JOURNAL OF TRAFFIC AND TRANSPORTATION ENGINEERING (ENGLISH EDITION) 2016; 3 (3): 262–270

Available online at www.sciencedirect.com

ScienceDirect

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

Original Research Paper

Development and application of traffic accident density estimation models using kernel density estimation

Seiji Hashimoto [a], Syuji Yoshiki [b,*], Ryoko Saeki [c], Yasuhiro Mimura [d], Ryosuke Ando [e], Shutaro Nanba [a]

[a] Graduate School of Environmental and Life Science, Okayama University, Okayama 700-0807, Japan
[b] Department of Civil Engineering, Fukuoka University, Fukuoka 814-0180, Japan
[c] Fukuyama Consultants Co. Ltd., Fukuoka 812-0013, Japan
[d] International Development Consultants Co. Ltd., Aichi 460-0008, Japan
[e] Toyota Transportation Research Institute, Aichi 471-0024, Japan

ARTICLE INFO

Article history:
Received 14 May 2015
Received in revised form
30 December 2015
Accepted 6 January 2016
Available online 10 May 2016

Keywords:
Traffic safety
Kernel density estimation (KDE)
Hotspots
Zone 30

ABSTRACT

Traffic accident frequency has been decreasing in Japan in recent years. Nevertheless, many accidents still occur on residential roads. Area-wide traffic calming measures including Zone 30, which discourages traffic by setting a speed limit of 30 km/h in residential areas, have been implemented. However, no objective implementation method has been established. Development of a model for traffic accident density estimation explained by GIS data can enable the determination of dangerous areas objectively and easily, indicating where area-wide traffic calming can be implemented preferentially. This study examined the relations between traffic accidents and city characteristics, such as population, road factors, and spatial factors. A model was developed to estimate traffic accident density. Kernel density estimation (KDE) techniques were used to assess the relations efficiently. Besides, 16 models were developed by combining accident locations, accident types, and data types. By using them, the applicability of traffic accident density estimation models was examined. Results obtained using Spearman rank correlation show high coefficients between the predicted number and the actual number. The model can indicate the relative accident risk in cities. Results of this study can be used for objective determination of areas where area-wide traffic calming can be implemented preferentially, even if sufficient traffic accident data are not available.

* *Corresponding author*. Tel.: +81 92 8716631x6494; fax: +81 92 8656031.
E-mail address: syoshiki@fukuoka-u.ac.jp (S. Yoshiki).

Peer review under responsibility of Periodical Offices of Chang'an University.
http://dx.doi.org/10.1016/j.jtte.2016.01.005

1. Introduction

Traffic accident frequency has been decreasing in Japan in recent years. Nevertheless, many accidents still occur on residential roads. Therefore, it is necessary to implement measures for traffic calming on roads. Area-wide traffic calming measures, including Zone 30 (Institute for Road Safety Research, 2004), are especially effective and have often been implemented throughout urban areas. However, for implementation of these measures in Japan, it is impossible to adopt a universal application methodology that merely implements these measures in an entire urban area because borders between urban and rural areas are ill-defined in Japan. In contrast to foreign countries, urban areas in Japan spread out throughout a city, because Japan is an island nation. It has not been exposed to the threat of different regional ethnic groups and it is also not necessary to surround urban areas with walls. This urban feature makes it easy to construct sprawling cities with ever-increasing populations. Consequently, urban areas have been constructed in Japan with ill-defined borders separating urban and rural areas. Moreover, it is difficult to implement measures quickly because of budget constraints. Therefore, it is necessary to ascertain areas where implementation is the most preferred.

Locations of traffic accidents are crucially important information for the implementation of traffic safety measures. Location information of arterial roads is readily available because traffic is typically heavy and traffic accidents often occur on such roads. However, insufficient information is available about such locations on residential roads because traffic accidents occur rarely and incidentally on such roads. Moreover, traffic accident data are difficult to obtain from police departments of many Japanese cities. It is difficult to ascertain the distribution of traffic accidents in all cities including residential areas. Consequently, when area-wide traffic calming measures are implemented in a city, determining which areas these measures should be implemented preferentially must depend on an experience-based subjective view.

The authors develop an estimative model of traffic accident density from GIS data, which are commonly available data by the public and private sector, including population, road factors, and spatial factors. The model enables objective and easy determination of areas to implement area-wide traffic calming preferentially, even if traffic accident data are not available.

This study examined relations between traffic accidents and city components of population, road factors, and spatial factors. Then a model was developed to estimate traffic accident density.

A sufficient amount of traffic accident data must be accumulated to develop the model because traffic accidents in residential areas occur rarely and incidentally. The possibility exists that analysis based on only a few years of data impairs the predictive accuracy. Therefore, this study specifically uses kernel density estimation (KDE) described by Silverman (1986), which can deal with comprehensive estimation of the distribution based on a finite data sample.

