Spatial and temporal distribution of pedestrian crashes in Melbourne metropolitan area

Alireza Toran Pour, Sara Moridpour, Abbas Rajabifard and Richard Tay

Abstract

About 1100 vehicle-pedestrian crashes occur in Melbourne metropolitan area every year. Identifying the temporal and spatial patterns of pedestrian injuries is essential to enhance the safety of these vulnerable road users. In this paper, Decision Tree (DT) and interactive DT are applied to identify the influence of temporal, spatial and personal characteristics on vehicle-pedestrian crash severity. DT is a simple but powerful form of data analyses using machine learning technique. Result of DT indicates that time of crash is the most significant variable in classifying and predicting the severity of vehicle-pedestrian crashes in Melbourne metropolitan area. According to this model, accidents occurring between 7:00 p.m. and 6:00 a.m. are more severe than other times. Moreover, spatial correlation shows that there are positive correlation between time and location of crashes. Kernel Density Estimation (KDE) is applied to explore the spatial distribution of vehicle-pedestrian crashes. KDE results show that most vehicle-pedestrian crashes between 7:00 p.m. and 6:00 a.m. occur around hotels, clubs and bars. Safety measures should be applied around these areas to assist in preventing and reducing the severity of vehicle-pedestrian crashes.

1. INTRODUCTION

Background

Pedestrians are known as vulnerable road users in road safety literature because they are more likely to be harmed or injured in traffic crashes. Pedestrians are about four times more likely to be injured in traffic crashes than other road users (Elvik et al. 2009). In addition, because their body is exposed and unprotected in traffic crashes, they are 23 times more likely to be killed than vehicle occupants (Miranda-Moreno, Morency & El-Geneidy 2011). According to the World Health Organisation’s report, every year about 1.24 million people are...
killed in traffic accidents in the world and more than 22% of these deaths are pedestrians (WHO 2013). In Australia, vehicle-pedestrian crashes account for more than 13% of total fatal crashes. Every year, pedestrians are involved in about 1100 traffic crashes in Melbourne and about 38 pedestrians are killed in these traffic accidents, which comprise about 18% of total pedestrian fatalities in Australia (ABS 2010; VicRoads 2016). Pedestrians and other vulnerable road users are specifically targeted in the recent Road Safety Agenda of the Victorian government (VicRoads 2015). Design and implementation of effective countermeasures to improve the safety of these vulnerable road users will require a better understanding of the major crash contributing factors as well as the temporal-spatial patterns of vehicle-pedestrian crashes.

Objectives and scope of study
The aim of this paper is to identify the influence of temporal and spatial variables on vehicle-pedestrian crash severity. Decision Tree (DT) and Cross-Validation (CV) techniques are used to analyse these temporal and spatial patterns in vehicle-pedestrian crashes in Melbourne metropolitan area. In addition, spatial autocorrelation is applied in examining the vehicle-pedestrian crashes in Geographic Information Systems (GIS) to identify any dependency between time and location of these crashes. Kernel Density Estimation (KDE) is then used in the GIS environment to determine spatial patterns of vehicle-pedestrian crashes at different time periods. The novelty of this study is in the application of these three techniques together to explore the spatial and temporal patterns of crashes.

Furthermore, this research will apply interactive DT to identify the significant crash contributing factors. It will provide valuable insights that can be used to guide decisions in allocating human, financial and time resources to enhance road safety of pedestrians. In addition, the influence of Points Of Interest (POIs) on vehicle-pedestrian crashes is analysed in this study. POIs are locations that people may find useful or interesting to visit (e.g. hotels, shops, parks). Despite its importance in generating and attracting vehicular and pedestrian traffic, little research has been done to examine the role of POIs in vehicle-pedestrian collisions.

In the next section of the paper, the literature on vehicle-pedestrian crashes will be reviewed, with a focus on temporal and spatial analysis. In section three, an introduction to DTs, spatial autocorrelation and KDE is presented together with a description of the data and methodology used in this research. The results of this research will be presented afterwards. The final section of the paper will provide a summary of the outcomes and presents directions for future research.

2. LITERATURE REVIEW
The review of published research on pedestrian crashes found many studies that merely focused on spatial pattern of pedestrian crashes. In those studies, different statistical models were developed to identify spatial variables that influence pedestrian crashes (Schneider, Ryznar & Khattak 2004; Siddiqui, Abdel-Aty & Choi 2012). For instance, Siddiqui et al. (2012) applied a Bayesian spatial technique to model pedestrian and bicycle crashes in traffic analysis zones and found a spatial correlation between pedestrian and bicycle crashes. In other studies, KDE is applied to identify pedestrian crash patterns and black spots (Schneider et al. 2004; Tuong & Somenahalli 2011).

On the other hand, relatively more studies have conducted spatial and temporal analyses of motor vehicle crashes. Black (1991) applied temporal, spatial and spatial-temporal autocorrelation analysis techniques to examine highway collisions on an Indiana toll road in U.S. He applied von Neumann’s ratio, Moran’s I, nearest-neighbour analysis, and a spatial-temporal autocorrelation coefficient to show the applicability of these techniques in temporal and spatial collision analysis. In another study, Levine, Kim and Nitz (1995) examined spatial patterns in motor vehicle crashes in Honolulu, U.S. They used GIS analysis to describe the spatial distribution of crash locations in their study area. In addition, Andrey and Yagar (1993) conducted a temporal analysis to examine the collision risk during and after rain events in Calgary and Edmonton in Canada. They applied a matched sample approach to examine the crash data between 1979 and 1983. Literature review shows that KDE widely is used to identify vehicle crash black spots (Sandhu et al. 2016; Thakali, Kwon & Fu 2015; Xie & Yan 2008, 2013). Moreover, spatial autocorrelation has been applied in some research to show the correlation between variables influencing vehicle crashes (Flahaut et al. 2003; Shen, Li & Si 2016).

