, which stands for the days elapsed from hurricane impact to the end of the study period. For Hurricanes Irene and Sandy, starting days are August 28, 2011 and October 30, 2012, respectively. As shown in Fig. 7, during Hurricane Irene, the curves for roadway recovery reached one in two days for nearly all the zones. Full recovery of the subway system took longer than the roadway system for most zones. Compared with Hurricane Irene, Hurricane Sandy recovery for both modes required much longer recovery time. Subway system recovery in the case of Sandy is also slower than roadway system. Spatial patterns are also presented in Fig. 7, roadway curves were not fully recovered at the end of study period for zones 1 to 4. For zone 5, roadway system recovered on day 10, zone 6 and Safe zone recovered on Days 6 and 5, respectively. Subway recovery curves remain flat for high-risk zones. With decreasing rates of zonal vulnerability, subway curves become steeper. For zone 1 (refer to subsection “Evacuation Management Data” for zone division details), only 25 % of subway recovery was completed on day 11. Patterns for all other zones are similar, and subway ridership recovered on day 10 or 11.

The above results show that the process of multi-mode post-hurricane recovery can be captured by using logistic functions. The initial recovery rate of zones which are prone to hurricane-related risk such as zone 1 is lower than those of others, and it takes longer time for such zones for full recovery. Road network is found to have better resilience than subway network, since subway recovery has later initial starting point, lower initial

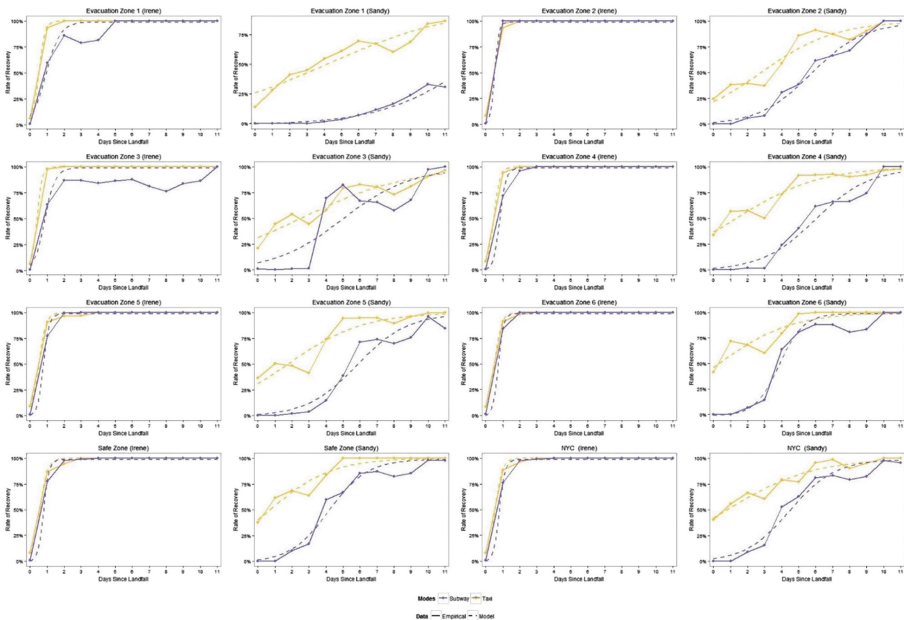


Fig. 7. Empirical and modeled response curves.

percentage and longer recovery period. One of the possible reasons is that failure of one single subway station/line always influences the entire system, whereas this is not the case for the roadway system due to the availability of more alternative routes.

4 Conclusion

This paper provides a comprehensive overview of data-oriented emergency management/planning in the complex urban systems by summarizing five case studies conducted using the big urban data of New York/New Jersey metropolitan area. There are great opportunities for the development of data-driven methods to obtain innovative solutions to the problems of emergency management and planning. The main findings from these case studies conducted by the research team are as follows:

(1) Evacuation behavior analysis

The use of the structural equation modeling is proposed to identify the potential relationship between the evacuation decision and the evacuation destination choices. A weak relationship is found between the evacuation decision and the evacuation destination choices based on the survey data.

(2) Evacuation zone prediction

The random forest has better performance in learning the relationship between current pre-determined evacuation zones and hurricane-related factors. The evacuation zones in the 2050 s and 2090 s are predicted using the random forest and are expected to expand along with the sea level rises.

(3) Traffic incident analysis

It is found that the proportion of debris, downed trees, flooding and weather related incidents increases significantly during the hurricane-impacted period. Based on developed incident frequency and incident duration models, a Monte Carlo simulation method is used to simulate the incident-induced capacity losses for the whole road network during the evacuation period.