KDE has been used for traffic accident analysis and widely as a visualization tool. For example, parameters of traffic accident prediction models have been estimated mainly based not on KDE but on raw count data in Japan. Yu et al. (2014) recently reported that KDE outperformed other hazardous road segment identification methods. Therefore, the accuracy of traffic accident prediction model might be improved by using KDE.

This study aims to develop traffic accident density models based on KDE as an explained variable. Additionally, the contribution of these models is evaluated from practical and academic perspectives. Regarding practicality, the applicability of these models to other cities is examined. Academically speaking, this study examines the improvement in applicability of using KDE as an explained variable instead of using raw count data as an explained variable.

For this study, a model using KDE as an explained variable is a KDE model. One using raw count data as an explained variable is a raw count data model.

2. Literature review

This section presents a review of the literature about KDE application to traffic accident analysis. As explained above, traffic accidents occur rarely and incidentally in residential areas. When traffic accident hotspots are analyzed based on raw data, the possibility exists that potential hotspots are not detected. Therefore, KDE has been used to detect traffic accident hotspots.

The first report using KDE for traffic accident data was made by Banos and Huguenin-Richard (2000), who mapped the distribution of child pedestrian accidents using KDE. Similarly, several studies have identified spatial clusters of accidents through KDE (Anderson, 2009; Pulugurtha et al., 2007; Schneider et al., 2004). Furthermore, several studies have used KDE to analyze traffic accidents spatially and temporally (Blazquez and Celis, 2013; Plug et al., 2011). Krisp and Durot (2007) mapped a distribution of wildlife–vehicle accidents using KDE. In fact, KDE has been evaluated for detection of traffic accident hotspots. It has also been compared with other methods (Erdogan et al., 2008; Yu et al., 2014). Network kernel density estimation (Network KDE), a method for adapting KDE as a function of networks, has been increasing recently. Loo et al. (2011) and Xie and Yan (2008, 2013) used Network KDE to analyze traffic accidents. The literature reveals that numerous studies have investigated traffic accident analysis by using KDE. However, Xie and Yan (2013) have reported that previous studies have remained at the level of using KDE mainly as a tool for visualization.

Many previous studies have examined relations between traffic accidents and city components (Kim et al., 2006; Noland and Quddus, 2004; Pulugurtha et al., 2013; Quddus, 2008; Wier et al., 2009), and have developed models to estimate traffic accident risk (Hadayeghi et al., 2010; Marshall and Garrick, 2011; Moeinaddini et al., 2014; Rifaat et al., 2011), but the models were developed based on the prior few years of accident data, even though traffic accidents rarely occurred on residential roads. This fact can reduce the accuracy of prediction. Therefore, this study specifically addresses KDE for developing models. Additionally, Yu et al. (2014) reported that, recently, KDE outperformed other hazardous road segment identification methods. It is assumed that analysis based not on finite traffic accident data but on KDE can improve the prediction model applicability.

However, the parameters of traffic accident prediction model have not been estimated by KDE, and have not been evaluated for applicability of the model. Therefore, this study examines whether KDE are useful for the development of a predictive model.

3. Case study

Toyota City and Okayama City were selected as cases for this study. Density estimate models were developed based on the Toyota City database, and the possibility of their application was clarified by assessing their performance for Okayama City.

Toyota City, located in northern Aichi Prefecture, Japan, has an area of 918 km^2, comprising an urban area, suburban areas, and hilly and mountainous areas. It has about 400,000 residents. Okayama City, located in southeastern Okayama Prefecture, Japan, has area of 789 km^2, with geography of various types resembling that of Toyota City. It has about 700,000 residents.

Table 1 presents an outline of the traffic accident database. As the table shows, there are 23,998 accidents occurred between 1999 and 2007 in Toyota City, and 41,833 accidents occurred between 2006 and 2010 in Okayama City. As mentioned in Chapter 1, traffic accident data are difficult to obtain from police departments in many Japanese cities. On the basis of our collaboration with Aichi and Okayama Prefectural Police on the reduction of traffic accidents, traffic accident data have been provided with special permission.

Fig. 1 presents spatial patterns of all accident types (e.g., vehicle–pedestrian accidents, vehicle–vehicle accidents, and single-vehicle accidents) and vehicle–pedestrian accidents in Toyota City (1999–2007). Traffic accidents occur widely in western areas, where urban functions and populations are concentrated.

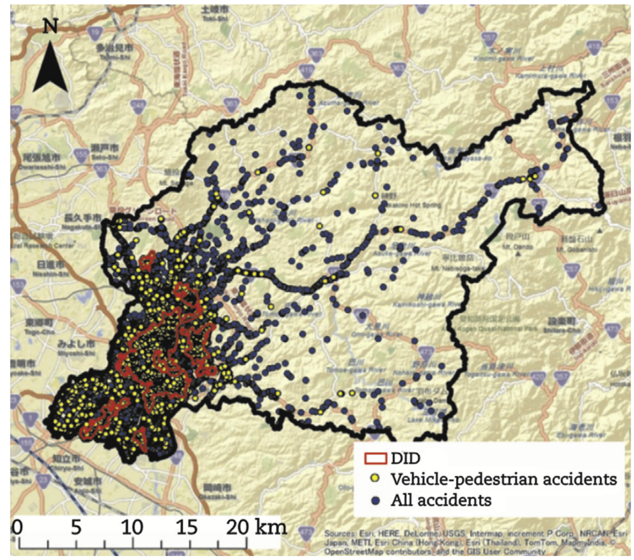

Fig. 1 – Spatial pattern of traffic accident in Toyota City (1999–2007).

