

2017

Data-Driven Spatial Modeling for Quantifying Networkwide Resilience in the Aftermath of Hurricanes Irene and Sandy

Yuan Zhu

Kun Xie


Kaan Ozbay

Fan Zuo

Hong Yang

Old Dominion University, hyang@odu.edu

Follow this and additional works at: https://digitalcommons.odu.edu/msve_fac_pubs

 Part of the [Emergency and Disaster Management Commons](#), [Infrastructure Commons](#), [Transportation Commons](#), and the [Transportation Engineering Commons](#)

Repository Citation

Zhu, Yuan; Xie, Kun; Ozbay, Kaan; Zuo, Fan; and Yang, Hong, "Data-Driven Spatial Modeling for Quantifying Networkwide Resilience in the Aftermath of Hurricanes Irene and Sandy" (2017). *Modeling, Simulation & Visualization Engineering Faculty Publications*. 8.

https://digitalcommons.odu.edu/msve_fac_pubs/8

Original Publication Citation

Zhu, Y., Xie, K., Ozbay, K., Zuo, F., & Yang, H. (2017). Data-driven spatial modeling for quantifying networkwide resilience in the aftermath of hurricanes Irene and Sandy. *Transportation Research Record*, 2604, 9-18. doi:10.3141/2604-02

Data-Driven Spatial Modeling for Quantifying Networkwide Resilience in the Aftermath of Hurricanes Irene and Sandy

Yuan Zhu, Kun Xie, Kaan Ozbay, Fan Zuo, and Hong Yang

In recent years, the New York City metropolitan area was hit by two major hurricanes, Irene and Sandy. These extreme weather events disrupted and devastated the transportation infrastructure, including road and subway networks. As an extension of the authors' recent research on this topic, this study explored the spatial patterns of infrastructure resilience in New York City with the use of taxi and subway ridership data. Neighborhood tabulation areas were used as the units of analysis. The recovery curve of each neighborhood tabulation area was modeled with the logistic function to quantify the resilience of road and subway systems. Moran's I tests confirmed the spatial correlation of recovery patterns for taxi and subway ridership. To account for this spatial correlation, citywide spatial models were estimated and found to outperform linear models. Factors such as the percentage of area influenced by storm surges, the distance to the coast, and the average elevation are found to affect the infrastructure resilience. The findings in this study provide insights into the vulnerability of transportation networks and can be used for more efficient emergency planning and management.

Hurricanes are one of the biggest natural disaster threats in the northeast corridor of the United States. New York City, which is located in the vulnerable northeast corridor area, experienced two major hurricanes in recent years. On August 2011, Hurricane Irene made landfall in Brooklyn, New York. One year later, Hurricane Sandy landed in New Jersey, south of New York City. As shown in Figure 1, both hurricanes caused the inundation of coastal areas of New York City and resulted in different impact levels on the city's transportation services and infrastructure: Hurricane Irene led to inland flooding and temporary suspension of citywide public transit.

Y. Zhu, Department of Civil and Urban Engineering, 6 MetroTech Center, RH 404A; K. Xie, Department of Civil and Urban Engineering and Center for Urban Science and Progress, 1 MetroTech Center; K. Ozbay, Department of Civil and Urban Engineering and Center for Urban Science and Progress, 6 MetroTech Center, RH 404; F. Zuo, CitySmart Laboratory at Urban Intelligent Transportation Systems Center, 6 MetroTech Center, RH 404A, Tandon School of Engineering, New York University, Brooklyn, NY 11201. H. Yang, Department of Modeling, Simulation, and Visualization Engineering, Batten College of Engineering and Technology, Old Dominion University, 4700 Elkhorn Avenue, Norfolk, VA 23529. Corresponding author: Y. Zhu, yuan.zhu@nyu.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2604, 2017, pp. 9–18.
<http://dx.doi.org/10.3141/2604-02>

Since most of the infrastructure was intact, public transit was back to normal 1 day after landfall. Hurricane Sandy, however, turned into one of the most costly natural disasters in the city's recent history. Unlike with Hurricane Irene, several subway stations and tunnels were flooded, especially those located in Lower Manhattan and Coney Island (1). Although the Metropolitan Transportation Authority restored half of the major services within a week after landfall, it took several months for stations seriously damaged to be fully functional because of the mass erosion of the power supply and tube structure by saltwater. Both hurricanes also caused disruption to and destruction of the highway network. Major bridges and tunnels were closed, and several tunnels were flooded during Hurricane Sandy.

After the disruption and devastation caused by the hurricanes, researchers started to show an increasing interest in strengthening the city infrastructure to avoid, or at least to mitigate, the effect of future coastal storms. It was, therefore, necessary to evaluate the resilience of roadway and transit networks as measured by their vulnerability to storm surge. The current six-category evacuation zone system based on New York City's hurricane contingency plan identifies possible impacts on the city districts (2). A recent study by the authors of this paper (3–5) explored the recovery patterns of highway and subway networks and developed multilayer models for evacuation zones in New York City (3). In this paper, logistic curves—which are frequently used for evacuation demand modeling—were used for recovery modeling. Results showed a clear relationship between recovery patterns and evacuation zone characteristics, and it seemed plausible that the road network was more resilient than the subway system. However, since zones of the same category are widely distributed, it is difficult to quantify different levels of impact on areas in the same category, and it is not trivial to distinguish damage caused to highway or subway networks separately.

As a follow-up to the previous paper in which the analysis was done in relation to the evacuation zones of New York City, the goal of this study was to model the resilience of roadway and transit systems in relation to the city's individual neighborhoods and to conduct statistical spatial analysis to explore the intercorrelation of zonal resilience (3). In addition, this study explored the resilience of the same network for two events, namely, Hurricanes Sandy and Irene. Compared with previous models based on evacuation zones, the new models can better reveal the spatial distribution of recovery characteristics and make it possible to predict the resilience of highway and transit networks according to the geographic location and hurricane intensity.

