Estimating Urban Traffic Safety and Analyzing Spatial Patterns through the Integration of City-wide Near-miss Data: A New York City Case Study

Chuan Xu[ab], Jingqin Gao[b], Fan Zuo[b], Kaan Ozbay[b]
[a] Southwest Jiaotong University [b] New York University

C2SMARTER CONNECTED COMMUNITIES WITH SMART MOBILITY TO EQUITABLY REDUCE CONGESTION | NYU | 西南交通大学 Southwest Jiaotong University

Paper 24-04985

Contact: xuchuan7@gmail.com，jg3146@nyu.edu

Abstract

Backgrounds: City-wide near-miss data can be beneficial for traffic safety estimation. In this study, we evaluate urban traffic safety and examine spatial patterns by incorporating city-wide near-miss data (59,277 near misses).

Methods: Our methodology employs a grid-based method, the Empirical Bayes (EB) approach, and spatial analysis tools including global Moran's I and local Moran's I .

Results:

- The study findings reveal that near misses have the strongest correlation with observed crash frequency among all the variables studied. Interestingly, the ratio of near-misses to crashes is roughly estimated to be 1957:1, providing a potentially useful benchmark for urban areas. For other variables, an increased number of intersections and bus stops, along with a greater road length, contribute to a higher crash frequency. Conversely, residential and open-space land use rates show a negative correlation with crash frequency. Through spatial analysis, potential risk hotspots including roads linking bridges and tunnels, and avenues bustling with pedestrian activity, are highlighted.
- The study also identified negative local spatial correlations in crash frequencies, suggesting significant safety risk variations within relatively short distances. By mapping the differences between observed and predicted crash frequencies, we identified specific grid areas with unexpectedly high or low crash frequencies.

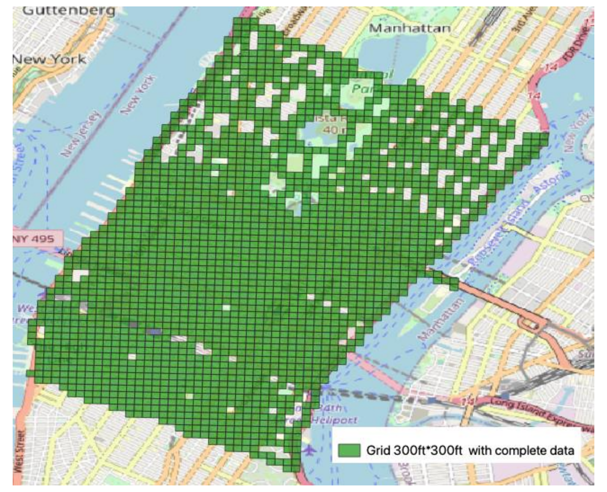
Implications:

These findings highlight the crucial role of near-miss data in urban traffic safety policy and planning, particularly relevant with the imminent rise of autonomous and connected vehicles. By integrating near-miss data into safety estimations, we can develop a more comprehensive understanding of traffic safety and, thus, more effectively address urban traffic risks.

METHODOLOGY

Grid-based method

The grid cell-based method divides a geographical area into a grid of uniformly sized and shaped cells. This cell size was chosen because it closely aligns with the standard block width in Manhattan (264 ft) and the block length (900 ft) is divisible by 300 ft .



Empirical Bayes method

The Empirical Bayes (EB) method was introduced to estimate crash frequency. It combines the observed crash frequency from real-world data and the expected crash frequency predicted by Safety Performance Function (SPF).

$$N_{eb} = \omega \times N_{spf} + (1-\omega) \times N_{ob}$$

$$\omega = \frac{1}{1 + N_{spf}/\varphi}$$

where N_{eb} is the estimated crash frequency in each grid using the EB method, N_{spf} is the estimated crash frequency in each grid by SPF, N_{ob} is the observed crash frequency in each grid, ω is the weight, and φ is the dispersion parameter of the SPF model