4. Development of estimating traffic accident density models

This chapter presents development of models for traffic accident density estimation explained by using GIS data. Parameters of these models based not only on KDE but also on raw count data are estimated to compare the applicability of KDE models with that of raw count data models in the next chapter.

4.1. *KDE*

A well-established method used to identify spatial patterns is KDE, which calculates the density of events around each point, scaled by the distance from the point to each event. KDE describes a smooth and continuous surface map of risk targets because a discrete density surface is made continuously by interpolation. Therefore, this method can compensate for a paucity of data.

Table 1 – Outline of traffic accident database.

Content	Toyota City	Okayama City
Source	Toyota Transportation Research Institute	Okayama Prefectural Police
Year	1999–2007	2006–2010
Number of accidents	23,998	41,833
GIS file format	Points	Points
Main content	• Date, time, and place of accident • Accident types (vehicle–pedestrian accidents, vehicle–vehicle accidents, and single-vehicle accidents) • Severity of accident (fatal accident, severe injury, and minor injury)	

A general density estimation function is shown in Eq. (1).

$$f(x) = \frac{1}{nh}\sum_{i=1}^{n}\frac{K(x - x_i)}{h} \tag{1}$$

where x_i stands for the value of the variable x at location i, n signifies the total number of locations, h denotes the bandwidth or smoothing parameter, K represents the kernel function, as explained in an earlier report (Silverman, 1986).

Several kernel density functions have been proposed. According to Yu et al. (2014), previous studies generally have indicated that the kernel density function selection did not affect the results significantly. However, bandwidth h significantly affects the results. No impeccable measure exists for determining the bandwidth. In this study, the method Ito et al. (2010) reported is used for the estimation.

The method is described as follows: (1) Increase bandwidth at 50 m intervals. (2) Multiple regression analyses (explained variable: estimates of traffic accident density based on KDE) are conducted for each bandwidth. (3) Adopt the bandwidth at which the adjusted coefficient of determination (adjusted R^2) is the highest.

Fig. 2 shows the relations among R^2, adjusted R^2, and bandwidth. 250 m bandwidth is chosen because the figure shows the highest point at 250 m. Fig. 3 portrays the results with applying KDE (bandwidth 250 m) to accidents of all types between 1999 and 2007 in Toyota City.

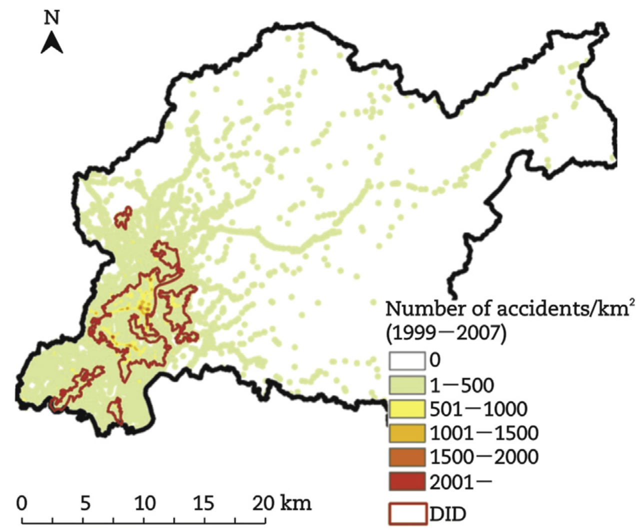

Fig. 3 – Results with applying KDE (bandwidth 250 m) to accidents of all types between 1999 and 2007 in Toyota City.

4.2. *Data descriptions*

Data descriptions are shown as below.

- Toyota City is divided into several areas in 250 m mesh. Consequently, the city has a total of 14,460 meshes. An estimation of traffic accident density mesh-unit data (explained variable) and the component mesh-unit data (explanatory variables) are created for each mesh.
- The unit of density estimates produced using the KDE method is the accident number per square kilometer. The density per year is obtained by dividing the KDE, which applies KDE to accidents between 1999 and 2007, by 9 years. The unit of density per year is the accident number per square kilometer per year. Furthermore, the explained variable is obtained by dividing the density per year by 16 because the mesh is 250 m × 250 m. The unit of the explained variable is the accident number per mesh per year.
- The explained variable for raw count data models is obtained by counting accidents for each mesh and dividing that result by 9 years. The unit of the explained variable is the accident number per mesh per year.

