Modeling Multimodal Crash Frequency

Yongping Zhang, Wen Cheng
Department of Civil Engineering
California State Polytechnic University, Pomona

CAL POLY POMONA

Tom Vo
Southern California Association of Governments

SCAG
INNOVATING FOR A BETTER TOMORROW

ESRI UC Conference
July 11, 2017
San Diego, CA

Outline

1. Introduction
2. Methodology
3. Data
4. Results
5. Conclusions
6. Future Work

Introduction

- Encouraging individuals to use active transportation, including walking and bicycling, brings with it a societal obligation to protect commuters as they engage in these modes of travel.

- Despite the health and environmental benefits, choosing walking or cycling as the desired mode exposes cyclists and pedestrians to safety risks.

Introduction

- Urban mobility and safety for all modes of transportation are key elements in the development of safer traffic environment.

- Reaching this goal requires the implementation of multimodal approaches and a shift towards non-motorized modes of transportation, namely walking and cycling.

Introduction

- The traffic safety field has employed separate temporal and spatial correlations for simultaneous estimation of crash outcomes.
- However, this is no or little research considering both dimensions of time and space, as well as the associated interactions, for the multivariate models.
- To this end, this study proposed two multivariate spatial-temporal models.

Methodology

- This study analyzed four different transportation mode users-involved crashes occurring at the TAZ's of the City of Irvine in California.
- The process involved development of multivariate spatial-temporal models and compare their modeling and site ranking performance with three other competing multivariate models assuming a Poisson-Lognormal distribution for crash counts.
- All models in the study were developed using the Full Bayes approach.

Methodology

- Model 1: Multivariate Poisson-Lognormal Model (MVPLN)
- Model 2: Multivariate Poisson-Lognormal with Time Trend (MVPLNT)
- Model 3: Multivariate Poisson-Lognormal Spatial Model (MVPLNS)
- Model 4: Multivariate Poisson-Lognormal Spatial-Temporal Model (MVPLNST) with Fixed Time Coefficient
- Model 5: Multivariate Poisson-Lognormal Spatial-Temporal Model (MVPLNST) with Varying Time Coefficients for Crash Types

Methodology

Goodness-of-Fit of the Models

- The Deviance Information Criterion (DIC) was employed to assess the complexity and fit of the models.

- The DIC is computed as the sum of the posterior mean deviance and estimated effective number of parameters:

$$DIC = \overline{D} + p_D$$

Data

1. Crashes which occurred in the City of Irvine (2006–2012)
2. Compared with other geographic units such as block groups and census tracts, TAZs have benefits of better homogeneity and easy integration into the transportation planning process
3. There are 203 TAZs in the City
4. Four different transportation mode-related crashes were collected from SWITRS (California Statewide Integrated Traffic Records System) which include pedestrian, bicyclist, motorcyclist and vehicle only crashes
5. Shape file of TAZ boundary and TAZ characteristics were provided by SCAG

Data

1. The numbers of various transportation mode-involved crashes were used as the dependent variables.
2. DVMT was utilized as the exposure variable.
3. The explanatory variables were the predictors commonly used in previous regional safety analyses which include socioeconomic, transportation-related, and environment-related factors, and so on.
4. In addition, the distance matrix containing distances among various TAZ centroids were also collected from SCAG for the estimation of distance-based spatial random effect.

Data

The variables used for model development and the associated descriptive statistics