Agueiro-Valverde and Jovanis (2006) applied Full Bayes (FB) hierarchical models with spatial and temporal effects and space-time interactions to examine injury and fatal crashes in Pennsylvania, U.S. They found spatial correlation in their crash data and that this correlation was more important in road segment and intersection-level crash models than county-level FB models. In another study, Li, Zhu and Sui (2007) used a GIS-based Bayesian
approach to analyse the spatial-temporal patterns of motor vehicle crashes in Houston, U.S. They found the spatial-temporal analysis method to be useful in identifying and ranking roadway segments with high risk of vehicle crashes.

Plug, Xia and Caulfield (2011) used spatial, temporal and spatiotemporal techniques in GIS to study single vehicle crash patterns in Western Australia. In this study, they used visualisation techniques such as KDE and different plots. In this technique, linear, spider, and circular graphs were applied to illustrate the temporal patterns of crashes. Their results showed that there were significant differences in spatial and temporal patterns of single vehicle crashes.

The spatial-temporal analyses conducted thus far have mainly examined motor vehicle crashes as a whole and did not focus on vulnerable road users. In one of the few studies that focussed on vulnerable road users, Blasquez and Celis (2013) applied this approach to examine vehicle crashes involving child pedestrians in Santiago, Chile. In this study, they applied KDE to identify the critical areas for child pedestrian safety. They then applied Moran’s Index to identify the correlation between spatial and temporal variables for those crashes.

In summary, the review of published literature revealed that there are limited studies focusing on temporal and spatial analyses of motor vehicle crashes, and fewer numbers that focussed on pedestrians and other vulnerable road users. Since vehicle-pedestrian crashes would have significantly different crash characteristics from vehicle-vehicle crashes, studies focusing on vehicle-pedestrian crashes would provide useful insights to improve the safety of these vulnerable road users. KDE and Moran’s I are applied in some studies; however, applying DT and interactive DT is a new idea in temporal and spatial analysis of traffic crashes. In addition, the literature shows that different approaches have been used to show the distribution of crashes. However, the effect of POIs has not been considered. It is the first time that POI is considered as an important factor in vehicle-pedestrian crash modelling and analysis.

3. METHODOLOGY

The dataset used in this study is consisting of spatial and non-spatial variables. The non-spatial variables are used to develop the DT model, identify the influencing variables or explanatory variables on vehicle-pedestrian crashes and analyse the temporal distribution of these crashes. The spatial variables are used to display spatial patterns of these crashes while the DT model is applied to explore the influence of time and location of vehicle-pedestrian crashes. Different spider plots are used to show the temporal distribution of these crashes at different time periods. Furthermore, Moran’s Index and KDE are used to show spatial autocorrelation between locations and times of crashes and to identify areas with high risks of vehicle-pedestrian crashes. Figure 1 shows the methodology of this research.

**Decision Tree (DT)**

In this paper, DT model is applied to identify the contribution of time to pedestrian crash severity. A simple decision tree is a set of procedures used to classify the input training data into more homogenous subgroups using generated rules or decisions called nodes (Friedl & Brodley 1997). A DT that is used to predict a categorical variable is called a classification tree. A decision tree developed for predicting a continuous variable is called a regression tree, while a classification and regression tree can be used to develop both classification and regression trees. One advantage of using the DT is that it is a non-parametric model and there is no need to pre-define any underlying relationship or assumptions between the dependent and independent variables. Moreover, in statistical models such as logit or probit models, when the contributions of various risk factors need to be
assessed, it is important to try to account for the correlation between risk factors. Otherwise, the contributions of a set of risk factors to severity of crashes may be counted more than one time (Elvik et al. 2009). Therefore, the potential for severity reduction by improving or controlling the risk factors may be overestimated. In contrast, when DT analysis is applied, the correlation problems between independent variables would not be a great concern (Chang & Wang 2006; Kashani & Mohaymany 2011).

Figure 2 shows the principle of the DT method in developing the classification tree. The process starts from the root node, which contains all the objects in the data set. The root node is split into two nodes by a rule. These two new nodes are called child nodes. In general, a rule must be chosen to make the resulting child nodes as homogeneous as possible. In the next step, this process is repeated and each child node is assumed as a parent node. This process is continued until no further split can be made (either all child nodes are homogeneous or a user-defined minimum number of objects in the node is reached). These final nodes are called terminal nodes or leaves and they have no branches.

The principle to split a node in DT is based on decreasing impurity (heterogeneity or variance) in the terminal nodes. There are different measures of node impurity. However, one of the most popular measures is the Gini criterion. Gini criterion is defined as:

$$i(t) = \sum_{j \in J} p(i|t)p(j|t)$$  \hspace{1cm} (1)

where:

- $i(t)$ is a measure of the impurity of node $t$.
- $p(i|t)$ is the node’s proportions (the cases in node $t$ belonging to class $j$).

The partitioning will be terminated when all possible threshold values for all explanatory variables (splitters) have been assessed to find the greatest improvement in the purity score of the resultant nodes. The quality of the split is used to measure the decrease in impurity between a parent node and its children, and it is defined as follows:

$$\Delta i(s,t) = i(t) - p_Ri(t_R) - p_Li(t_L)$$  \hspace{1cm} (2)

where:

- $s$ is a candidate split.
- $p_L$ and $p_R$ are the proportions of observations of parent node $t$ related to child node $t_L$ and $t_R$ respectively.

The splitter related to the maximum $\Delta i(s,t)$ is the optimum impurity reduction.