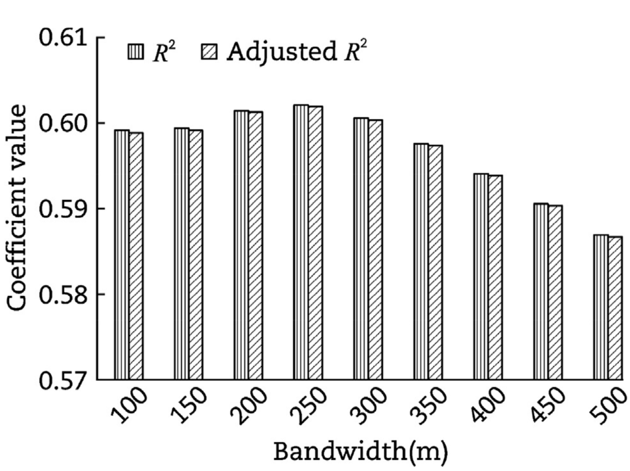

Fig. 2 – Relations between coefficients and bandwidth.

This study was conducted to estimate the traffic accident distribution in the city based on city characteristics. According to Fell (1976), traffic accident factors fell into three broad categories, including human causal chain, vehicle causal chain, and environmental causal chain. Among them, explanatory variables should be selected from the environmental causal chain, if models are applied to ascertain which area should have preferential implementation. Environmental causal chain can be subdivided to include structural factors (e.g., road structure), traffic factors (e.g., traffic stream), and weather factors (e.g., weather condition). Structural factors universally affect the occurrence of traffic accidents. Many data of structural factors are readily available because they are provided as GIS data. Therefore, models with structural factors have wide applications. It is assumed that structural factors should be valued above other factors, thus, structural variables are specifically examined. Variables provided as GIS data in Japan are chosen in Table 2.

- Explained variables are based on accident data between 1999 and 2007, but Table 2 presents explanatory variables in different period. A data year discrepancy exists between the explained variable and explanatory variables. However, that discrepancy has small effects on

Table 2 – Explanatory variable data and associated sources.

Explanatory variable	Description	VIF
Source: e–Stat (Statistics Bureau of General Affairs Agency)		
Population	Total population in mesh based on 2010 population census of Japan	–
Pop younger than 15 years	Pop younger than 15 years in mesh based on 2010 population census of Japan	3.01
Pop aged 15–64 years	Pop 15–64 years old in mesh based on 2010 population census of Japan	–
Pop aged over 65 years	Pop over 65 years old in mesh based on 2010 population census of Japan	–
Pop aged over 75 years	Pop over 75 years old in mesh based on 2010 population census of Japan	3.32
Source: ArcGIS data collection premium series 2010 road network (ESRI Japan)		
Intersection	Number of intersections, including unsignalled intersection in mesh	–
Signalized intersection	Number of signalized intersections in mesh	1.67
Intersection on prefectural road	Number of intersections on prefectural road in mesh	1.21
Intersection on national road	Number of intersections on national road in mesh	1.29
Road length (m)	Total length of road in mesh	–
Road length (road width: 3.0–5.5 m) (m)	Total road width: 3.0–5.5 m in mesh	1.60
Road length (road width: 5.5–13.0 m) (m)	Total road width: 5.5–13.0 m in mesh	1.51
Road length (road width: over 13.0 m) (m)	Total road width: over 13.0 m in mesh	1.13
Source: Zmap–AREA II, ZENRIN residential maps 2011 (ZENRIN)		
Building area (m^2)	Total building area in mesh	1.92
Source: national land numerical information download service (ministry of land, infrastructure and transport)		
Public facilities	Number of public facilities, including public offices, schools, post offices, and social welfare facilities	1.11
Healthcare facilities	Number of hospitals, health clinics, and dental clinics in each mesh	1.22

Table 3 – Negative binomial regression models (accidents in all areas).