Spatial Analysis

Global Moran's I and local Moran's I were utilized to analyze the spatial autocorrelation of the estimated crash frequency. Global Moran's I test is widely used to measure how related the values of a variable are based on the locations and the values of their neighbors. The Local Indicators of Spatial Association (LISA) test assumes that global Moran's I is a summation of individual cross-products

$$Z_I = (I - E[I])/SD[I]$$

$$I_i = \frac{z_i}{\left(\frac{\sum_i z_i^2}{N}\right)} \sum_j w_{ij} z_j$$

- $E[I]$ is the expectation of I and $SD[I]$ is the standard deviation of I. A positive Z_I indicates the observation distribution is spatially clustered
- Local Moran's I for the observation z_i, z_j in cell i, j with weight matrix w_{ij} and N observations

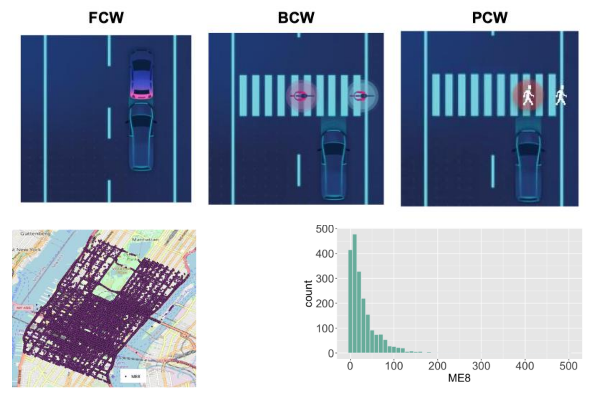
DATA PREPARATION

Near misses

The near-miss data is extracted from Mobileye collision warning events that were reported by vehicles equipped with Mobileye Advanced driver assistance systems (ADAS) solution.

Forward Collision Warning (FCW)
Pedestrian Collision Warning (PCW)
Bicyclist Collision Warning (BCW)

FCW indicates a potential vehicle-to-vehicle collision, detected up to 80 meters ahead and active for speeds between 1 km/h and 200 km/h. The TTC threshold for FCW is triggered at 2.7 seconds. Both BCW and PCW involve potential collisions with bicyclists and pedestrians, respectively, detected up to 28 meters ahead and active for speeds between 1 km/h and 50 km/h. The TTC threshold for these warnings is 2 seconds.



DATA PREPARATION

Crash data

The historical motor vehicle crash data for this analysis was obtained from Open NYC

To maintain consistency with the near-miss data, the crash data was filtered to include incidents occurring between July 5, 2022, and December 31, 2022.

Other data

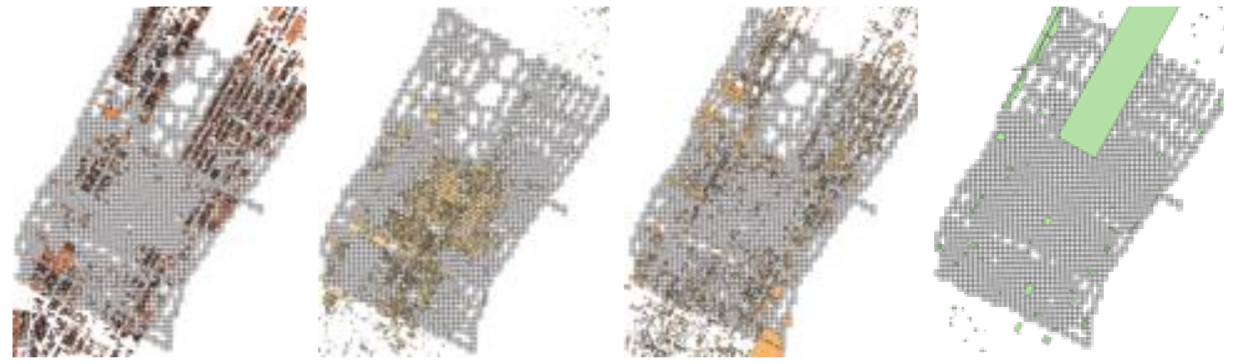
Traffic exposure data
Road network & transport facility
Land use
Population density

Annual Average Daily Traffic (AADT) and Vehicle Miles Traveled (VMT) data are used as approximations for traffic exposure.