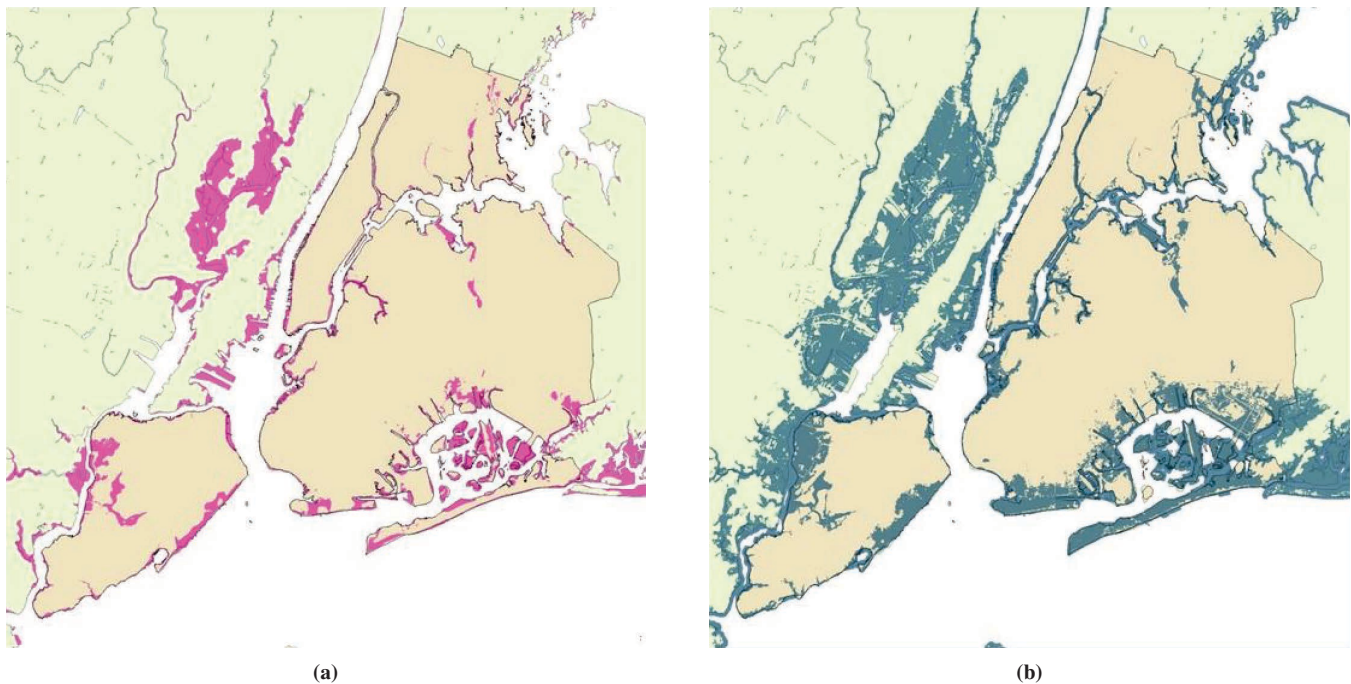


FIGURE 1 Areas influenced by storm surges during Hurricanes Irene and Sandy in New York City: (a) Hurricane Irene and (b) Hurricane Sandy (23).

LITERATURE REVIEW

Transportation infrastructure—including road networks, subway stations, and tunnels alike—is faced with disruptions caused by natural disasters such as hurricanes. In recent years, researchers have begun to show interest in the ability of transportation systems to withstand and recover from the disruptions, and the concept of resilience was introduced. Heaslip et al. pointed out two key factors of resilience: How can the system maintain a demonstrated level of service? And how long does it take for a system to restore to a demonstrated level of service (6)? Similarly, Bruneau et al. introduced the “resilience triangle” to quantify three key issues of resilience: the possibility of failure, the severity of outcome, and the duration of recovery (7). They defined the area of the triangle as loss of resilience (LoR), which can be mathematically represented by Equation 1:

$$\text{LoR} = \int_{t_0}^{t_1} [100 - Q(t)] dt \quad (1)$$

where $Q(t)$ is the time-dependent quality of the infrastructure (7). Therefore, LoR can be determined by the depth of the initial disruption and the speed of quality restoration, as the key issues stated above.

Transportation System Resilience

Testa et al. measured the resilience of the highway network of the metropolitan area of New York City by testing the topological graph properties under various scenarios of link removal (8). According to Donovan and Work, the New York City taxi data set can be used to measure roadway resilience of the city during Hurricane Sandy by measuring the deviation of normalized travel times between four regions of the city (9).

Hosseini and Barker used the Bayesian network approach to quantify resilience as a function of adaptive and restorative capacities, and the model was demonstrated in a case of inland waterway ports (10). Adjetey-Bahun et al. developed a simulation-based model to quantify resilience of the mass railway transit system in Paris (11). The model evaluates system resilience during perturbation by quantifying passenger delay and load. Simulation results indicated the resiliency of the system, which is consistent with observations. D’Lima and Medda used a mean-reverting stochastic model to explore daily fluctuations of the London Underground in regard to subway lines (12).

Logistic functions, as first proposed by Belgian mathematician Pierre-Francois Verhulst in 1838 to analyze population growth in Belgium, were widely used in pre- and posthurricane studies (13). The concept of the S-curve was introduced by Lewis to describe the evacuation pattern before hurricanes (14). Hobeika et al. suggested the use of the logistic curve based on behavior research (15). Fu et al. used a post-Hurricane Floyd survey of South Carolina to model the evacuation response curve (16). The same models were proved to be effective to estimate evacuation demand in Hurricane Andrew. Li et al. used traffic count data of Cape May County, New Jersey, during Hurricane Irene to build an empirical response curve, which showed a better fit with the logistic function (17). The logistic function was also used as a demand generation approach by Ozbay and Yazici (18).

Spatial Analysis of Transportation Networks

Spatial analysis is widely used in the safety assessment of transportation networks. Tasic and Porter built an areawide model for Chicago to evaluate the spatial association of safety issues and multi-model transportation infrastructure and found a strong relationship between crashes and the availability of transportation service (19).

Xie et al. developed an incident duration model for Hurricane Sandy and confirmed spatial dependencies of durations of neighboring incidents (20). Spatial error and spatial lag models were further developed to indicate factors that affect the duration of an incident.

In this paper, previously proposed methods of resilience quantification and logistic modeling are used for New York City by subdividing the city into small units based on neighborhood tabulation areas (NTAs). Then, factors affecting recovery patterns and resilience are identified and analyzed. Based on results of this highly detailed spatial resilience modeling approach, spatial dependence tests and further statistical modeling efforts are made to study resilience characteristics for roadway and subway systems of New York City for two hurricanes.

DATA

To analyze the resilience of New York City's highway and transit networks, two types of data sets are used. One is the city's taxi trips data, which were made available by the New York City Taxi and Limousine Commission (9, 21). This data set contains taxi trips from 2010 to 2013. Each trip record includes time and location information of pickups and drop-offs. The second data set consists of subway ridership data obtained from the data feed of the Metropolitan Transportation Authority; this turnstile data set is stored in individual weekly text files containing hour-by-hour counts along with other related spatiotemporal information (22). Each row in the weekly file contains a record of entry and exit counts as well as the remote unit (station) and control area (turnstile) that the counter belongs to. In normal situations, counter readings of each turnstile are recorded every 4 h, but the time of the reading differs for the stations. To obtain the ridership for each subway station, it is necessary to convert counter readings to turnstile ridership by subtracting the last and first reading of a day and then calculating the sum of all turnstiles. Although the subway data set has the Staten Island Railway fields, insufficient records are

found in the study periods; therefore, the Staten Island transit network is excluded from the analysis.