Accident places	Accidents in all areas					
Data types	KDE			Raw count data		
Accident types	All	Vehicle–pedestrian	Minor	All	Vehicle–pedestrian	Minor
Model	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Pop aged younger than 15 years	0.005	0.011	0.005	0.005	0.020	0.005
	0.000**	0.028*	0.001**	0.002**	0.000**	0.002**
Pop aged over 75 years	−0.015	0.023	−0.015	−0.018	0.002	−0.018
	0.000**	0.093	0.000**	0.000**	0.882	0.000**
Signalized intersection	0.310	0.471	0.317	0.502	0.346	0.512
	0.000**	0.031*	0.000**	0.000**	0.040*	0.000**
Intersection on prefectural road	0.133	0.094	0.137	0.165	0.102	0.166
	0.000**	0.452	0.000**	0.000**	0.287	0.000**
Intersection on national road	0.087	−0.005	0.094	0.130	0.194	0.131
	0.000**	0.962	0.000**	0.000**	0.004**	0.000**
Road length (road width: 3.0–5.5 m) (m)	0.001	0.001	0.001	0.001	0.000	0.001
	0.000**	0.256	0.000**	0.000**	0.236	0.000**
Road length (road width: 5.5–13.0 m) (m)	0.002	0.002	0.002	0.003	0.002	0.003
	0.000**	0.013*	0.000**	0.000**	0.001**	0.000**
Road length (road width: over 13.0 m) (m)	0.004	0.001	0.004	0.005	0.002	0.005
	0.000**	0.800	0.000**	0.000**	0.108	0.000**
Building area (m^2)	6.61×10^{-5}	4.99×10^{-5}	6.68×10^{-5}	6.10×10^{-5}	5.01×10^{-5}	6.03×10^{-5}
	0.000**	0.001**	0.000**	0.000**	0.000**	0.000**
Public facilities	0.394	0.483	0.375	0.279	0.530	0.285
	0.000**	0.014*	0.000**	0.000**	0.002**	0.000**
Healthcare facilities	0.344	0.328	0.336	0.406	0.431	0.404
	0.000**	0.038*	0.000**	0.000**	0.002**	0.000**
Constant	−3.647	−8.285	−3.698	−3.768	−7.560	−3.794
	0.000**	0.000**	0.000**	0.000**	0.000**	0.000**
Number of observations	8670	8670	8670	8670	8670	8670
Log-likelihood	−3497.630	−122.173	−3398.555	−3352.313	−218.458	−3277.432
AIC	7019.259	268.346	6821.110	6728.627	460.917	6578.864

Note: upper row means estimated coefficient; lower row means p-value (** means p-value <0.01, * means p-value <0.05).

results because no great urban structure alteration occurred after 1999 in Toyota City. Therefore, these data for analyses are used.

4.3. Relations between traffic accidents and city characteristics

Sixteen model types were developed by combining accident places, accident types, and data types of the explained variable. First, accidents were classified based on accident locations, such as accidents in all areas, accidents only in densely-inhabited districts (DID), and accidents only on city roads. Analysis of accidents in all area is intended to develop a model that can reflect the accident distribution throughout the city. Analysis of accidents in densely-inhabited district is intended to develop a model that is specialized for estimating the accident distribution in DID. Analysis of accidents only on city roads is intended to develop a model that assigns importance to accidents in residential areas. Second, accidents were classified based on accident types, all types of accidents (e.g., vehicle–pedestrian accidents, vehicle–vehicle accidents, single-vehicle accidents), vehicle–pedestrian accidents, and minor accidents. Severe injury accidents and fatal accidents can not be analyzed because accident data are scarce. Third, models are developed to estimate traffic accident density based on KDE and raw count data. Negative binomial regression is adopted for this analysis. It is commonly used to model count data (traffic accident data).

The explanatory variable of the analysis target is selected to avoid multicollinearity. Multicollinearity is checked by calculating the variance inflation factors (VIFs). VIFs are less than 10 when the explanatory variables, such as population, pop aged 15–64 years, pop aged over 65 years, intersection, and road length, are removed. Table 2 presents the results for the VIFs. Tables 3–5 present the results of parameters estimation using negative binomial regression. Meshes in roadless areas (road length is 0 m in mesh) are removed from analysis. Parameters of models 14 and 17 can not be estimated because of the scarcity data for vehicle–pedestrian accidents on city roads.

Comparison between the KDE model and the raw count data model shows that the respective significant variables of the models are not much different. Furthermore, comparison of accident types, all with minor, shows that significant variables and parameter values of the models are not much different between these two accident types because many accidents are minor accidents. Understandably, variations in the number of intersections and road length have significant effects on many models. Public facilities and healthcare

Table 4 – Negative binomial regression models (accidents only in DID).

Accident places	Accidents only in DID					
Data types	KDE			Raw count data		
Accident types	All	Vehicle–pedestrian	Minor	All	Vehicle–pedestrian	Minor
Model	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12
Pop aged younger than 15 years	0.005	0.006	0.005	0.006	0.013	0.006
	0.003**	0.263	0.002**	0.000**	0.000**	0.001**
Pop aged over 75 years	−0.005	0.023	−0.006	−0.010	−0.001	−0.010
	0.263	0.097	0.207	0.040*	0.953	0.045*
Signalized intersection	0.277	0.449	0.290	0.388	0.382	0.387
	0.000**	0.047*	0.000**	0.000**	0.029*	0.000**
Intersection on prefectural road	0.083	0.078	0.086	0.112	0.038	0.110
	0.028*	0.533	0.022*	0.003**	0.700	0.003**
Intersection on national road	0.035	−0.008	0.032	0.056	0.136	0.057
	0.240	0.941	0.296	0.055	0.051	0.050
Road length (road width: 3.0–5.5 m) (m)	0.000	0.000	0.000	0.000	0.000	0.000
	0.657	0.564	0.614	0.446	0.390	0.545
Road length (road width: 5.5–13.0 m) (m)	0.001	0.001	0.001	0.001	0.001	0.001
	0.001**	0.351	0.002**	0.000**	0.207	0.000**
Road length (road width: over 13.0 m) (m)	0.002	0.000	0.002	0.002	0.001	0.002
	0.000**	0.845	0.001**	0.000**	0.327	0.000**
Building area (m^2)	4.77×10^{-5}	2.93×10^{-5}	4.83×10^{-5}	4.52×10^{-5}	3.09×10^{-5}	4.53×10^{-5}
	0.000**	0.344	0.000**	0.000**	0.172	0.000**
Pubic facilities	0.167	0.406	0.137	0.124	0.445	0.110
	0.047*	0.038*	0.105	0.147	0.007**	0.202
Healthcare facilities	0.265	0.274	0.267	0.276	0.337	0.271
	0.000**	0.071	0.000**	0.000**	0.009**	0.000**
Constant	−1.461	−5.560	−1.509	−1.713	−4.918	−1.715
	0.000**	0.000**	0.000**	0.000**	0.000**	0.000**
Number of observations	872	872	872	872	872	872
Log-likelihood	−1245.269	−105.205	−1219.951	−1218.782	−175.099	−1204.231
AIC	2514.538	234.41	2463.902	2461.565	374.198	2432.463