The road network data for New York City was obtained from Data.gov.

Land use data were acquired from the NYC Department of City Planning (NYC DCP) Map PLUTO

Population data was obtained from the U.S. Census Bureau.



(a) Residential area (b) Commercial area (c) Mixed R & C (d) Open area

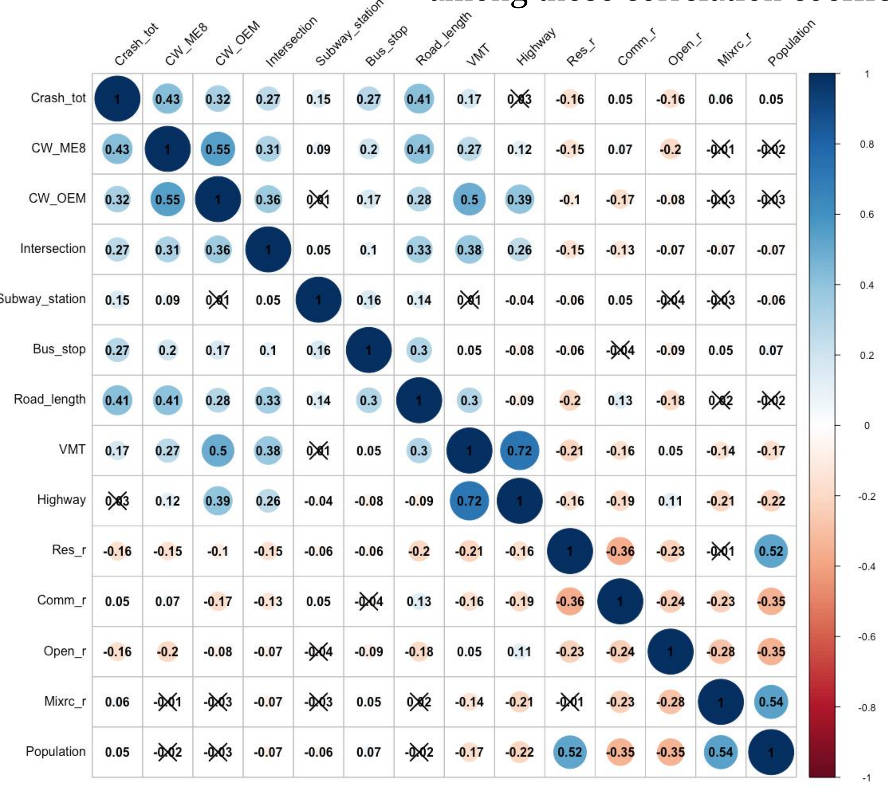
Input Variable Summary

Variable	Mean	S.D.	Median	Min	Max
Crash_tot	1.43	2.02	1	0	15
CW_ME8	30.87	38.30	19	0	504
CW_OEM	1.24	2.01	0	0	17
Intersection	0.93	1.40	1	0	18
Subway_station	0.026	0.17	0	0	2
Bus_stop	0.34	0.65	0	0	4
Road_length (ft)	407.97	176.82	355.22	3.64	1032.05
VMT (mi*veh)	1259	1869	723	0	15777
Highway	0.08	0.27	0	0	1
Res_r	15%	20%	5%	0%	98%
Comm_r	16%	23%	2%	0%	87%
Open_r	9%	26%	0%	0%	100%
Mix_rc_r	14%	16%	8%	0%	91%
Population	259	226	227	0	1367

RESULTS AND DISCUSSION

Correlation Analysis

- The correlation coefficient of crash count and ME8 near misses (CW_ME8) is 0.43, which is the highest among those correlation coefficients for crash count.