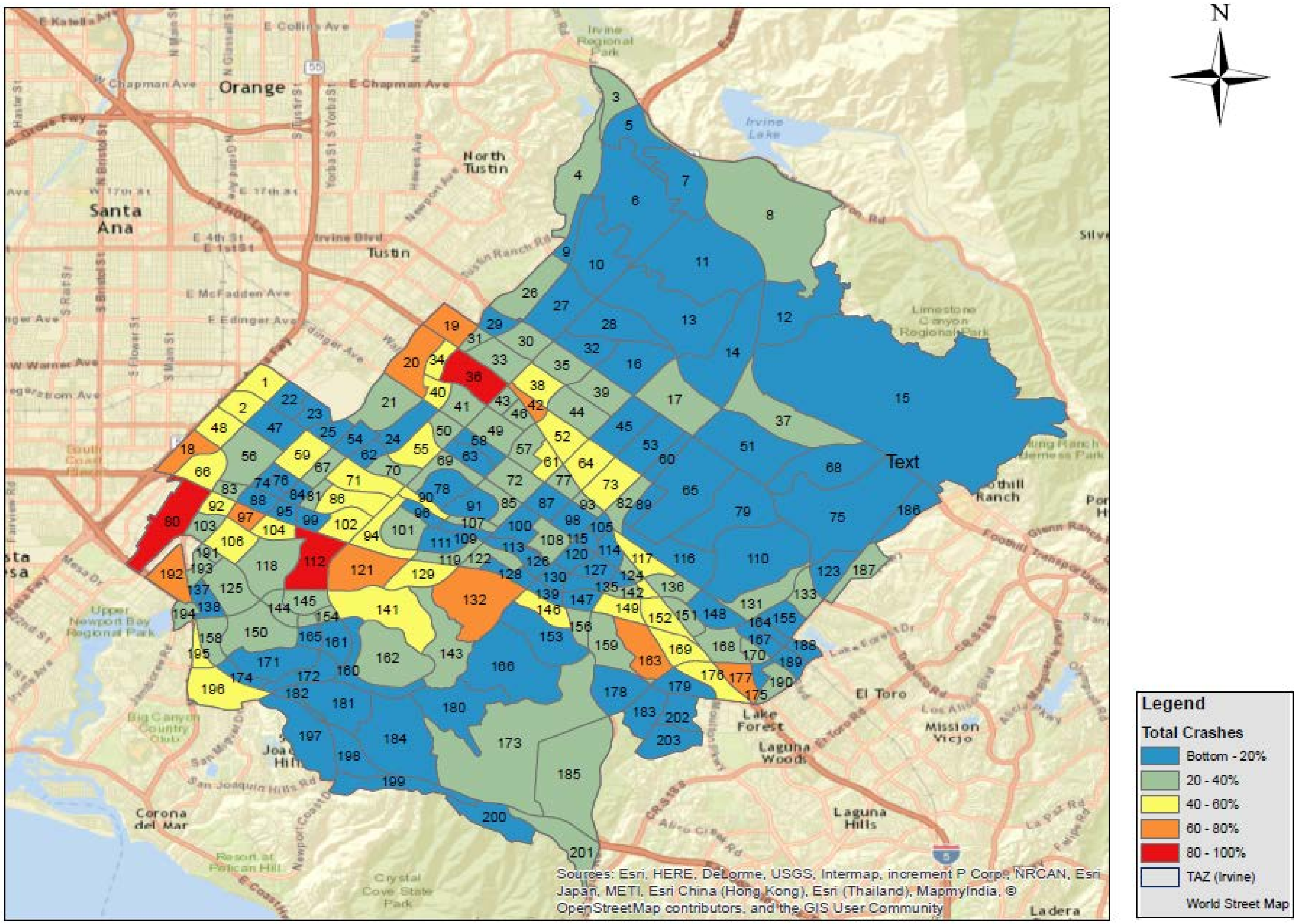
Variables	Description	Mean	Std. Dev.	Min.	Max.
Veh	Total vehicle-only crashes	44.05	56.19	0	394
MC	Total motorcycle-involved crashes	2.67	4.25	0	24
Bike	Total bike-involved crashes	2.59	3.24	0	14
Ped	Total pedestrian-involved crashes	1.35	2.09	0	15
DVMT	Daily vehicle miles traveled	54,262.44	56,156.84	112.6	276,079.90
Acre	TAZ Area in acre	282.9	431.75	0.69	5,062.95
Median	Median house income ($)	48,440.78	50,635.10	0	183,347
Pop_den	Population density by area	6.18	7.96	0	32.4
HH_den	Household density by area	2.34	3.15	0	13.62
Emp_den	Employment density by area	10.34	17.43	0	121.1
Ret_den	Retail job density	0.79	2.02	0	17.45
% age 5_17	% of population age 5-17	8.64%	8.78%	0	27%
% age 18_24	% of population age 18-24	5.79%	7.42%	0	40%
% age 24_64	% of population age 24-64	38.35%	36.12%	0	95%
% age 65+	% of population age 65 or older	6.25%	10.21%	0	83%

Data

The variables used for model development and the associated descriptive statistics