Note: upper row means estimated coefficient; lower row means *p*-value (** means *p*-value <0.01, * means *p*-value <0.05).

Table 5 – Negative binomial regression models (accidents only on city roads).

Accident places	Accident only on city road					
Data types	KDE			Raw count data		
Accident types	All	Vehicle–pedestrian	Minor	All	Vehicle–pedestrian	Minor
Model	Model 13	Model 14	Model 15	Model 16	Model 17	Model 18
Pop aged younger than 15 years	0.008	–	0.008	0.006	–	0.006
	0.000**	–	0.000**	0.000**	–	0.001**
Pop aged over 75 years	−0.010	–	−0.009	−0.013	–	−0.013
	0.042*	–	0.064	0.009**	–	0.013*
Signalized intersection	0.511	–	0.505	0.699	–	0.700
	0.000**	–	0.000**	0.000**	–	0.000**
Intersection on prefectural road	−0.205	–	−0.195	−0.296	–	−0.303
	0.000**	–	0.000**	0.000**	–	0.000**
Intersection on national road	−0.284	–	−0.277	−0.357	–	−0.348
	0.000**	–	0.000**	0.000**	–	0.000**
Road length (road width: 3.0–5.5 m) (m)	0.001	–	0.001	0.001	–	0.001
	0.000**	–	0.000**	0.000**	–	0.000**
Road length (road width: 5.5–13.0 m) (m)	0.002	–	0.002	0.003	–	0.003
	0.000**	–	0.000**	0.000**	–	0.000**
Road length (road width: over 13.0 m) (m)	0.003	–	0.003	0.004	–	0.004
	0.000**	–	0.000**	0.000**	–	0.000**
Building area (m^2)	5.98×10^{-5}	–	6.02×10^{-5}	5.62×10^{-5}	–	5.61×10^{-5}
	0.000**	–	0.000**	0.000**	–	0.000**
Public facilities	0.431	–	0.458	0.313	–	0.301
	0.000**	–	0.000**	0.000**	–	0.000**
Healthcare facilities	0.392	–	0.386	0.455	–	0.469
	0.000**	–	0.000**	0.000**	–	0.000**
Constant	−4.483	–	−4.541	−4.421	–	−4.441
	0.000**	–	0.000**	0.000**	–	0.000**
Number of observations	8670	–	8670	8670	–	8670
Log-likelihood	−1981.142	–	−1914.613	−2064.87	–	−2020.42
AIC	3986.285	–	3853.226	4153.739	–	4064.84

Note: upper row means estimated coefficient; lower row means p-value (** means p-value <0.01, * means p-value <0.05).

facilities also have a significant effect on many models. It can be inferred that these facilities attract many pedestrians and vehicles, and that traffic accidents occur frequently around these facilities. Consequently, these variables have significant effects on many models.

5. Application of traffic accident models

The preceding chapter has developed a traffic accident density estimation models based on a database for Toyota City. This chapter presents the evaluation of the contributions of the models from practical and academic perspectives. The applicability of these models is assessed for other cities. Additionally, the improvement in applicability of KDE models compared with that of raw count data models is examined.

Okayama City is separated into areas of 250 m mesh, therefore, it has 12,697 meshes. The component mesh-unit data are created for each mesh.

The number of accidents in Okayama City is predicted by Model 1 as an example. Predicted values are rounded to integers. As described in chapter 1, the purpose of this study is to develop models that enable the determination of areas objectively and easily, and which indicate where to implement area-wide traffic calming preferentially. Therefore, each mesh is ranked according to the actual number of accidents. Similarly, each mesh is ranked according to the predicted number of accidents. Fig. 4 presents the order by the actual number of traffic accidents (5-year average) and their predicted number. As the figure shows, the model can predict numerous traffic accidents in the south, but predict fewer in the north.