In DT models, pruning a tree is removing the branches with little impact on predictive value of the tree. In other words, pruning is used to simplify the structure of the DT, make a smaller tree and prevents over-fitting. The pruning process starts with the maximal tree and DT uses the cost-complexity algorithm to decrease misclassification rate of tree. This algorithm starts with defining the misclassification cost for a node and a tree. Equation 3 shows the definition of misclassification.

$$r(t) = 1 - p(j|t)$$  \hspace{1cm} (3)

In this equation, $r(t)$ is the misclassification for node $t$, and the tree misclassification cost $R(T)$ can be defined as:

$$R(T) = \sum r(t)p(t)$$  \hspace{1cm} (4)

Misclassification rate is a measure of the total leaf impurity in a decision tree. This rate varies between
Relative importance of variables is a key output of DT models. The importance of a variable in DT model is defined as Equation 5:

$$VIM(X) = \sum_{t=1}^{T} \frac{n_t}{N} \Delta Gini(S(x_t, t))$$

where

- $\Delta Gini(S(x_t, t))$ is the reduction in the Gini index at node $t$ that is achieved by the splitting variable $x_t$,
- $n_t/N$ is the proportion of the observation in the dataset that belongs to the node $t$,
- $T$ is the total number of nodes,
- $N$ is the total number of observations.

In DT models, normalised importance of each variable with respect to the class variable is obtained by dividing $VIM(X)$ by the largest value obtained for a variable.

One disadvantage of DT models is that they use a random seed number to develop trees and identify the influencing variables in crashes. This means that DTs use a random sample data to develop the tree, and unselected data is not considered in developing the tree model. Thus, if the DT is applied to one dataset for many times, the results may be unstable and different trees might be generated by small changes in the selected training data.

To increase the accuracy of DT models, Cross-Validation (CV) technique is applied to identify the datasets used to develop the DT. CV is a statistical method to evaluate and compare different learning algorithms. In this technique, the data is divided into two segments: one used for learning or training a model and the rest of the data is used for validating the model. One of the CV techniques is K-fold CV with repeated rounds. In this method, the data is portioned into $k$ equal size segments or folds. Subsequently, $k$ iterations of training and validation are performed and in each iterations $k-1$ folds are used to train the data and one fold is used for validation. This procedure is repeated $n$ times to improve the accuracy of model. In this study, according to literature review a 10-fold with 2 repeated rounds is used to develop a more accurate and robust DT model (Blockeel & Struyf 2002; Depren et al. 2005; Kohavi 1995).

**Spatial autocorrelation**

After identifying the contribution of time and location on pedestrian accidents, a spatial autocorrelation test is performed to determine the spatial dependency between time and location of vehicle-pedestrian crashes. In this research, the Moran’s Index is used to find this autocorrelation. Moran’s I measures the spatial autocorrelation between two features, which in this study, are the location and time of vehicle-pedestrian crashes. For a set of features and their associated attributes, Moran’s I values range from −1 (indicating perfect dispersion or random) to +1 (perfect correlation) which is shown in Figure 3. This index can be calculated using Equation 6.

$$I = \frac{n}{S_0} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} c_i c_j \sum_{i=1}^{n} c_i^2$$

(6)
where:

- $c_i$ = deviations of an attribute for feature $i$ from its mean $\mu_i$
- $w_{ij}$ is the spatial weight between features $i$ and $j$
- $n$ is equal to the number of features
- $S_0$ is the aggregate of all spatial weights:

$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}$$  \hspace{1cm} (7)

The $z_i$ score that indicates whether or not we can reject the null hypothesis (there is no spatial clustering) is computed as:

$$z_i = \frac{I - E[I]}{\sqrt{V[I]}}$$  \hspace{1cm} (8)

where:

- $E[I] = -1/(n - 1)$
- $V[I] = E[I^2] - E[I]^2$  \hspace{1cm} (9)

**Identifying areas with high risk of crash**

Kernel density estimation involves placing a symmetrical surface over each point and then evaluating the distance from the point to a reference location and then summing the value for all the surfaces for that reference location. This procedure is repeated for successive points. This allows us to place a kernel over each observation, and give us the density estimate of the distribution of collision points. This surface has a maximum value at the reference point and this value decreases with increases in distance from the reference point and reaches zero at the radius distance from reference point (Pulugurtha, Krishnakumar & Nambisan 2007). One common mathematical function used is:

$$f(x,y) = \frac{1}{nh^2} \sum_{i=1}^{n} K\left(\frac{d_i}{h}\right)$$  \hspace{1cm} (10)

where:

- $f(x,y)$ is the density estimation at location $(x,y)$
- $n$ is the number of observations
- $h$ is the bandwidth or kernel size or smooth parameter
- $K$ is the kernel function

While in a simple density method, a circular neighbourhood is considered around each cell, in kernel method, the research area is divided into predetermined number of cells. Thus, the kernel method draws a circular neighbourhood around each feature point (here each vehicle-pedestrian crash). Equation 10 which ranges between 1 at the position of the crash and 0 at the neighbourhood boundary is used in this research (see Figure 4).

There are different types of kernel functions, such as Gaussian, Quartic, Conic, negative exponential, and epanichennêk (Kuter, Usul & Kuter 2011; Levine 2002). According to the literature, the accuracy of kernel function type, $k$, is less important than the impact of bandwidth, $r$, in KDE (Loo, Shenjun & Jianping 2011; O'Sullivan & Unwin 2014; Schabenberger & Gotway 2004; Silverman 1986; Xie & Yan 2008). In this research, the Quartic kernel which is one of the three most common types of kernel functions is applied (Schabenberger & Gotway 2004; Xie & Yan 2008). The specific form of the Quartic kernel function is:

$$K\left(\frac{d_i}{h}\right) = K\left(1 - \frac{d_i^2}{h^2}\right)$$  \hspace{1cm} (11a)

$$K\left(\frac{d_i}{h}\right) = 0$$  \hspace{1cm} (11b)

In Equation 11a,

- $k$ is the kernel function
- $d_i$ is the distance between the location $(x,y)$ and the location of $i^{th}$ observation.