Since the aim was to track recovery patterns for both hurricanes, 12 days after landfall was chosen as the study period. Specifically, August 28 to September 8, 2011, for Hurricane Irene, and October 29 to November 10, 2012, for Hurricane Sandy. For comparison purposes, data sets of the same periods of previous years are used. Since traffic in New York City has a significant day-of-the-week pattern, days closest to the days of the week in the study period are used.

The taxi and subway data sets include noisy and erroneous records, and it is crucial to select the appropriate part and filter the data. According to Donovan and Work, there are significant errors in the taxi data set, including missing or unrealistic coordinates and impossible travel times or speeds (9). For subway trips, errors include extremely low or high ridership values, which are caused by counter reset as a result of maintenance; these need to be filtered out.

Other data sets used in this study include a socioeconomic demographic data set for New York City obtained from the U.S. Census Bureau (Source: <http://factfinder.census.gov>), surge area data for both hurricanes from the Federal Emergency Management Agency (23), and New York City elevation data (24). For modeling purposes, these data sets were further featured into NTA levels (25). Table 1 presents the description and descriptive analysis of key variables. The explanatory variables are grouped into three categories: geographic, socioeconomic, and transportation. The computation of dependent variables listed in Table 1 is introduced in the next section.

MODELING NEIGHBORHOOD-BASED RECOVERY PATTERNS

The main objective of this section is to propose recovery models and identify coefficients for all neighborhoods and, then, find spatial correlations of the model parameters. Travel modes and weather events are modeled separately.

TABLE 1 Description and Descriptive Analysis of Key Variables ($N = 195$)

Variable Group	Description	Mean	SD
Dependent variable			
TI_LoR	LoR for the taxi system during Hurricane Irene	0.447	0.393
SI_LoR	LoR for the subway system during Hurricane Irene	0.855	0.597
TS_LoR	LoR for the taxi system during Hurricane Sandy	0.858	1.189
SS_LoR	LoR for the subway system during Hurricane Sandy	4.787	2.063
Geographic			
Near_Dist	Distance to coast (10^3 ft)	5.617	4.251
Elevation	Average elevation (ft)	78.970	36.367
Pct_Surge	Percentage of area influenced by storm surges	0.107	0.192
Manhattan	1 if in Manhattan, 0 otherwise	0.149	0.357
Brooklyn	1 if in Brooklyn, 0 otherwise	0.262	0.441
Queens	1 if in Queens, 0 otherwise	0.297	0.458
Bronx	1 if in the Bronx, 0 otherwise	0.195	0.397
Socioeconomic			
Population	Total population in 2010 (10^3)	42.047	22.484
Edu_Bac	Population with bachelor's degree or higher (10^3)	9.704	10.117
Avg_Income	Average income (10^3 \$)	73.994	35.890
Employment	Number of employed (10^3)	19.371	11.457
Schools	Number of schools	14.056	10.011
Roads_Mi	Length of roadways (mi)	48.083	28.431
Veh_Own	Number of families with private vehicles (10^3)	6.992	3.903
Transportation			
Sub_Time	Subway access time (min)	16.771	16.749
Bus_Stop	Number of bus stops	66.323	41.630

Using NTAs as Units of Analysis

First, the processed data sets of taxi trips and subway ridership will be mapped into subareas of New York City; thus, it is necessary to determine the unit of study from the very beginning. In this paper, neighborhoods of New York City with reference to NTAs are used as the geographic modeling units. The NTA is a set of polygons created by the New York City Department of City Planning and used to represent data from the Census and American Community Survey (25). Overall, there are 195 NTAs in New York City, and each NTA corresponds to one neighborhood with a unique ID and name. Compared with evacuation zones, there are two advantages of selecting NTAs. First, the NTA sizes are appropriate for the analysis, especially for subway data, because these areas are neither so big that they may cover more than one evacuation zone category, nor too small that they may not include even one subway station. Second, as mentioned above, unlike traffic analysis zones or census tracts, each NTA also has a familiar name, so it is much easier to follow the travel patterns with NTAs.

Outputs of the first step are the daily taxi trips and the subway ridership of each NTA for each hurricane and study period; these were later converted into time-dependent recovery rates. The rate of recovery is defined as the quotient of trips during a certain hurricane period divided by trips during a corresponding normal (control) period. The recovery rates of a 12-day period for all NTAs are calculated. Then, recovery rates are processed to conform to satisfy prerequisites of the logistic model. Values greater than one are rounded to one. Also, if the recovery rate reaches one, it is assumed that the area has already been recovered; then, the recovery rate is kept as one for the rest of the study period.

For New York City, most of the taxi trips are located in Manhattan, downtown Brooklyn, densely populated areas in Queens and the Bronx, and major airports. For other neighborhoods farther away from these areas, there are far fewer taxi trips. Also, subway service is not available in all of the NTAs. Therefore, NTAs with no data availability for specific travel modes are filtered out.

Modeling Resilience for Each NTA

This section briefly describes the functional form used for modeling recovery rates for each NTA, the performance of model calibration efforts, and the definition of zonal resilience. A more detailed discussion on this specific method is provided in Zhu et al. (3).

The basic logistic function is used for modeling evacuation curves, as shown in Equation 2:

$$P_t = \frac{1}{1 + e^{-\alpha(t-H)}} \quad (2)$$

where

- P_t = recovery rate of area by time t ,
- α = factor affecting slope of recovery rate, and
- H = half recovery time; in other words, the time system reaches half of service capacity.

Therefore, α and H can determine the shape of the S-curve, which reflects the recovery behavior and the resilience for each NTA.

The nonlinear least squares error (LSE), as shown in Equation 3, is used to fit the model by comparing the difference between the modeled function and empirically obtained data points.

$$\text{LSE} = \sum_{t=t_0}^{t_1} (y_t - P_t)^2 \quad (3)$$

where y_t is the observed recovery rate of day t and P_t is the logistic function (Equation 2). The values of t_0 and t_1 are 0 and 11. The objective is to minimize LSE, the difference between observed and estimated recovery rates. For subway and taxi trips in each NTA, distinct pairs of model parameters (α and H) are calibrated to minimize S .

Another critical factor that needs to be identified is LoR, which can be calculated with the abovementioned model in Equation 1 (7). With the use of the logistic function P_t to replace Q_t , Equation 1 can be rewritten as

$$\text{LoR} = \int_{t_0}^{t_1} \left[1 - \frac{1}{1 + e^{-\alpha(t-H)}} \right] dt \quad (4)$$

where LoR is the loss of resilience from the time of the original hurricane impact; LoR is the area enclosed by the logistic function, y axis, and line $x = 1$ (100%).

The logistic functions are built for most areas, except for the following situations: (a) the recovery rates of the entire study period are one (in this case, the area was not affected by the storm surge and LoR is zero) and (b) the recovery rates of the period are zero, which was the case in the transit networks of certain NTAs in Hurricane Sandy, in which subway restoration took longer than the study period. The LoR was maximum, therefore, and the value was 11, in that case.