- This indicates that the crash count (both total crashes and injury and fatal crashes) and near misses have a moderate positive linear relationship and near misses may play an important part in the crash prediction models

RESULTS AND DISCUSSION

Crash Prediction Model

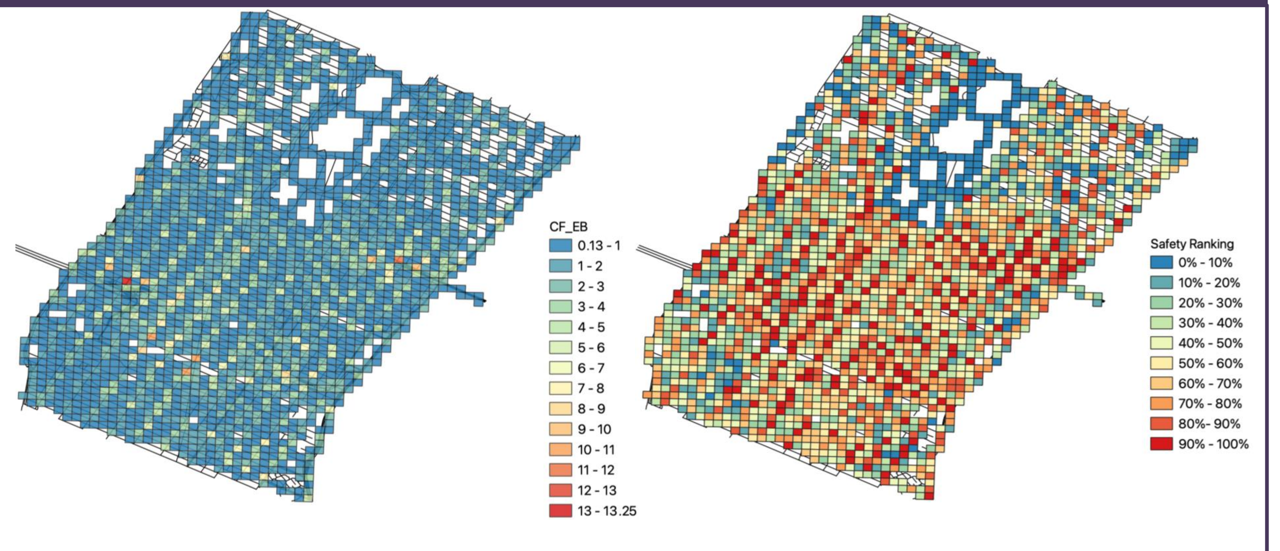
We utilized the preliminarily selected variables to construct an NB model. The StepAIC function in R was employed to identify the optimal variables for input.

	Estimate	Marginal Effects	Std. Error	z value	Pr(>|z|)
Intercept	-0.88		0.09	-9.389	0.00**a
CW_ME8	0.01	0.0074	0.00	10.694	0.00**
Intersection	0.12	0.1322	0.02	6.893	0.00**
Subway_station	0.20	0.2225	0.14	1.506	0.132
Bus_stop	0.19	0.2062	0.04	4.878	0.00**
Road_length	0.01	0.0065	0.00	10.152	0.00**
Res_r	-0.73	-0.8001	0.17	-4.419	0.00**
Open_r	-1.04	-1.1295	0.16	-6.358	0.00**
Mix_rc_r	0.28	0.3002	0.17	1.587	0.112
Ψb	1.53				
Std. err	0.12				
AIC	6045.2				
2*log-likelihood	-6025.2				