In Equation 11b, $K$ is a scaling factor and its purpose is to ensure the total volume under Quartic curve...
Spatial and temporal distribution of pedestrian crashes in Melbourne metropolitan area

12

is 1. The common values used for $K$ include $3/\pi$ and $3/4$.

Many studies have shown that the selection of bandwidth or smoothing parameter in KDE is subjective (Bil, Andrášik & Janoška 2013; Plug et al. (2011). Different studies have selected different bandwidth values according to area of study and size of dataset. In general, according to 

\[ f(x,y) \text{ with large bandwidth values (} h \rightarrow \infty \text{) will decrease the density (} f(x,y) \rightarrow 0 \text{) and will show significant smoothing and low-density values (over smooth). In contrast, a small bandwidth value will result in less smoothing (under smooth), producing a map that depicts local variations in point densities (Chainey, Reid & Stuart 2002).}

In our research, different bandwidth values are examined to find the most appropriate bandwidth value. Results of KDE analysis with different bandwidth values showed that KDE related to 600 m bandwidth value is not very small and smooth. Therefore, 600 m is selected as KDE bandwidth value for hotspot analysis. The density of vehicle-pedestrian crashes is displayed by continuous surfaces in a raster map. In this map, which shows the results of KDE, lighter shades represent locations with lower crash density, while darker shades indicate areas with higher density of crashes.

4. DATA DESCRIPTION

In this research, data for all vehicle-pedestrian crashes on public roadways in the Melbourne metropolitan area from 2004 to 2013 were extracted from the Victoria interactive crash statistics application (CrashStats). In Victoria, only crashes resulting in deaths or injuries are legally required to be reported to the police. A total of 12 279 vehicle-pedestrian crashes were recorded in the Melbourne metropolitan area over the 10-years period. After removing the incomplete and invalid data, 9872 pedestrian crashes were used to investigate the variables contributing to the vehicle-pedestrian crash severity. The dataset includes personal characteristics, vehicle characteristics, road and environment conditions, and temporal characteristics.

In addition, this research considered the influence of POIs on vehicle-pedestrian crashes to explore the impacts of different types of POIs on crash severity. These points are extracted from the Australian Urban Research Infrastructure Network (AURIN). AURIN is the largest single resource for accessing diverse types and sources of data, spanning the physical, social, economic and ecological aspects of Australian cities, towns and communities.

In CrashStats, crash severity is divided into four levels: fatal, serious injury, other injury, and non-injury crashes. A fatal crash refers to a crash that has at least one death within 30 days while in a serious injury crash, at least one person is sent to the hospital. Fatal and serious injuries comprised 49.2% of total vehicle-pedestrian crashes (3.2% for fatal and 46% for serious injury) while the remaining 50.8% are reported as other injuries. Table 1 shows a summary of the classified data used in this research for different levels of vehicle-pedestrian crash severity. In this research, 10 explanatory variables are classified into 4 categories describing the accident time, personal characteristics, road variables, and location type for each recorded crash.

5. RESULTS AND DISCUSSION

DT model

Figure 5 shows the results of DT model, with nine terminal nodes. This means that it was not possible to define new rules to split these nodes and find more homogenous nodes. According to this figure, the time of crash, speed limit, pedestrian age, and POI are the primary splitters in this model. This implies that these variables are critical in classifying the severity of vehicle-pedestrian crashes in the Melbourne metropolitan area.

Interestingly, the initial split at first node is based on the time of crash variable. This reveals that the most important influencing variable to classify the severity of vehicle-pedestrian crashes is the time of crash. Furthermore, Figure 5 shows that 25% of the pedestrian crashes occurred between 7:00 p.m. and 6:00 a.m. (night time). However, a considerable percentage (about 60%) of these accidents was fatal or serious injury. This indicates that the risk of severe crashes occurring during this period is higher than other times.

Part of the reason for the above result is provided by the next split node. For vehicle-pedestrian crashes which occurred during the night, speed is the second splitter variable. According to Figure 5, pedestrians are more likely to be killed or harmed in a crash during night and on roads with a speed limit 80 km/h or higher. This figure reveals that pedestrian crashes during the night and on roads with a speed limit of more than 80 km/h or higher are more likely to be fatal (19%) or injury crashes (62%). This is a significant result that shows that both time and speed limit are significant risk factors in vehicle-pedestrian crashes.

On the other hand, for vehicle-pedestrian crashes which occurred during the day, the pedestrian’s age is the most significant variable to classify.
the severity of the crash. Figure 5 shows that for day time crashes, the severity of traffic crash for ageing pedestrians (>65) is higher than all other age groups. Also, for this age group, POI is a significant classifier variable. This finding reveals that for day time crashes, aged pedestrians are more likely to be injured or killed (about 60%) and for this group of pedestrians, the type of POI is an important factor. However, for other age groups, speed and POI are selected to classify the data for following branches. These results indicate that road safety strategies such as reducing travel speed close to POIs (e.g. government offices, playgrounds and health centres) can improve road safety and decrease the