Empirical Analysis of Resilience

Since this study covered four recovery patterns of two networks for two distinct weather events and each of them contained submodels of most NTAs, it is not practical to show this multilayer model in a table format. Instead, recovery characteristics are visualized on a map of New York City with NTAs to show the three above-mentioned critical factors (α , H , and LoR), and each row has maps of four recovery scenarios [Irene highway (taxi), Sandy highway (taxi), Irene subway, and Sandy subway]. To be able to provide a side-by-side comparison, subplots are created with the same scale for four scenarios. Another point worth mentioning is the selection of a color gradient. Plots of all four terms use a green and red gradient, but the colors of the start and end points varied for the terms, and a greener plot always stands for a better recovery situation or the goodness of fit of the models. For α , as a higher value stands for a steeper slope of recovery function, greener colors are used for the higher values. For H , the gradient is from green to red as H is an indicator of recovery time. LoR plots use red to show higher values, which stand for a worse curve fit of the empirical data, and a higher loss of resilience. As mentioned above, neighborhoods with no data input are excluded from the modeling, as shown in gray in the figures.

Parameter α from the modeling results is visualized in the first row of Figure 2. It can be seen that highway and transit networks have higher α values for Hurricane Irene compared with Hurricane Sandy, implying a faster speed (lower travel time) in the aftermath of Hurricane Irene. In Figure 2, *a* and *b*, the α values of Manhattan and the coastal neighborhoods are lower than those from the inland neighborhoods. In addition, for Hurricane Irene, most of the inland areas in Brooklyn and Queens are green, whereas only a small

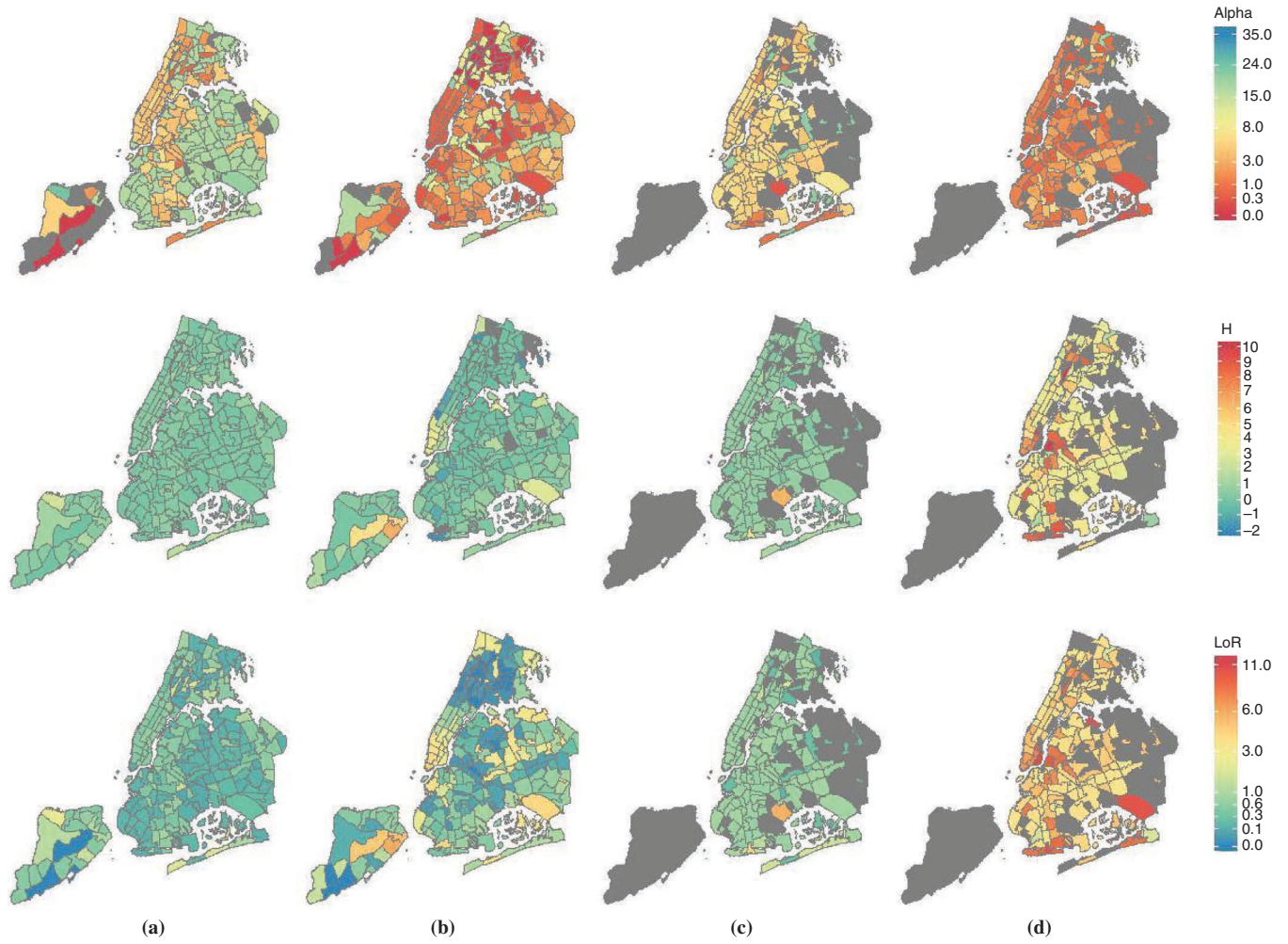


FIGURE 2 Parameter alpha (slope of recovery rate), H (time recovery reaches half of service capacity), and LoR: (a) Irene highway (taxi), (b) Sandy highway (taxi), (c) Irene subway, and (d) Sandy subway.

proportion of those areas is shown in green for Hurricane Sandy. It can be inferred that the magnitude of the disruption of the highway network according to taxi data is greater for Hurricane Sandy. Compared with the recovery of the highway network, the α values of the subway network are relatively lower for both hurricanes. Also, in Figure 2c, the α values of most areas are similar, except for a few NTAs, in which subway stations or depots suffered from the storm surge. The values for Hurricane Sandy for the entire city are significantly low, as shown from the wide range of red-colored zones in Figure 2d.

Parameter H is shown in the second row of Figure 2. As mentioned above, H stands for the time that the network recovery reaches half of the service capacity; therefore, a lower value of H implies a shorter recovery time. According to Figure 2, a and c , during Hurricane Irene, H values are below 1 and nearly identical for most NTAs. That finding means that highway and transit networks were back to full capacity in 2 days after the Hurricane Irene landfall because of the limited impact of that storm. For Hurricane Sandy, as expected, the subway network has much higher H values. However, the highway network in some neighborhoods tends to have lower H values than for Hurricane Irene. Particularly, certain NTAs in the Bronx have negative H values. The negative value of H means that the initial recovery

rate of the NTA is already greater than 50%. One possible reason for that outcome is that these areas were not affected by the hurricane, as shown in Figure 1a. However, because of the suspension of subway service, more travel demand might have been diverted to the taxi mode.