(4) Evacuation simulation

An hour-by-hour evacuation simulation model is proposed based on a large-scale macroscopic network model, with consideration of incident-induced capacity losses. Compared with the scenario with full capacity, the evacuation travel times for capacity loss scenario are significantly higher, and are closer to the ones calculated from the historical taxi trip data in the same period.

(5) Resilience assessment

The process of multi-modal post-hurricane recovery can be captured by using logistic functions. The initial recovery rate of evacuation zones which are prone to hurricane-related risk is found to be lower than those of others. It is also found that road network has better resilience than subway network due to its operational, physical and topographical characteristics.

Acknowledgments. The work is partially funded by New York State Resiliency Institute for Storms & Emergencies, Urban Mobility & Intelligent Transportation Systems (UrbanMITS) laboratory, Center for Urban Science and Progress (CUSP), Civil and Urban Engineering at New

York University (NYU) as well as University Transportation Research Center (UTRC) at City College of New York (CUNY). The contents of this paper reflect views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents of the paper do not necessarily reflect the official views or policies of the agencies.

References

1. Gregory, K.: City Adds 600,000 People to Storm Evacuation Zones. <http://www.nytimes.com/2013/06/19/nyregion/new-storm-evacuation-zones-add-600000-city-residents.html>. Access 21 July 2015
2. Xie, K., Ozbay, K., Yang, H.: Spatial analysis of highway incident durations in the context of Hurricane Sandy. *Accid. Anal. Prev.* **74**, 77–86 (2015)
3. Donovan, B., Work, D.: Using coarse GPS data to quantify city-scale transportation system resilience to extreme events. In: Transportation Research Board 94th Annual Meeting, Washington DC (2015)
4. Work, D., Donovan, B.: 2010–2013 New York City taxi data. <http://publish.illinois.edu/dbwork/open-data/>
5. Metropolitan Transportation Authority, MTA turnstile data. <http://web.mta.info/developers/turnstile.html>
6. New York Metropolitan Transportation Council, Best Practice Model. <http://www.nymtc.org/project/bpm/bpmindex.html>
7. Li, J., Ozbay, K.: Empirical evacuation response curve during Hurricane Irene in Cape May County, New Jersey. *Transp. Res. Rec. J. Transp. Res. Board* **2376**(1), 1–10 (2013)
8. Carnegie, J., Deka, D.: Using hypothetical disaster scenarios to predict evacuation behavioral response. In: Proceedings of the Transportation Research Board 89th Annual Meeting (2010)
9. Yang, H., Morgul, E.F., Ozbay, K., Xie, K.: Modeling evacuation behavior under hurricane conditions. In: Transportation Research Board, Washington, DC (2016)
10. Wilmot, C., Meduri, N.: Methodology to establish hurricane evacuation zones. *Transp. Res. Rec. J. Transp. Res. Board* **1922**, 129–137 (2005)
11. Viera, A.J., Garrett, J.M.: Understanding interobserver agreement: the kappa statistic. *Fam. Med.* **37**(5), 360–363 (2005)
12. Xie, K., Ozbay, K., Zhu, Y., Yang, H.: A data-driven method for predicting future evacuation zones in the context of climate change. In: Transportation Research Board, Washington, D.C (2016)
13. Van Vuuren, D.P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G.C., Kram, T., Krey, V., Lamarque, J.-F.: The representative concentration pathways: an overview. *Clim. Change* **109**, 5–31 (2011)
14. Xie, K., Wang, X., Huang, H., Chen, X.: Corridor-level signalized intersection safety analysis in Shanghai, China using Bayesian hierarchical models. *Accid. Anal. Prev.* **50**, 25–33 (2013)
15. Garib, A., Radwan, A., Al-Deek, H.: Estimating magnitude and duration of incident delays. *J. Transp. Eng.* **123**(6), 459–466 (1997)
16. Mooney, C.Z.: Monte carlo simulation, Sage (1997)
17. Zhu, Y., Ozbay, K., Xie, K., Yang, H., Morgul, E.F.: Network modeling of hurricane evacuation using data driven demand and incident induced capacity loss models. In: Proceedings of the Transportation Research Board, Washington, D.C (2016)
18. Caliper, TransCAD - Transportation planning software. <http://www.caliper.com/tcovu.htm>

19. Ozbay, K., Yazici, M., Iyer, S., Li, J., Ozguven, E.: Use of regional transportation planning tool for modeling emergency evacuation: case study of northern New Jersey. *Transp. Res. Rec. J. Transp. Res. Board* **2312**, 89–97 (2012)
20. Yazici, M.A., Ozbay, K.: Evacuation modeling in the United States: does the demand model choice matter? *Transport Rev.* **28**(6), 757–779 (2008)
21. Zhu, Y., Ozbay, K., Xie, K., Yang, H.: Using big data to study resilience of taxi and subway trips for Hurricanes Sandy and Irene. In: *Transportation Research Record*, Washington, D.C (2016)