Variables	Description	Mean	Std. Dev.	Min.	Max.
K12	K12 student enrollment	0.39	1	0	5.52
College	College student enrollment	0.11	1	0	12.59
Int34_den	Intersection density (3- and 4- legs)	0.12	0.12	0	0.62
BKlnACC	Bike lane access (1=if a TAZ has bike lane)	0.92	0.28	0	1
BL_den	Bike lane density	3.4	1.8	0	7.26
Rail	1=at least one rail station in a TAZ	0.01	0.1	0	1
TTbus_D	Total Bus Stop Density	0.05	0.09	0	0.53
Exbus_D	Stop density for Express Bus and BRT	0.002	0.007	0	0.06
HFLbus_D	High-Frequency Bus Stop Density (local bus headway <= 20 mins)	0.001	0.004	0	0.03
WalkAcc	Walk Accessibility	3.87	9.46	0	74.53
% Arterial	Percent of main arterial (45-55mph) of TAZ	10.61%	17.33%	0	80%
Distance	Distance among TAZ centroids (in miles)	4.06	2.09	0.16	11.78

Data: the distribution of the all TAZs and associated crash counts



Results: Model Comparison

Crash Types	Variables	Model 1: MVPLN	Model 2: MVPLNT	Model 3: MVPLNS	Model 4: MVPLNST (fixed time coefficient)	Model 5: MVPLNST (varying time coefficients)
Vehicle	Constant	**-10.29 (0.07)**	**-10.12 (0.11)**	**-10.27 (0.08)**	**-10.19 (0.09)**	**-10.13(0.13)**
	% age 18_24	**2.32 (0.46)**	**2.35 (0.47)**	**2.16 (0.46)**	**2.35 (0.48)**	**2.41 (0.48)**
	% age 65+	**1.24 (0.29)**	**1.23 (0.31)**	**1.24 (0.29)**	**1.25 (0.30)**	**1.25 (0.30)**
	College	**-0.18 (0.03)**	**-0.18 (0.03)**	**-0.18 (0.03)**	**-0.18 (0.03)**	**-0.18 (0.03)**
	% Arterial	**-0.44 (0.19)**	**-0.44 (0.20)**	**-0.47 (0.19)**	**-0.43 (0.20)**	**-0.46 (0.20)**
	BL_den	**0.17 (0.02)**	**0.17 (0.02)**	**0.17 (0.02)**	**0.17 (0.02)**	**0.17 (0.02)**
	Yearly trend	NA	**-0.04 (0.02)**	NA	**-0.03 (0.01)**	**-0.04 (0.02)**
Motorcycle	Constant	**-12.81(0.19)**	**-12.89(0.22)**	**-12.74 (0.18)**	**-12.71(0.20)**	**-12.93(0.25)**
	% age 18_24	1.10 (1.04)	0.99 (1.03)	0.85 (0.99)	1.08 (1.06)	1.12 (1.04)
	% age 65+	**1.86 (0.60)**	**1.82 (0.60)**	**1.78 (0.59)**	**1.85 (0.60)**	**1.85 (0.60)**
	College	-0.09 (0.07)	-0.09 (0.07)	-0.09 (0.07)	-0.09 (0.07)	-0.09 (0.07)
	% Arterial	**-1.11 (0.42)**	**-1.10 (0.43)**	**-1.17 (0.41)**	**-1.10 (0.43)**	**-1.11 (0.42)**
	BL_den	0.07 (0.05)	0.07 (0.04)	0.07 (0.04)	0.07 (0.05)	0.08 (0.04)
	Yearly trend	NA	0.03 (0.03)	NA	**-0.03 (0.01)**	0.03 (0.03)