The hurricane-induced LoR values are shown in the third row of Figure 2. It can be observed that both networks were quite resilient during Hurricane Irene, compared with the high LoR values in Hurricane Sandy. The overall LoR values for the taxi data tend to be lower than those based on the subway data, which were given in the conclusion section of the previous study (3). In addition, the distribution of the LoR values appears to be more spatially correlated for the highway network. As shown in Figure 2b, neighborhoods located in the Bronx are found to be more resilient than those in Manhattan and Brooklyn. Also, from uptown to downtown Manhattan, the LoR values gradually increase. The south tip of Manhattan has the highest LoR values; that finding is consistent with the map of the Sandy surge zones presented in Figure 1. The resilience of subway ridership is not as correlated spatially as it is for taxi trips. However, the LoR for zones with damaged critical subway infrastructure is still significantly higher, such as those in Lower Manhattan or Coney Island.

SPATIAL ANALYSIS OF HURRICANE-INDUCED LOSS OF RESILIENCE

Spatial Dependence Test for LoR

From Figure 2, the spatial clustering of LoR can be visually observed. To quantitatively analyze the spatial dependence of LoR, the Moran's I test proposed by Moran (1948) was conducted (26). Given its simplicity and intuitiveness, the Moran's I test has been widely used to measure the spatial autocorrelation of continuous observations (27–30). The test was used in a recent study by Xie et al. to measure the spatial dependence of highway incident durations (20). The Moran's I in matrix form is defined as in Xie et al. in Equation 5 (20):

$$I = \left(\frac{N}{S_0} \right) \left(\frac{\mathbf{d}'\mathbf{W}\mathbf{d}}{\mathbf{d}'\mathbf{d}} \right) \quad (5)$$

where

- \mathbf{d}' = transpose vector of \mathbf{d} ,
- \mathbf{d} = vector of deviations of LoR values from mean,
- \mathbf{W} = spatial weights matrix between each pair of NTAs,
- N = total number of NTAs, and
- S_0 = aggregation of spatial weights $\sum_{i=1}^N \sum_{j=1}^N w_{ij}$.

If the distance between the centroids of NTAs i and j is less than the threshold distance, the spatial weight w_{ij} is defined by the inverse distance between them. Otherwise, the spatial weight w_{ij} is set to be 0. The minimum threshold distance that could ensure that all NTAs have at least one neighbor was used (31).

The pseudo p -value obtained from the permutation test is recommended to assess the significance of Moran's I (32). The pseudo- p -value is defined as $M + 1/S + 1$, where M is the number of instances with Moran's I equal to or greater than that of the observed data and S is the total number of permutations. A total of 999 permutations were performed to compute the pseudo p -value.

The results of Moran's I tests for highway and subway systems during Hurricanes Irene and Sandy are presented in Table 2. See Xie et al. for definitions of statistics $E[I]$, $SD[I]$, and z_I (20). It is found that all pseudo p -values are less than .05, and thus the spatial dependence of LoR can be confirmed. If spatial dependence is neglected in estimating LoR, biased statistical inferences will result.

Interpolating Missing LoR

To build the spatial model, the missing values in the input data have to be interpolated. The main task is to estimate the missing LoR

TABLE 2 Results of Moran's I Tests

LoR Value	I	$E[I]$	$SD[I]$	z_I^a	Pseudo p -Value
TI_LoR	0.1176	-0.0052	0.0035	3.5273	.0070
TS_LoR	0.1138	-0.0052	0.0310	3.8025	.0060
SI_LoR	0.3184	-0.0052	0.0345	9.3733	.0010
SS_LoR	0.0093	-0.0052	0.0209	4.7621	.0050

NOTE: $E[I]$ = expectation of I ; $SD[I]$ = standard deviation of I ; z_I = spatially clustered distribution of the observations; TI = taxi Irene; TS = taxi Sandy; SI = subway Irene; SS = subway Sandy.

^aPositive values indicate that the distribution of the observation are spatially clustered.

values in subway data. Typically, if there is no subway station in one NTA, travelers tend to use the stations in nearby neighborhoods, and their choices of stations are directly related to the distance. Therefore, for an NTA without direct subway service, its resilience could be represented by those of all nearby stations. The inverse distance weighting method is used to interpolate missing LoR data. The function of inverse distance weighting is specified in Equation 6.

$$f(x) = \frac{\sum_i w_i(x) y_i}{\sum_i w_i(x)} \quad w_i(x) = \left(\frac{1}{\|x - x_i\|} \right)^p \quad (6)$$

where x_i are points with LoR values y_i . The default value of exponent p is 2; however, to avoid the bulls-eye effect (see http://www.gitta.info/ContiSpatVar/en/html/Interpolatio_learningObject2.xhtml), a value of 1 is used.

Inverse distance weighting can be used only for a point where missing values are surrounded by known values. Missing values not between two observations (particularly NTAs adjacent to Nassau County of Long Island) cannot be interpolated. Instead, it is assumed that subway riders would go to the nearest NTA with subway service. Therefore, the resilience of such zones is assumed to be the same as that of the nearest accessible NTA.

Spatial Modeling of LoR

In this section, the linear model, the spatial error model, and the spatial lag model are proposed to estimate the LoR. The maximum likelihood estimation method is used for model calibration. See Xie et al. for more details on model specification and estimation (20).

Linear Model

A linear relationship is assumed between LoR and explanatory variables. In matrix form, it can be expressed as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I}) \quad (7)$$

where

- \mathbf{y} = vector of LoR values;
- \mathbf{X} = vector of explanatory variables, such as surge percentage, average elevation, and population;
- $\boldsymbol{\beta}$ = vector of regression coefficients to be estimated;
- σ^2 = variance; and
- \mathbf{I} = identity matrix.

In the linear model, the error term $\boldsymbol{\varepsilon}$ is assumed to be independent and identically distributed with mean zero and a constant variance.

Spatial Error Model

In the spatial error model, spatial dependence is captured via spatial error correlation (omitted variables at one site can affect the dependent variable itself and its neighboring sites). The spatial error model in matrix form can be specified as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \lambda \mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I}) \quad (8)$$

In the spatial error model, the overall error is represented by two components, namely, $\boldsymbol{\varepsilon}$, which is a spatially uncorrelated error term, and \mathbf{u} , which is a spatially dependent error term. The spatial autoregressive parameter λ indicates the extent to which \mathbf{u} of observations are correlated.

Spatial Lag Model

In the spatial lag model, spatial dependence is captured through spatial error correlation effects and spatial spillover effects (observed variables at one site can affect the dependent variable itself and its neighboring sites). The spatial lag model in matrix form can be specified as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \rho \mathbf{W}\mathbf{y} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I}) \quad (9)$$

where $\rho \mathbf{W}\mathbf{y}$ is a spatially lagged dependent variable, ρ is a spatial autoregressive parameter, and the rest of the notation is as before. The assumption of error term $\boldsymbol{\varepsilon}$ is the same as that in the linear model.