<table>
<thead>
<tr>
<th>Category</th>
<th>Variable</th>
<th>Injury severity (%)</th>
<th>Fatal</th>
<th>Serious injury</th>
<th>Other injury</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temporal Variables</td>
<td>Day of week</td>
<td>Sunday</td>
<td>0.3</td>
<td>4.8</td>
<td>4.7</td>
<td>9.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Monday</td>
<td>0.4</td>
<td>6.3</td>
<td>7.2</td>
<td>13.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Tuesday</td>
<td>0.4</td>
<td>6.6</td>
<td>8.7</td>
<td>15.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wednesday</td>
<td>0.4</td>
<td>7.0</td>
<td>7.5</td>
<td>15.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thursday</td>
<td>0.6</td>
<td>7.2</td>
<td>8.3</td>
<td>16.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Friday</td>
<td>0.5</td>
<td>8.0</td>
<td>8.6</td>
<td>17.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Saturday</td>
<td>0.5</td>
<td>6.1</td>
<td>5.7</td>
<td>12.3</td>
</tr>
<tr>
<td></td>
<td>Month</td>
<td>January</td>
<td>0.2</td>
<td>2.9</td>
<td>2.9</td>
<td>6.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>February</td>
<td>0.2</td>
<td>3.6</td>
<td>3.9</td>
<td>7.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>March</td>
<td>0.3</td>
<td>3.8</td>
<td>4.5</td>
<td>8.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>April</td>
<td>0.3</td>
<td>3.7</td>
<td>4.1</td>
<td>8.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>May</td>
<td>0.3</td>
<td>4.2</td>
<td>4.9</td>
<td>9.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>June</td>
<td>0.3</td>
<td>4.5</td>
<td>4.7</td>
<td>9.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>July</td>
<td>0.4</td>
<td>4.5</td>
<td>4.7</td>
<td>9.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>August</td>
<td>0.3</td>
<td>4.4</td>
<td>4.4</td>
<td>9.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>September</td>
<td>0.2</td>
<td>3.6</td>
<td>4.5</td>
<td>8.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>October</td>
<td>0.3</td>
<td>3.7</td>
<td>4.3</td>
<td>8.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>November</td>
<td>0.2</td>
<td>3.4</td>
<td>3.9</td>
<td>7.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>December</td>
<td>0.3</td>
<td>3.6</td>
<td>3.9</td>
<td>7.8</td>
</tr>
<tr>
<td>Temporal Variables</td>
<td>Time</td>
<td>0-6 (night)</td>
<td>0.6</td>
<td>5.0</td>
<td>3.5</td>
<td>9.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7-9 (First Peak)</td>
<td>0.4</td>
<td>6.3</td>
<td>8.9</td>
<td>15.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10-15 (off-Peak)</td>
<td>0.8</td>
<td>15.6</td>
<td>20.1</td>
<td>36.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16-18 (Second Peak)</td>
<td>0.6</td>
<td>10.4</td>
<td>11.8</td>
<td>22.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19-23 (night)</td>
<td>0.8</td>
<td>8.6</td>
<td>6.6</td>
<td>16.0</td>
</tr>
<tr>
<td>Personal Characteristics</td>
<td>Pedestrian age</td>
<td>≤ 18</td>
<td>0.3</td>
<td>8.9</td>
<td>10.4</td>
<td>19.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19-24</td>
<td>0.3</td>
<td>6.2</td>
<td>7.3</td>
<td>13.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25-34</td>
<td>0.4</td>
<td>7.7</td>
<td>8.9</td>
<td>17.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>35-44</td>
<td>0.4</td>
<td>4.8</td>
<td>6.5</td>
<td>11.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>45-54</td>
<td>0.3</td>
<td>4.0</td>
<td>5.8</td>
<td>10.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>55-64</td>
<td>0.3</td>
<td>4.1</td>
<td>4.4</td>
<td>8.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>65-74</td>
<td>0.4</td>
<td>4.1</td>
<td>3.6</td>
<td>8.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>75+</td>
<td>1.0</td>
<td>6.0</td>
<td>3.9</td>
<td>10.9</td>
</tr>
<tr>
<td></td>
<td>Pedestrian gender</td>
<td>Male</td>
<td>2.0</td>
<td>24.8</td>
<td>24.8</td>
<td>51.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Female</td>
<td>1.2</td>
<td>20.8</td>
<td>25.7</td>
<td>47.7</td>
</tr>
</tbody>
</table>

(Table 1 continued next page)
Spatial and temporal distribution of pedestrian crashes in Melbourne metropolitan area