Results: Model Comparison

Crash Types	Variables	Model 1: MVPLN	Model 2: MVPLNT	Model 3: MVPLNS	Model 4: MVPLNST (fixed time coefficient)	Model 5: MVPLNST (varying time coefficients)
Bike	% age 18_24	**3.96 (0.93)**	**4.02 (0.95)**	**3.64 (0.88)**	**4.07 (0.97)**	**3.96 (0.91)**
	% age 65+	**1.98 (0.53)**	**2.04 (0.55)**	**1.98 (0.52)**	**2.05 (0.54)**	**1.99 (0.52)**
	College	**-0.12 (0.05)**	**-0.12 (0.05)**	**-0.11 (0.05)**	**-0.12 (0.05)**	**-0.12 (0.05)**
	% Arterial	**-6.39 (0.79)**	**-6.32 (0.79)**	**-6.19 (0.75)**	**-6.28 (0.80)**	**-6.40(0.79)**
	BL_den	**0.25 (0.05)**	**0.25 (0.05)**	**0.24 (0.04)**	**0.25 (0.05)**	**0.25 (0.04)**
	Yearly trend	NA	**0.13 (0.03)**	NA	**-0.03 (0.01)**	**0.13 (0.03)**
Pedestrian	Constant	**-14.63 (0.29)**	**-14.80 (0.33)**	**-14.60 (0.29)**	**-14.58 (0.30)**	**-14.79 (0.35)**
	% age 18_24	**5.24 (1.05)**	**5.35 (1.08)**	**4.72 (1.04)**	**5.32 (1.07)**	**5.32 (1.08)**
	% age 65+	1.38 (0.74)	1.46 (0.76)	1.41 (0.73)	1.46 (0.76)	1.45 (0.76)
	College	**-0.96 (0.53)**	**-0.96 (0.53)**	**-0.92 (0.50)**	**-0.96 (0.53)**	**-0.97 (0.53)**
	% Arterial	**-2.63 (0.74)**	**-2.60 (0.74)**	**-2.70 (0.72)**	**-2.60 (0.75)**	**-2.62 (0.74)**
	BL_den	**0.39 (0.06)**	**0.39 (0.06)**	**0.38 (0.06)**	**0.39 (0.06)**	**0.39 (0.06)**
	Yearly trend	NA	0.02 (0.04)	NA	**-0.03 (0.01)**	0.02 (0.04)
Goodness-of-fit	$\bar{D}$	9009.04	8963.96	9009.44	8971.7	8989.45
	p_D	1151.25	1148.3	1077.07	1083.76	1048.25
	DIC	10160.6	10112.3	10086.5	10055.5	10037.7

Results: Correlation among Crash Types

	Heterogeneity (ε_{ij})				Independent Spatial (u_{ij})				Interacted Spatial (δ_{ij})			
	Veh	MC	Bike	Ped	Veh	MC	Bike	Ped	Veh	MC	Bike	Ped
Veh	1.00	0.12	0.00	0.25	1.00	0.48	0.04	0.02	1.00	0.95	0.02	0.03
MC	0.04	1.00	0.00	0.00	-0.06	1.00	0.00	0.26	0.00	1.00	0.00	0.00
Bike	**0.71**	**0.75**	1.00	0.00	**0.16**	**0.26**	1.00	0.00	**0.18**	**0.42**	1.00	0.00
Ped	-0.03	**0.36**	**0.57**	1.00	**0.18**	0.05	**0.23**	1.00	**0.17**	**0.13**	**0.24**	1.00

Results:

Performance Comparison in Site Ranking

Methods	Sensitivity (top5%)	Specificity (top 5%)	Kappa (3 groups)	Kappa (4 groups)	Kappa (5 groups)	MAD
Method 1	90%	99.5%	0.978	0.963	0.948	0.498
Method 2	100%	100%	0.985	0.965	0.960	0.482
Method 3	90%	99.5%	0.985	0.967	0.957	0.479
Method 4	100%	100%	0.985	0.967	0.962	0.476

Conclusions

- The model fitness results from DIC revealed that the proposed models significantly improved the model fitting by pooling strength from the neighbors, consideration of time trend, as well as their interactions.
- Among the two proposed models, the model with mode-varying time coefficients was observed to be better. The influential factors for all the models were the same.
- The relative site ranking performance using different evaluation criteria with increasing group divisions consistently showed the strong positive correlation between modeling and site ranking performances.
- In other words, the models with closer DIC's tend to yield more similar ranking results.

Future Work

- This study was focused at the TAZ level with a set of influential variables. Somewhat different results may be expected for other geographic areas, like block level, county, or smaller entities like intersections.
- The fitness of models was assessed by employing the DIC. Other techniques could be utilized for such assessment like MAD, MSPE, among others.
- Cross validation techniques would also help verify the expected advantages at crash prediction.
- This study used a linear time-space interaction for development of models. The fitness and performance of other time-space relationships could be explored and compared with the proposed models.

Thank you!

Questions?