Model Assessment

To measure the goodness of fit of a model R^2 is generally used (33). However, since residuals of spatial models are not independent of each other, it is not appropriate to compare spatial models with R^2 . Instead, criteria-based likelihood estimation methods can be used, such as maximum likelihood and Akaike information criterion (AIC) developed by Akaike (34) or the Bayesian information criterion (BIC) first proposed by Schwarz (35). Equations 10 and 11 specify the terms AIC and BIC:

$$\text{AIC} = -2\text{LL}_{\max} + 2k \quad (10)$$

$$\text{BIC} = -2\text{LL}_{\max} + k \ln(N) \quad (11)$$

where

LL_{\max} = maximum log likelihood obtained according to Xie et al. (20),
 k = parameter number, and
 N = sample size.

If the AIC and BIC differences between two models are greater than 4, then the two models can be regarded as considerably different; differences greater than 10 provide strong evidence that the model with a lower AIC and BIC should be favored (36, 37).

Results of LoR Models

Results of three modeling strategies in relation to R^2 , AIC, and BIC are displayed in Table 3. According to Table 3, spatial error and spatial lag models have greater R^2 values compared with classic linear modeling. However, as mentioned above, because of the dependence of residuals, R^2 should be used with caution. The likelihood-based criteria of AIC and BIC are presented as well. For scenarios of highway (taxi) Irene, subway Irene, and subway Sandy, BIC differences are greater than 4; that result means the spatial error model is considerably better than the spatial lag model. It also indicates that the spatial autoregressive process occurs mainly in the error term. It can be seen from Table 3 that models estimating LoR during Sandy perform better than those of Irene. It can be revealed that the spatial correlation of LoR is stronger in Hurricane Sandy than in Hurricane Irene. The modeling results of the taxi network are also better than those of the subway. Overall, the behavior of each model is consistent with the findings of the empirical analysis presented in the paper.

Table 4 shows the modeling results of spatial error and spatial lag models. The autoregressive parameters λ in the spatial error model and ρ in the spatial lag model are also reported. The selected factors for modeling vary on four occasions, and Pct_Surge is found to be the major contributor for all four scenarios. The spatial error model is used to evaluate effects of the variables. For interpreting the signs of the coefficient in Table 4, a positive sign implies an expected increase in LoR, whereas a negative sign suggests an expected decrease. According to Tavassoli Hojati et al., the coefficient exponents can be used to measure the percentage change in the dependent variable with one unit change of the explanatory variables (38).

According to spatial error and spatial lag models shown in Table 4, in all four occasions, the taxi LoR values during Hurricane Irene are positively related to Pct_Surge because the human activity and the service status of the infrastructure were directly affected by the landfall. As shown in Table 4, the taxi LoR values in Hurricane Irene are also positively determined by Sub_Time; the probable reason is that in areas far from transit service, people rely more on taxi service, and then the lack of alternative modes causes less resiliency in service recovery. The zone signs in Queens and Brooklyn are negatively related to the LoR values; that finding implies that the taxi ridership of the two boroughs was more resilient during Hurricane Irene. But that conclusion applies only to Hurricane Irene, considering the limited impact it had on the two boroughs. According to Table 4,

TABLE 3 Model Comparisons

Statistic Measure	TI_LoR			SI_LoR			TS_LoR			SS_LoR		
	Linear	Spatial Error	Spatial Lag	Linear	Spatial Error	Spatial Lag	Linear	Spatial Error	Spatial Lag	Linear	Spatial Error	Spatial Lag
R^2	.138	.162	.152	.115	.118	.115	.348	.354	.359	.290	.292	.292
AIC	169.344	165.726	169.085	252.666	252.293	254.632	548.389	547.062	547.715	653.923	653.522	655.615
BIC	185.709	182.091	188.723	278.03	277.657	283.166	568.027	566.699	570.626	669.776	669.374	674.638

TABLE 4 Modeling Results of LoR

Variable	Taxi						Subway					
	Spatial Error			Spatial Lag			Spatial Error			Spatial Lag		
	Coefficient	SE	p-Value	Coefficient	SE	p-Value	Coefficient	SE	p-Value	Coefficient	SE	p-Value
Hurricane Irene												
Constant	0.4766	0.0367	<.0001	0.5897	0.0912	<.0001	0.9619	0.1431	<.0001	0.934	0.2111	<.0001
Pct_Surge	0.4119	0.3067	.1792	0.4097	0.3318	.2169	0.7172	0.5383	.1828	0.5893	0.542	.2769
Queens	-0.3145	0.0518	<.0001	-0.3656	0.075	<.0001	na	na	na	na	na	na
Brooklyn	-0.1558	0.0504	.0202	-0.1875	0.0675	.0055	na	na	na	na	na	na
Sub_Time	0.0052	0.0014	.0002	0.0061	0.0018	.0006	na	na	na	na	na	na
Near_Dist	na	na	na	na	na	na	-0.0108	0.0089	.2222	-0.0113	0.0093	.2273
Elevation	na	na	na	na	na	na	-0.0026	0.0011	.0158	-0.0025	0.0012	.0355
Veh_Own	na	na	na	na	na	na	0.0443	0.0142	.0018	0.0447	0.0148	.0025
Roads_Mi	na	na	na	na	na	na	-0.0032	0.0017	.0544	-0.0032	0.0017	.0613
Population	na	na	na	na	na	na	-0.0036	0.0026	.1587	-0.0037	0.0027	.1591
Bus_Stop	na	na	na	na	na	na	0.0019	0.0011	.074	0.0018	0.0011	.0965
λ	-0.317	0.1663	.0566	na	na	na	-0.1281	0.1837	.4854	na	na	na
ρ	na	na	na	-.2335	.1578	.139	na	na	na	.0326	.1634	.8418
Hurricane Sandy												
Constant	-0.1539	0.2758	.5768	-0.2051	0.2623	.4343	4.6873	0.415	<.0001	4.3525	0.8037	<.0001
Pct_Surge	1.0737	0.4244	.0114	1.0173	0.4099	.0131	4.4688	0.7079	<.0001	4.34	0.7374	<.0001
Near_Dist	-0.0238	0.0194	.2474	-0.0193	0.0179	.2827	na	na	na	na	na	na
Population	-0.0097	0.0032	.0026	-0.0089	3.18 E-06	.0052	na	na	na	na	na	na
Avg_Income	0.0153	0.0021	<.0001	0.0143	0.0021	<.0001	na	na	na	na	na	na
Roads_Mi	0.0061	0.0026	.0169	0.005	0.0025	.0469	na	na	na	na	na	na
Elevation	na	na	na	na	na	na	-0.0028	0.0039	.476	-0.0024	0.0038	.5211
Veh_Own	na	na	na	na	na	na	-0.1491	0.0417	.0004	-0.1444	0.0421	.0006
Employment	na	na	na	na	na	na	0.0393	0.0146	.00707	0.0378	0.0145	.0093
λ	0.1421	0.1435	.3221	na	na	na	0.0941	0.1676	.5743	na	na	na
ρ	na	na	na	.1689	.1169	.1486	na	na	na	.069	.1448	.6336

NOTE: SE = standard error; na = not applicable.

the values of LoR for the highway network during Hurricane Sandy were positively related to Avg_Income and Roads_Mi. This is an interesting finding showing that LoR values are also related to the zonal income level and the roadway density. Normally, the areas of higher average income in New York City are located either in uptown Manhattan or areas in other boroughs with a considerably lower density (such as Dyker Heights, Brooklyn), where residents prefer to use a taxi for travel, so an extreme event could have a more significant effect on taxi trips in such areas. The way Avg_Income affects LoR values can be explained by the fact that a hurricane might cause greater disruption to areas with longer roadway mileage.