### Table: Injury severity (%)

<table>
<thead>
<tr>
<th>Category</th>
<th>Variable</th>
<th>Fatal</th>
<th>Serious injury</th>
<th>Other injury</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal Characteristics</td>
<td>Driver age ≤ 18</td>
<td>0.1</td>
<td>1.5</td>
<td>1.3</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>19-24</td>
<td>0.6</td>
<td>7.8</td>
<td>7.1</td>
<td>15.4</td>
</tr>
<tr>
<td></td>
<td>25-34</td>
<td>0.7</td>
<td>9.9</td>
<td>11.3</td>
<td>22.0</td>
</tr>
<tr>
<td></td>
<td>35-44</td>
<td>0.6</td>
<td>9.5</td>
<td>11.0</td>
<td>21.1</td>
</tr>
<tr>
<td></td>
<td>45-54</td>
<td>0.6</td>
<td>7.6</td>
<td>8.9</td>
<td>17.1</td>
</tr>
<tr>
<td></td>
<td>55-64</td>
<td>0.5</td>
<td>5.1</td>
<td>6.4</td>
<td>12.0</td>
</tr>
<tr>
<td></td>
<td>65-74</td>
<td>0.1</td>
<td>2.7</td>
<td>2.8</td>
<td>5.6</td>
</tr>
<tr>
<td></td>
<td>75+</td>
<td>0.1</td>
<td>1.8</td>
<td>2.1</td>
<td>4.0</td>
</tr>
<tr>
<td>Driver gender</td>
<td>Male</td>
<td>2.4</td>
<td>30.7</td>
<td>32.6</td>
<td>65.7</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>0.9</td>
<td>15.1</td>
<td>18.1</td>
<td>34.1</td>
</tr>
<tr>
<td>Road characteristics</td>
<td>Speed limit ≤ 50</td>
<td>0.7</td>
<td>17.6</td>
<td>23.3</td>
<td>41.7</td>
</tr>
<tr>
<td></td>
<td>60-70</td>
<td>1.7</td>
<td>23.3</td>
<td>22.5</td>
<td>47.5</td>
</tr>
<tr>
<td></td>
<td>80-120</td>
<td>0.8</td>
<td>3.3</td>
<td>1.8</td>
<td>5.9</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>0.0</td>
<td>1.7</td>
<td>3.2</td>
<td>4.9</td>
</tr>
<tr>
<td>Node type</td>
<td>Intersection</td>
<td>1.4</td>
<td>23.6</td>
<td>28.5</td>
<td>53.5</td>
</tr>
<tr>
<td></td>
<td>Mid-block</td>
<td>1.8</td>
<td>22.2</td>
<td>22.2</td>
<td>46.2</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>0.0</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
</tr>
<tr>
<td>Location</td>
<td>Educational</td>
<td>1.0</td>
<td>14.9</td>
<td>16.1</td>
<td>32.0</td>
</tr>
<tr>
<td></td>
<td>Health services</td>
<td>0.5</td>
<td>8.9</td>
<td>9.7</td>
<td>19.3</td>
</tr>
<tr>
<td></td>
<td>Playground</td>
<td>0.4</td>
<td>5.0</td>
<td>4.9</td>
<td>10.3</td>
</tr>
<tr>
<td></td>
<td>Accommodation</td>
<td>0.2</td>
<td>4.1</td>
<td>4.6</td>
<td>8.9</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>1.1</td>
<td>13.1</td>
<td>15.5</td>
<td>29.7</td>
</tr>
</tbody>
</table>

### Figure 5
The Decision Tree for pedestrian crashes
severity of vehicle-pedestrian crashes. Furthermore, Table 2 shows the relative importance of different variables influencing vehicle-pedestrian crashes in Melbourne metropolitan area. According to this table, pedestrian age, POI, time of crash, and speed are the four most significant variables. These variables have the greatest (75%–100%) impact on vehicle-pedestrian crash severity.

In summary, the DT model (Figure 5 and Table 2) confirms that temporal and spatial variables (time and POI) have a significant impact on vehicle-pedestrian crash severity. Therefore, it is important to consider temporal and spatial characteristics of pedestrian crashes in crash or injury severity analysis to improve our understanding of the road safety issues for this group of road users.

Spatial autocorrelation

Spatial autocorrelation is applied to show the dependency between the location and time of vehicle-pedestrian crashes. Moran’s I is calculated multiple times with different increasing distance thresholds. Figure 6 shows the results of Moran’s I for vehicle-pedestrian crashes during day and night. This figure indicates that there is a positive correlation between time and location of vehicle-pedestrian crashes for day and night crashes. In addition, according to Figure 6, Moran’s I for 100 m and 150 m distance thresholds have the highest magnitude for day and night crashes, respectively. In addition, Figure 6 shows that the maximum Z-score is 3.0 for day and 5.6 for night crashes. In addition, Moran’s I relative to maximum Z-score is 0.03 for vehicle-pedestrian crashes during the day and 0.06 for night accidents.

These results for Z-score and Moran’s I illustrate that the dependency between crash time and location is more significant during the night (7:00 p.m. to 6:00 a.m.) than during the day. Furthermore, the Z-scores of 5.6 and 3.0 (p=0.00) for 150 m and 100 m radii indicate there is less than 1% chance that this clustered pattern could be the result of a random choice. These findings reveal that there is a significant relationship between the time and location of vehicle-pedestrian crashes. This relationship is more significant for crashes during the night.

Temporal analysis

In this research, spider plots are applied to analyse the temporal distribution of vehicle-pedestrian crashes in the Melbourne metropolitan area. Figure 7 illustrates how the vehicle-pedestrian crashes are distributed during different times of the day. This figure shows two peaks at 3:00 p.m. and 8:00 a.m. and a slight peak at 5:00 p.m. According to Victoria traffic monitor reports on 2013, there are two traffic peak period times (7:30 a.m. to 9:00 a.m. and 4:30 p.m. to 6:00 p.m.) and one off-peak period (10:00 a.m. to 3:00 p.m.) in the road network of Melbourne metropolitan area (VicRoads 2014). It should be noted that according to Figure 7, most
Spatial and temporal distribution of pedestrian crashes in Melbourne metropolitan area

pedestrian crashes occur during the off-peak traffic time (10:00 a.m. to 3:00 p.m.).

Interactive DT models provide the opportunity to classify the dataset according to a particular variable (e.g. speed, age, POI). Applying interactive DT for off-peak crashes shows that age of pedestrians is a significant variable to classify these crashes (Figure 8). This figure shows that for pedestrians older than 65 years of age who are involved in crashes, POI is an important variable. More than 70% of crashes for this category occurred around other POI, including aged cares, churches and offices. For the other age groups, speed and POI are the most important variables to classify crashes during off-peak traffic. For these pedestrians, education centres, general offices and local government offices are the important POIs.