Therefore, to clarify the applicability of KDE models and its improvement, the Spearman rank correlation coefficient between the mesh ranking according to the predicted number and the one ranking according to the actual number are calculated. Table 6 presents the calculation results of the Spearman rank correlation coefficient.

The predicted numbers of all accidents and minor accidents and their actual number have a strong positive correlation. However, the Spearman rank correlation coefficient of vehicle–pedestrian is not high. The models predicting all accidents and minor accidents are likely to show applicability to other cities. The improvement of the applicability of KDE models compared with that of raw count data models is examined next. Comparison of Spearman rank correlation

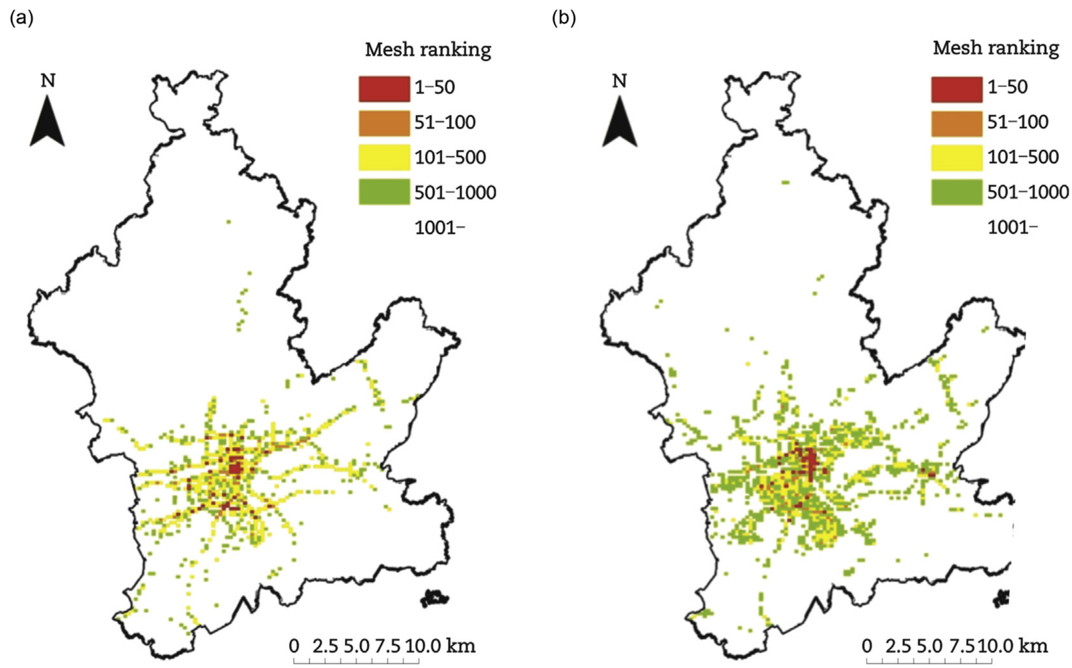

Fig. 4 – Mesh rank of the actual number of traffic accidents in Okayama City. (a) Rank of the actual number of accidents per year. (b) Rank of the predicted number of accidents per year.

coefficients shows that no great difference is found in the correlation coefficients between KDE models and raw count data models.

6. Conclusions

This study assessed relations between traffic accidents and city characteristics. Models were developed to estimate the traffic accident density. Additionally, this study examined the improvement in the applicability of using KDE as an explained variable compared with that of using raw count data.

The following results were obtained.

Table 6 – Spearman rank correlation coefficient between the mesh ranking according to the predicted number and the one ranking according to the actual number.

Accident location	Accident type	Data type	
		KDE model	Raw count data model
Accident in all areas	All	0.70	0.69
	Vehicle–pedestrian	0.38	0.36
	Minor	0.69	0.68
Accident only in DID	All	0.74	0.75
	Vehicle–pedestrian	0.40	0.38
	Minor	0.73	0.74
Accident only on city road	All	0.60	0.60
	Vehicle–pedestrian	–	–
	Minor	0.61	0.61

- KDE was used to examine the relation in an efficient manner. Sixteen model types were examined by combining accident places, accident types, and data types. Understandably, variables of the number of intersections and road length have a significant effect on many models. A significant effect was found for public facilities and healthcare facilities in many models. It can be inferred that these facilities attract many pedestrians and vehicles, and that traffic accidents occur frequently around these facilities. Consequently, these variables have a significant effect on many models.
- The applicability of density estimation for traffic accident models was examined. Results show that the Spearman rank correlation coefficient between the predicted number and the actual number is strong. Model predictions for all and minor accidents are likely to be applicable to other cities.
- Comparison using Spearman rank correlation coefficients reveals that no greater difference is found in the correlation coefficient between KDE models and raw count data models.

First, the practical contributions of this study are the following. The predicted numbers estimated by developed models and the actual numbers of accidents show a strong positive correlation. The model can reveal the relative accident risk in a city, even if traffic accident data are not available. The results of this study objectively indicate areas in which area-wide traffic calming should be implemented preferentially. Second, the academic contribution of this study is that KDE models achieve slight improvement in applicability among cities. This result shows that KDE data can develop a prediction model compared well with raw count data. For example, it often happens that information of crime locations is distributed to the public not in point data format but in density data format in Japan because of the need of the personal information protection. This study and its results can expand the application of density data in various fields.