The LoR values of the subway network in Irene is found to be negatively related to Near_Dist, Elevation, and Population, which can be seen in Table 4. The first two are direct indicators of vulnerability to storm surge. If the area is near the shore or if the elevation of an area is low, it is prone to hurricane landfall. Also the model reveals that NTAs with a higher population tend to have higher transit resilience, probably because of the high priority of system recovery. According to Table 4, transit resilience after Hurricane Sandy is positively related to employment and negatively related to Near_Dist and Veh_Own. The relationship between LoR and employment shows the relationship between resilience and land use. Hurricane Sandy did have a significant effect on commercial areas, such as Lower Manhattan, and caused severe business activity disruptions. In addition, subway network resilience is related to auto ownership, as areas with higher auto ownership are also more resilient in regard to transit, which is partially related to the fact that residents do not have to rely on public transit or to the insignificance of public transit as an alternative mode of travel. It is noticeable that the Veh_Own for Hurricane Irene is positively related to LoR. The main reason for the inconsistency may be that Hurricane Irene actually did not cause much damage to the system, so the system was immediately restored in the aftermath of the hurricane. It is reasonable to conclude that auto ownership affects LoR in a negative way.

CONCLUSION

In this study, an NTA-based statistically robust spatial model was proposed to identify characteristics of the recovery patterns for highway and subway networks in New York City. One major contribution of this study was the introduction of the notion of spatial dependence, which complements the empirical analysis of recovery patterns presented in the previous paper (3). Also, the estimated recovery models were built to represent the spatiotemporal recovery patterns with the use of the logistic function with two parameters, with which the loss of resilience (LoR) of each NTA could be calculated. Compared with evacuation zone-based modeling, neighborhood-based models can provide more detailed information about the variations in recovery behaviors. Moreover, instead of six logistic functions estimated for six evacuation zones in Zhu et al., the improved spatiotemporal model had 195 NTAs and corresponding recovery curves for both hurricanes (3). This new approach made it possible to conduct a comprehensive spatial analysis. The empirical analysis of modeling results demonstrated that the values of the estimated model parameters α , H , and LoR varied greatly by individual storms, transport modes, and spatial locations. The higher spatial clustering of resilience was observed during Hurricane Sandy, which had greater intensities.

The spatial dependence of LoR was also explored quantitatively in this study. With Moran's I test, it was confirmed that the LoR values were spatially correlated. Linear, spatial error, and spatial

lag models were used to estimate the LoR values with geographic, socioeconomic, and transportation features. The spatial error models outperformed the others by presenting smaller AIC and BIC values. Results indicate that the spatial autoregressive process occurs mainly in the error term. Omitted variables are the major cause of spatial correlation. Factors such as the percentage of area influenced by storm surges, the distance to the coast, and the average elevation are found to affect the infrastructure resilience with respect to hurricanes. It is likely that contributing factors to the infrastructure resilience when other disruptions, such as earthquakes and tornadoes, are confronted would be different.

As a result of the introduction of a smaller modeling unit for the zones and the study of spatial dependence, compared with previous studies, this paper is able to provide a deeper insight into the vulnerability of highway and transit networks in New York City (3–5). With socioeconomic and projected surge zone information, the spatial error and lag models for LoR can be used as an estimation tool of vulnerability assessment in response to future storms. These models can also be useful for government agencies and policy makers dealing with emergency management.

The results presented in this paper, however, may not be directly transferrable to other cities, considering the uniqueness and complexity of the transportation network in New York City. To predict the recovery performance of posthurricane recovery in other regions, this model needs to be recalibrated with empirical data or simulated data from regional multimodal network models.

The future improvement and calibration of this proposed method may consider other factors related to critical corridors, especially additional factors from highway and subway lines, since their recovery patterns may resemble each other in a common corridor. Another future research direction is to investigate the factors contributing to infrastructure resilience when other types of natural disasters are faced.

ACKNOWLEDGMENTS

This study was funded in part by CitySmart Laboratory at the Urban Intelligent Transportation Systems Center of New York University and the New York State Resiliency Institute for Storms and Emergencies.

REFERENCES

1. Metropolitan Transportation Authority. *Timeline of the Storm and Restoration of Service*. <http://web.mta.info/sandy/timeline.htm>. Accessed June 15, 2016.
2. City of New York. *NYC Hurricane Evacuation Zones*. <http://maps.nyc.gov/hurricane/>. Accessed May 20, 2016.
3. Zhu, Y., K. Ozbay, K. Xie, and H. Yang. Using Big Data to Study Resilience of Taxi and Subway Trips for Hurricanes Sandy and Irene. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2599, 2016, pp. 70–80. <http://dx.doi.org/10.3141/2599-09>.
4. Xie, K., K. Ozbay, Y. Zhu, and H. Yang. Case Studies for Data-Oriented Emergency Management/Planning in Complex Urban Systems. *LNCS Transactions on Large-Scale Data- and Knowledge-Centered Systems*, Vol. 9860, 2016, pp. 190–207.
5. Xie, K., K. Ozbay, Y. Zhu, and H. Yang. Data-Driven Method for Predicting Future Evacuation Zones in the Context of Climate Change. Presented at 95th Annual Meeting of the Transportation Research Board, Washington, D.C., 2016.
6. Heaslip, K., W. Louisell, J. Collura, and N. Urena Serulle. A Sketch Level Method for Assessing Transportation Network Resiliency to Natural Disasters and Man-Made Events. Presented at 89th Annual Meeting of the Transportation Research Board, Washington, D.C., 2010.