Figure 9 illustrates the temporal patterns of vehicle-pedestrian crashes for different age groups. According to this figure, different age groups have different crash patterns during the day and night. School age groups (0–18 years old) have two peaks at 8:00 a.m. and 3:00 p.m. when they go to school and return home. However, for pedestrians between 19 and 35 years of age, this figure shows that most of the crashes occur after 5:00 p.m., when most of them finished work. For pedestrians older than 65 years of age, 9:00 a.m. to 12:00 p.m. and 3:00 p.m. are high risk times. These findings reveal that off-peak traffic time is more important in vehicle-pedestrian crashes. Furthermore, these results indicate that different safety strategies must be considered for different age groups. For pedestrians at school age, improving safety around schools and increasing the traffic safety awareness of children (e.g. safety education, walking to or from schools under supervision, etc.) might be helpful in decreasing the risk of crash for this age group. However, for pedestrians between 19 and 35 years of age, safety strategies targeted at evening attractions, such as restaurant and bars, must be considered.

Figure 10 illustrates the distribution of vehicle-pedestrian crashes over different times of the day and days of the week. It is clear that the patterns of vehicle-pedestrian crashes are different during the 24 hours on weekdays and weekends. During the weekdays, most crashes occurred between 8:00 a.m. and 6:00 p.m., while many weekend crashes occurred late at night and early in the morning.
In addition, pedestrian crash patterns at different times on Friday are different from other weekdays. Figure 10 shows that there are more crashes between 8:00 p.m. and 11:00 p.m. on Friday night. Interestingly, there are different crash peaks for Saturdays and Sundays between 11:00 p.m. and 1:00 a.m.. One possible explanation for these results is that more people go out on Friday nights than Sunday nights. These results show that it is important to consider targeted pedestrian safety strategies for these periods of time, such as increasing the visibility of pedestrians at night by improving street lighting or encouraging pedestrians to carry a flashlight or wear retro-reflective clothing. Some strategies to manage alcohol consumption and drink walking should be considered, especially on Friday and Saturday nights.

Table 3 shows the distribution of crashes for each severity level across different days of the week. According to this table, vehicle-pedestrian crashes occur most frequently on Friday (17.1%), followed by Thursday (16.1%). Moreover, Thursdays with 18.4% and Saturdays with 16.9% have the highest percentages of fatal crashes. Also, Table 3 shows that most of the serious injury crashes occur on Thursdays (15.6%) and Fridays (17.4%).

Figure 11 shows the distribution of crashes with different severity levels for each day of the week. In addition, Table 3 and Figure 11 reveal that only about 22% of pedestrian crashes occur during the weekend. However, more than 50% of pedestrian crashes which occur on Saturdays and Sundays are serious or fatal injury crashes. Figure 11 shows that the percentage of fatal and severe injury crashes reaches the highest value on Saturday and Sunday (54% and 52%, respectively), followed by Friday.
Spatial and temporal distribution of pedestrian crashes in Melbourne metropolitan area

(50%). Traffic is usually lower on weekends and drivers tend to drive faster than on other days. This reason and the higher chance of pedestrian or driver intoxication are factors contributing to the higher severity of vehicle-pedestrian crashes at weekends. Therefore, it is essential to consider appropriate countermeasure, such as increasing drink-driving and drink-walking enforcement, for improving pedestrian safety during these days.

### Table 3
Distribution of crashes for each severity level across different days

<table>
<thead>
<tr>
<th>Severity</th>
<th>Sunday</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
<th>Saturday</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fatal</td>
<td>9.7%</td>
<td>13.8%</td>
<td>12.5%</td>
<td>13.1%</td>
<td>18.4%</td>
<td>15.6%</td>
<td>16.9%</td>
</tr>
<tr>
<td>Serious injury</td>
<td>10.4%</td>
<td>13.6%</td>
<td>14.3%</td>
<td>15.3%</td>
<td>15.6%</td>
<td>17.4%</td>
<td>13.2%</td>
</tr>
<tr>
<td>Other injury</td>
<td>9.2%</td>
<td>14.2%</td>
<td>17.2%</td>
<td>14.9%</td>
<td>16.4%</td>
<td>17.0%</td>
<td>11.2%</td>
</tr>
<tr>
<td>Total</td>
<td>9.8%</td>
<td>13.9%</td>
<td>15.7%</td>
<td>15.0%</td>
<td>16.1%</td>
<td>17.1%</td>
<td>12.3%</td>
</tr>
</tbody>
</table>

Spatial analysis

*Figure 12* illustrates the vehicle-pedestrian crash distribution throughout the Melbourne metropolitan area over different time periods. These figures show that vehicle-pedestrian crash patterns are different during day and night. According to these figures, vehicle-pedestrian crashes are not significant in some suburbs such as Mornington, Keysborough and Ringwood during the day. However, the numbers of crashes are considerable for these suburbs during the night.
KDE is applied to analyse the spatial distribution of vehicle-pedestrian crashes and identify locations with high risk of vehicle-pedestrian crashes in different periods of time. Figure 13 shows the spatial distribution of pedestrian crashes (a) 7:00 a.m. to 9:00 a.m., (b) 10:00 a.m. to 3:00 p.m., (c) 4:00 p.m. to 6:00 p.m., (d) 7:00 p.m. to 11:00 p.m., and (e) 00:00 to 6:00 a.m. in a magnified map of the Melbourne metropolitan area. These figures show that the distribution of vehicle-pedestrian crashes in Melbourne varies according to the time of day. In general, vehicle-pedestrian crashes are highly concentrated in the Melbourne central suburbs. However, these tend to relatively more disperse during the day but comparatively more concentrated in few areas around south-east suburbs during the night.

As shown by spider graph in Figure 13(f), vehicle-pedestrian crashes are most frequently reported to occur 8–9 a.m. and 3–6 p.m. when people who work in Central Business District (CBD) commute to work or leave their work places. The existence of many offices, shopping centres and two Universities in Melbourne CBD provide many origins and destinations for pedestrian trips. These POIs increase the number of trips and consequently the risks of vehicle-pedestrian crashes in these areas (see Figures 13a, 13b, and 13c). However, Figures 13d and 13e reveal that between 7:00 and 11:00 p.m. and 00:00 and 6:00 a.m., there are several high risk areas for pedestrians, including Melbourne CBD, St. Kilda, Windsor and Prahran.