Challenges for future investigation of this topic are the following.

- In this study, the developed model was applied to Okayama City because of the limitation of data. These models should be applied to other cities to verify their accuracy.
- Recently, Network KDE has been applied to traffic accident data. Regression analysis based upon the estimation of traffic accident density-based Network KDE as an explained variable is a subject for future analysis.

Acknowledgments

This research is supported by a grant from the Takata Foundation.

REFERENCES

Anderson, T.K., 2009. Kernel density estimation and K–means clustering to profile road accident hotspots. Accident Analysis & Prevention 41 (3), 359–364.

Banos, A., Huguenin-Richard, F., 2000. Spatial distribution of road accidents in the vicinity of point sources application to child pedestrian accidents. Geography and Medicine 8, 54–64.

Blazquez, C.A., Celis, M.S., 2013. A spatial and temporal analysis of child pedestrian crashes in Santiago, Chile. Accident Analysis & Prevention 50 (2), 304–311.

Erdogan, S., Yilmaz, I., Baybura, T., et al., 2008. Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar. Accident Analysis & Prevention 40 (1), 174–181.

Fell, J.C., 1976. A motor vehicle accident causal system: the human element. Human Factors: The Journal of the Human Factors & Ergonomics Society 18 (1), 85–94.

Hadayeghi, A., Shalaby, A.S., Persaud, B.N., 2010. Development of planning level transportation safety tools using geographically weighted Poisson regression. Accident Analysis & Prevention 42 (2), 676–688.

Institute for Road Safety Research, 2004. Zones 30: Urban Residential Areas. Institute for Road Safety Research, Leidschendam.

Ito, F., Itogawa, E., Umemoto, M., 2010. Occurrence factors from a microscale environmental characteristics point of view: case study of bag snatching in Itabashi Ward, Tokyo. Journal of Social Crime Safety Science (13), 109–118.

Kim, K., Brunner, I.M., Yamashita, E.Y., 2006. Influence of land use, population, employment, and economic activity on accidents. Transportation Research Record 1953, 56–64.

Krisp, J.M., Durot, S., 2007. Segmentation of lines based on point densities—an optimisation of wildlife warning sign placement in Southern Finland. Accident Analysis & Prevention 39 (1), 38–46.

Loo, B.P.Y., Yao, S., Wu, J., 2011. Spatial point analysis of road crashes in Shanghai: a GIS-based network kernel density method. In: The 19th International Conference on Geoinformatics, Shanghai, 2011.

Marshall, W.E., Garrick, N.W., 2011. Does street network design affect traffic safety? Accident Analysis & Prevention 43 (3), 769–781.

Moeinaddini, M., Asadi-Shekari, Z., Shah, M.Z., 2014. The relationship between urban street networks and the number of transport fatalities at the city level. Safety Science 62 (2), 114–120.

Noland, R.B., Quddus, M.A., 2004. A spatially disaggregate analysis of road casualties in England. Accident Analysis & Prevention 36 (6), 937–984.

Plug, C., Xia, J.C., Caulfield, C., 2011. Spatial and temporal visualisation techniques for crash analysis. Accident Analysis & Prevention 43 (6), 1937–1946.

Pulugurtha, S.S., Krishnakumar, V.K., Nambisan, S.S., 2007. New methods to identify and rank high pedestrian crash zones: an illustration. Accident Analysis & Prevention 39 (4), 800–811.

Pulugurtha, S.S., Duddu, V.R., Kotagiri, Y., 2013. Traffic analysis zone level crash estimation models based on land use characteristics. Accident Analysis & Prevention 50 (1), 678–687.

Quddus, M.A., 2008. Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data. Accident Analysis & Prevention 40 (4), 1486–1497.

Rifaat, S.M., Tay, R., de Barros, A., 2011. Effect of street pattern on the severity of crashes involving vulnerable road users. Accident Analysis & Prevention 43 (1), 276–283.

Schneider, R.J., Ryznar, R.M., Khattak, A.J., 2004. An accident waiting to happen: a spatial approach to proactive pedestrian planning. Accident Analysis & Prevention 36 (2), 193–211.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.

Wier, M., Weintraub, J., Humphreys, E.N., et al., 2009. An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. Accident Analysis & Prevention 41 (1), 137–145.

Xie, Z., Yan, J., 2008. Kernel density estimation of traffic accidents in a network space. Computers, Environment and Urban Systems 32 (5), 396–406.

Xie, Z., Yan, J., 2013. Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach. Journal of Transport Geography 31 (5), 64–71.

Yu, H., Liu, P., Chen, J., et al., 2014. Comparative analysis of the spatial analysis methods for hotspot identification. Accident Analysis & Prevention 66 (1), 80–88.