7. Bruneau, M., S.E. Chang, R.T. Eguchi, G.C. Lee, T.D. O'Rourke, A.M. Reinhorn, M. Shinozuka, K. Tierney, W.A. Wallace, and D. von Winterfeldt. A Framework to Quantitatively Assess and Enhance the Seismic Resilience of Communities. *Earthquake Spectra*, Vol. 19, No. 4, 2003, pp. 733–752. <http://dx.doi.org/10.1193/1.1623497>.
8. Testa, A. C., M. N. Furtado, and A. Alipour. Resilience of Coastal Transportation Networks Faced with Extreme Climatic Events. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2532, 2015, pp. 29–36. <http://dx.doi.org/10.3141/2532-04>.
9. Donovan, B., and D. Work. Using Coarse GPS Data to Quantify City-Scale Transportation System Resilience to Extreme Events. Presented at 94th Annual Meeting of the Transportation Research Board, Washington, D.C., 2015.
10. Hosseini, S., and K. Barker. Modeling Infrastructure Resilience Using Bayesian Networks: A Case Study of Inland Waterway Ports. *Computers & Industrial Engineering*, Vol. 93, 2016, pp. 252–266. <http://dx.doi.org/10.1016/j.cie.2016.01.007>.
11. Adjete-Bahun, K., B. Birregah, E. Châtelet, and J.-L. Planchet. A Model to Quantify the Resilience of Mass Railway Transportation Systems. *Reliability Engineering & System Safety*, Vol. 153, 2016, pp. 1–14. <http://dx.doi.org/10.1016/j.ress.2016.03.015>.
12. D'Lima, M., and F. Medda. A New Measure of Resilience: An Application to the London Underground. *Transportation Research Part A*, Vol. 81, 2015, pp. 35–46.
13. Linacre, L.J. *Why Logistic (Sigmoid) Ogive and Not Autocatalytic Curve?* <http://rasch.org/rmt/rmt64k.htm>. Accessed June 3, 2015.
14. Lewis, D.C. Transportation Planning for Hurricane Evacuations. *ITE Journal*, Vol. 55, No. 8, 1985, pp. 31–35.
15. Hobeika, A.G., A.E. Radwan, and D. Sivasailam. *A Computer Simulation Model for Rural Network Evacuation Under Natural Disasters*. National Emergency Training Center, Emmitsburg, Md. 1985.
16. Fu, H., C. G. Wilmot, H. Zhang, and E. J. Baker. Modeling the Hurricane Evacuation Response Curve. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2022, 2007, pp. 94–102. <http://dx.doi.org/10.3141/2022-11>.
17. Li, J., K. Ozbay, B. Martin, S. Iyer, and J.A. Carnegie. Empirical Evacuation Response Curve During Hurricane Irene in Cape May County, New Jersey. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2376, 2013, pp. 1–10. <http://dx.doi.org/10.3141/2376-01>.
18. Ozbay, K., and M.A. Yazici. Study of Networkwide Impact of Various Demand Generation Methods Under Hurricane Evacuation Conditions. Presented at 85th Annual Meeting of the Transportation Research Board, Washington, D.C., 2006.
19. Tasic, I., and R.J. Porter. Modeling Spatial Relationships Between Multimodal Transportation Infrastructure and Traffic Safety Outcomes in Urban Environments. *Safety Science*, Vol. 82, 2016, pp. 325–337. <http://dx.doi.org/10.1016/j.ssci.2015.09.021>.
20. Xie, K., K. Ozbay, and H. Yang. Spatial Analysis of Highway Incident Durations in the Context of Hurricane Sandy. *Accident Analysis and Prevention*, Vol. 74, 2015, pp. 77–86. <http://dx.doi.org/10.1016/j.aap.2014.10.015>.
21. Work, D., and B. Donovan. *2010–2013 New York City Taxi Data*. <http://publish.illinois.edu/dbwork/open-data/>. Accessed May 1, 2015.
22. Metropolitan Transportation Authority. *MTA Turnstile Data*. <http://web.mta.info/developers/turnstile.html>. Accessed June 1, 2015.
23. Federal Emergency Management Agency. *Surge Data of Hurricane Irene and Sandy*. ArcGIS REST Services Directory. 2013. http://services.fema.data.com/arcgis/rest/services/2012_Sandy. Accessed July 15, 2016.
24. Columbia University. Columbia University Spatial Data Catalog. NYC Digital Elevation Models. <http://culspatial.cul.columbia.edu/>. Accessed July 1, 2016.
25. NYC Department of City Planning. *New York City Neighborhood Tabulation Areas Metadata*. 2013. http://www.nyc.gov/html/dcp/pdf/bytes/nynta_metadata.pdf. Accessed June 15, 2015.
26. Moran, P.A. The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society. Series B. Methodological*, Vol. 10, No. 2, 1948, pp. 243–251.
27. Anselin, L. *Spatial Econometrics: Methods and Models*. Springer Netherlands. 1988. <http://dx.doi.org/10.1007/978-94-015-7799-1>.
28. Goodchild, M.F. *Spatial Autocorrelation*. Geo Books, Norwich, United Kingdom, 1986.
29. Getis, A., and J.K. Ord. The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, Vol. 24, No. 3, 1992, pp. 189–206. <http://dx.doi.org/10.1111/j.1538-4632.1992.tb00261.x>.
30. Xie, K., X. Wang, K. Ozbay, and H. Yang. Crash Frequency Modeling for Signalized Intersections in a High-Density Urban Road Network. *Analytic Methods in Accident Research*, Vol. 2, 2014, pp. 39–51. <http://dx.doi.org/10.1016/j.amar.2014.06.001>.
31. Mitchell, A. *The ESRI Guide to GIS Analysis, Volume 2: Spatial Measurements and Statistics*. ESRI Press, 2005.
32. Anselin, L., I. Syabri, and Y. Kho. GeoDa: An Introduction to Spatial Data Analysis. *Geographical Analysis*, Vol. 38, No. 1, 2006, pp. 5–22. <http://dx.doi.org/10.1111/j.0016-7363.2005.00671.x>.
33. Draper, N.R., and H. Smith. *Applied Regression Analysis*, 2nd ed., John Wiley & Sons, 1981.
34. Akaike, H. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, Vol. 19, No. 6, 1974, pp. 716–723.
35. Schwarz, G. Estimating the Dimension of a Model. *Annals of Statistics*, Vol. 6, No. 2, 1978, pp. 461–464. <http://dx.doi.org/10.1214/aos/1176344136>.
36. Burnham, K.P., and D.R. Anderson. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, 2002.
37. Kass, R.E., and A.E. Raftery. Bayes Factors. *Journal of the American Statistical Association*, Vol. 90, No. 430, 1995, pp. 773–795. <http://dx.doi.org/10.1080/01621459.1995.10476572>.
38. Tavassoli Hojati, A., L. Ferreira, S. Washington, and P. Charles. Hazard Based Models for Freeway Traffic Incident Duration. *Accident Analysis and Prevention*, Vol. 52, 2013, pp. 171–181. <http://dx.doi.org/10.1016/j.aap.2012.12.037>.

The contents of this paper reflect the views of the authors, who are responsible for the facts and accuracy of the data. The contents do not necessarily reflect the official views or policies of the agencies.

The Standing Committee on Critical Transportation Infrastructure Protection peer-reviewed this paper.