- ❖ Near-miss-to-crash ratio estimation
  - The marginal effect of near misses, as represented by CW_ME8, is 0.0074. This implies that in our study area, the near-miss to crash rate is approximately 135:1.
  - To further refine the total near-miss to crash rate, we leverage two statistics: that 10% of cars have the Advanced Driver Assistance System (ADAS) installed (*48*), and Mobileye holds a 69% market share in ADAS (*49*).
  - From this, we estimate that 6.9% of vehicles in New York City are equipped with Mobileye 8 Connected devices. This, in turn, allows us to calculate the total near-miss-to-crash ratio in NYC as roughly 1957:1.
- ❖ An increase in the number of intersections results in a rise in estimated crash frequency by 0.13.
- ❖ An increase of one bus stop per grid enhances crash frequency by approximately 0.21
- ❖ An increase in road length by 100 ft augments the estimated crash count by 0.65

Spatial Distribution

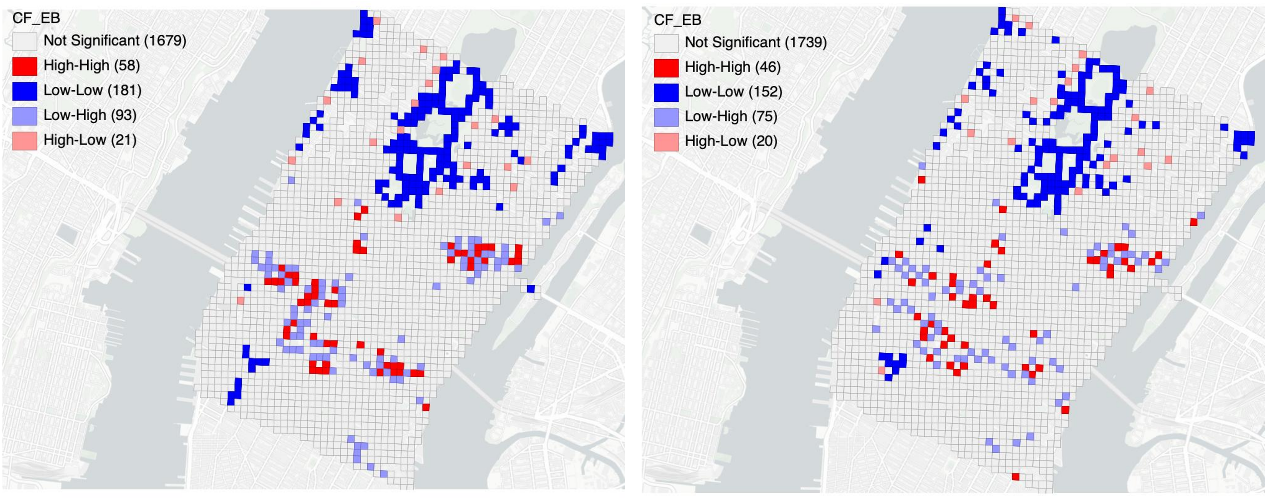
- First, grids encompassing linkage roads of bridges and tunnels exhibit a higher CF-EB. This is likely due to the complexity of the road network at these locations, with more frequent merging and diverging points leading to more frequent lane changes by drivers.
- Second, high-risk grids predominantly line the avenues from 5th Ave to 8th Ave. These avenues are characterized by high pedestrian activity due to the concentration of commercial establishments, public transit access points, and other urban amenities.