As mentioned before, alcohol consumption might be one of the important factors contributing vehicle-pedestrian crashes in these periods of time. Figure 14a shows the distribution of bars, restaurants and clubs in Melbourne Metropolitan area. As Figure 14a presents, there are many bars and restaurants in CBD. In addition, St. Kilda beach and bars around this beach are the destination of many pedestrians during weekend evenings. Furthermore, as Figures 12d, 12e, 14b and 14c show, Chapel Street is another area with high risk of vehicle-pedestrian crashes during the night. Figure 14a shows that there are many bars, clubs and restaurant in this street as well. Interestingly, comparing these figures shows that most areas with high density of vehicle-pedestrian crashes are also areas with a high number of bars and restaurants. Therefore, intoxication may be the main reason for vehicle-pedestrian crashes in these areas.

6. CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS
The main aim of this study was to examine the spatial and temporal distribution of vehicle-pedestrian crashes in Melbourne metropolitan area and to identify areas and times with high crash risk for these vulnerable road users. DT model was applied with 10 explanatory variables to explore their influence on vehicle-pedestrian crash severity. This model revealed that the time of crash is the most significant variable influencing the classification of vehicle-pedestrian crash severity. It also showed that speed limit and POIs are other
important variables for vehicle-pedestrian crash severity.

In addition, the analyses of spatial autocorrelation showed that there is a significant dependency between time and location of vehicle-pedestrian crashes. This dependency between time and location of crashes is greater during the night. Therefore, spider plots and KDE were applied to explore the temporal and spatial distribution of vehicle-pedestrian crashes in the Melbourne metropolitan area. Temporal analysis indicated that most crashes occur between 8:00 a.m. and 6:00 p.m. on weekdays. However, the frequency and severity of pedestrian crashes are higher during night time on weekends. About 22% of vehicle-pedestrian crashes occur on weekends (Saturdays and Sundays) and half of them are fatal or serious injury crashes.

Moreover, KDE analysis identified the areas with high risk of vehicle-pedestrian crashes in Melbourne metropolitan area for different time periods. Results of KDE identified that the risk of vehicle-pedestrian crashes is significant around the Melbourne CBD during the day. However, there are different vehicle-pedestrian crash patterns during the night and on weekends. The existence of bars, clubs and restaurants around these areas might increase pedestrian and vehicular traffic in these areas. Also, alcohol consumption by drivers or pedestrians might be the other important reason for vehicle-pedestrian crashes at nights during weekends in these areas. Evidence-based road safety strategies, such as drinking-driving and drink-walking enforcement, targeting these times and locations would be required to improve pedestrian safety.

7. REFERENCES


Blazquez, CA & Celis, MS 2013, 'A spatial and temporal analysis of child pedestrian crashes in Santiago, Chile', Accident Analysis & Prevention, vol. 50, no. 4, pp. 465-72.


Spatial and temporal distribution of pedestrian crashes in Melbourne metropolitan area

Alireza Toran Pour
Alireza Toran Pour received the B.Eng. in Civil Engineering from Bu-Ali Sina University, Iran. Also he earned M.Sc. degree in transportation engineering from Shahid Chamran University of Ahwaz, Iran. Ali is currently a PhD student in School of Engineering at RMIT University, Melbourne. His research interests are in road safety with emphasis on the applications of Geographic Information Systems, data analysis, and Intelligent Transpiration Systems.

Sara Moridpour
Sara Moridpour holds a Bachelor of Civil Engineering and Master’s degree in Transportation Planning and Engineering from Sharif University of Technology, Iran. She received her PhD degree from Monash University. She has 10 years of work and research experience in the field of traffic and transport. Her main research interests include driving behaviour modelling and analysis, micro simulation, transport network modelling and optimisation, transport infrastructure asset management and traffic safety. She has been working in the School of Civil, Environmental and Chemical Engineering, RMIT University, since 2010.

Richard Tay
Richard Tay is a VC Senior Research Fellow at RMIT University. His expertise is in applying multidisciplinary theory and evidence based approaches to improving road safety, and other transportation or safety related issues. He serves as Editor-in-Chief of International Journal of Transportation; Associate Editor of Journal of Advanced Transportation, Journal of Transportation Safety and Security, Asia-Pacific Journal of Transport, and Urban, Planning & Transport Research; Guest Editor of Transportation Research Part E, Open Transportation Journal, KSCE Journal of Civil Engineering and Applied Health Economics and Health Policy; and Editorial Board Member of Analytic Methods in Accident Research, Accident Analysis and Prevention, Journal of the Australian College of Road Safety, Open Transportation Journal, Canadian Journal of Transportation, and International Journal of Sustainable Transportation.

Abbas Rajabifard
Professor Abbas Rajabifard is Director of the Centre for Spatial Data Infrastructures and Land Administration at the Department of Infrastructure Engineering, the University of Melbourne. He is the immediate Past President of the GSDI Association, and was Vice Chair of Working Group 3 of the United Nations supported Permanent Committee on GIS Infrastructure for Asia and the Pacific (PCGIAP), a member of ICA-Spatial Data Standard Commission, and a member of the Victorian Spatial Council. He has been an Executive Board member and national representative to the PCGIAP (1994-1998), member of International Steering Committee for Global Mapping Project (1997-2001) and a member of the UN-ESCAP Group of Experts to develop Guidelines on GIS Standardisation for Asia-Pacific (1995).

CONTACT
Alireza Toran Pour
PhD student
School of Engineering
RMIT University
GPO Box 2476
Melbourne, 3001, Australia
Email: <alireza.toranpour@rmit.edu.au>