(a) the EB estimated crash frequency map (b) the safety ranking map

Spatial Autocorrelation

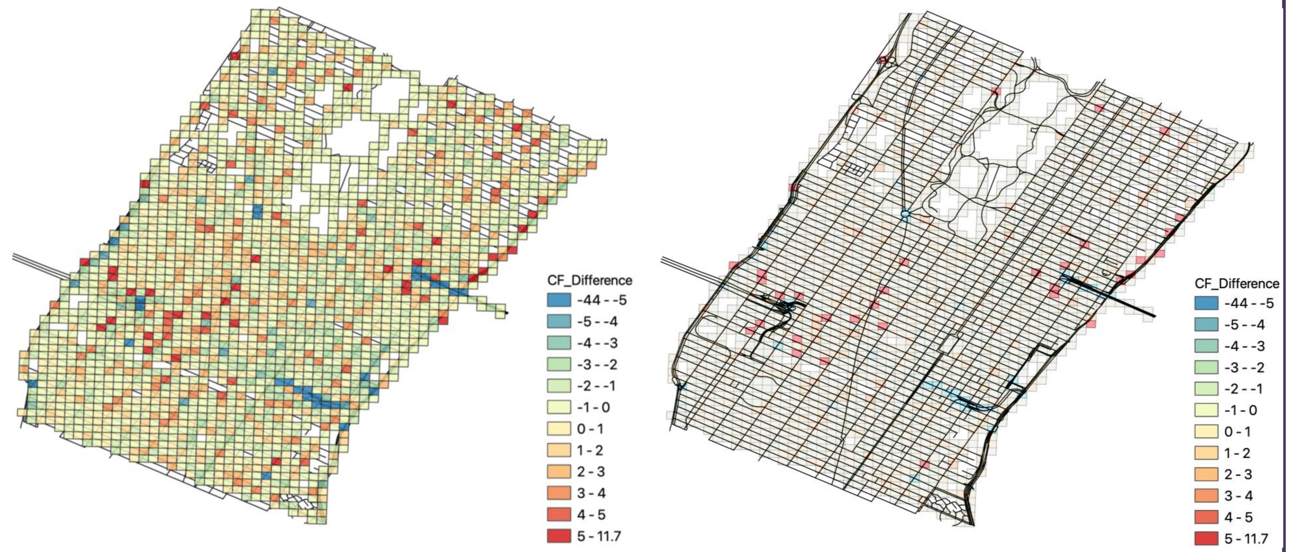
- Moran's I test: A statistically significant spatial correlation
- LISA Map: High-high and low-low locations have positive local spatial correlation.
- The maps also display fewer high-low locations but a significant number of low-high locations. High-low clusters primarily surround low-low clusters, such as the Central Park area. Low-high clusters are predominantly situated near some high-high locations, such as the Queensboro Bridge linkage road area and Midtown East, between 34th Street and 42nd Street.



(a) K-nearest neighbors (k=8) (b) K-nearest neighbors (k=4)

Observation and Prediction Difference

- Dark red grids represent areas with a higher-than-expected crash frequency. These grids are predominantly located near two tunnels, in the vicinity of Penn Station, and along the highway. For the grids situated on the highway, the road network is characterized by risk factors such as curves and interchanges.
- On the other hand, the dark blue grids, which depict locations with fewer accidents than predicted, are situated on or near the linkage roads of bridges and tunnels (Lincoln tunnel, Queens midtown tunnel, and Queensboro bridge), or covering a roundabout (Columbus Circle). The road networks within these grids are often complex and with large road lengths.



(a) grid view (b) road network view

CONCLUSIONS

- Near misses have the highest correlation with crash frequency among all the variables
- Other variables such as the number of intersections, number of bus stops, road length, residential land use rate, and open space land use rate significantly influence crash frequency in New York City.
- The near-miss-to-crash ratio, estimated to be 1957:1 for the study urban area, can serve as a potential benchmark for other urban environments.
- The analysis of network and facility-related variables revealed that a higher number of intersections and bus stops, and longer road length, all contribute to an increased crash frequency.
- Residential and open-space land use rates were negatively correlated with crash frequency, likely due to lower traffic volumes, fewer complex intersections, and more traffic calming measures in these areas.
- Spatial analysis highlighted areas with high and low crash frequencies, revealing potential risk hotspots, such as linkage roads of bridges and tunnels, and avenues with high pedestrian activity.
- The observation and prediction difference map further identified grids with unexpectedly high or low crash frequencies, offering valuable insights for targeted interventions.