Macro Level Model Development for Safety Assessment of Road Network Structures

Xuesong Wang, PhD
Associate Professor
School of Transportation Engineering
Tongji University
4800 Cao’an Road, Jiading District
Shanghai, 201804, China
Phone: +86-21-69583946
Email: wangxs@tongji.edu.cn
(Corresponding author)

Yu Jin
Shanghai Urban Planning and Design Research Institute
Shanghai, 200040, China
Phone: +86-21-62473288-0421
Fax: +86-21-62477739
Email: jinyu@supdri.com

Mohamed Abdel-Aty, PhD
Professor
Department of Civil, Environmental & Construction Engineering
University of Central Florida
Orlando, FL 32816-2450, U.S.A.
Phone: (407) 823-5657
Email: m.aty@ucf.edu

Paul J. Tremont, PhD
Visiting Professor
School of Transportation Engineering
Tongji University
4800 Cao’an Road, Jiading District
Shanghai, 201804, China
Phone: +86-21-69583946
Email: pjtremont@hotmail.com

Word Count: 5497
Tables and Figures (10 at 250 words each): 2000
Total Count: 7497

November 2011

Prepared for Presentation at the TRB Annual Meeting and Possible Publication in the Journal of the Transportation Research Board
Traffic safety is beginning to receive increasing attention at the stage of transportation planning. Although road network features are an essential aspect of transportation planning, studies of the safety effects of network patterns still remain limited. In this study, macro level safety models were developed to explore the relationship between crash occurrence and several underlying variables including demographic, land use, and road network variables. Prior efforts to model network structures have been hampered by difficulties quantifying network properties. In this study, different indices (e.g., Meshedness Coefficient, Closeness Centrality) of network structures were developed to examine the network structure effects on zonal level safety. In many cases, a large percentage of the crash locations (especially for arterial crashes) were not related to the Traffic Analysis Zones (TAZs) where drivers lived. To ensure proper linkage of crashes and zonal level features, we propose to model crashes of each TAZ for non-state maintained arterials (i.e., off-system), and state-maintained arterials (i.e., on-system), separately. We developed several Conditional Autoregressive (CAR) Bayesian models that incorporated the spatial correlation of nearby zones. Estimation results showed that crashes occurring on non-state maintained roads correlate more closely to the zonal network structure, and the demographic characteristics inside the TAZ, compared to crashes occurring on state-maintained arterials that correlate more closely with the traffic and road features of the major roads. The categorical variable generated from the Meshedness Coefficient performed well in capturing the nature of network patterns as they relate to off-system road crashes. This study shows that both the roadway type and the structure of the road network should be considered when developing TAZ level safety models.

**Keywords:** Transportation safety planning, network structure, Traffic Analysis Zone (TAZ), modeling strategy, Bayesian estimation, Conditional Autoregressive (CAR) model
1. INTRODUCTION

Traffic crashes remain the leading cause of death for ages 4 to 34, and are one of the top 10 causes of death for all other ages in the U.S. (1). The economic and social costs of these crashes also place an unacceptably large burden on society. One approach to addressing this problem is to incorporate safety into the transportation planning process. This is a way to address the safety problem proactively. During the past few decades, various researchers have started to explore this area (2; 3; 4), and several studies have been undertaken to develop safety prediction models at the Traffic Analysis Zone (TAZ) level. The intent has been to investigate the relationship between TAZ level crash frequencies and TAZ level demographic, street network, land use, and traffic characteristics (4; 5; 6; 7). These TAZ based models provide a global approach that, together with traditional transportation planning, can help transportation planners and engineers assess the safety of alternative planning programs.

As an essential aspect of transportation planning, the roadway network pattern plays an important role in safety by affecting the characteristics of traffic flow and driving behavior. A few previous studies have analyzed the safety effect of roadway network structure. Sun and Lovegrove (8), and Rifaat and Tay (9) analyzed the effect of TAZ road network patterns on crash frequency and severity. In their studies, road networks were classified into several categories by visual inspection. This method has the disadvantage of being subjective and time consuming, and consequently cannot be widely applied. Recently, the contribution of physicists (10; 11; 12) to the development of quantitative indices of network structures has led to advances in the science of complex networks. These indices provide objective measures of network configurations, and can be rapidly computed for area roadway structures. As a result, safety assessments based on roadway structures in TAZs are now able to effectively support the planning process.

Previous network studies (8; 9) used total crashes occurring for each TAZ to establish the safety performance function relating crash frequencies to TAZ related variables under the assumption that crashes in each TAZ are primarily related to the TAZ characteristics. However, this assumption might be easily violated for crashes occurring along arterials since drivers involved in these crashes often do not live in the specific TAZ to which the crash is assigned. The objectives of this paper are: (1) to introduce indices that can be used to quantify the roadway network structure and thereby allow its level of safety to be estimated, (2) to investigate the modeling strategy at the macro level to accurately assess the safety of road network structures, and (3) to perform a preliminary safety assessment of different road network structures.
2. LITERATURE REVIEW

Macro Level Safety Modeling

Traffic Analysis Zone (TAZ) level safety analysis stems from early studies on the spatial distribution of crashes. In 1995, Levine et al. used motor vehicle crash data from Honolulu to examine crash spatial distribution patterns (13), and developed a spatial lag model to examine the relationship between motor vehicle crashes for census block groups and trip generating activities (14). Since then, many studies have been conducted to identify factors contributing to crashes at the macro level. More recently, several TAZ level safety prediction models were developed and applied to accommodate conventional transportation planning (4; 5; 6; 7).

The factors investigated in these models include four classes: social demographic factors, traffic characteristics, roadway factors, and land use features. For social demographic factors, total population, population density, household number, and employment status within the TAZ were usually considered (6; 7; 8; 15; 16). For traffic characteristics, vehicle kilometers (or miles) traveled (VKT, VMT) were usually identified as the most important variable, while average speed, and average V/C ratio (V represent volume, and C represent capacity) were sometimes found to be significant (6; 8; 16). For roadway features, the length and density for a specific type of road or intersection, and the average curvature of roadways were identified to be significant (6; 7; 8; 15; 16). For land use, the safety effects of different land use types (residential, commercial, comprehensive development, container storage, and county park) were investigated and found to be significant (15; 17).

In the late 1990s, Dutch researchers found that their limited access Sustainable Road Safety (SRS) street patterns can reduce the risk of both total and severe collisions (18; 19). As different roadway patterns usually result in different safety levels, and the same road density may present different patterns, the importance of examining the effect of street pattern on safety became clear. In 2006, Lovegrove and Sayed used macro level safety prediction models to identify the safest road network patterns. They found two conceptual networks, the 3-way offset, and the modified Dutch SRS network, were safer than conventional grid and cul-de-sac patterns (20). Sun and Lovegrove examined the safety effects of street patterns at the community level. They found the fused grid network and the 3-way offset pattern to be safer than the grid network, cul-de-sac pattern, and Dutch SRS patterns (8). Rifaat and Tay investigated the effect of street patterns on collision severity and found that the loops and lollipops street patterns are safer than grid street patterns (9).

The main limitation of the above studies is that street patterns were classified using visual inspection because no objective quantifiable index was available. However, the visual inspection method is time consuming and subjective, and this limits its usefulness. This lack of an appropriate index to classify street patterns has hampered the usefulness of network analytical approaches to
safety. Development of such an index would improve the estimation of safety as it relates to street patterns, and would also simplify the application of various models to different street patterns.

**Network Structure**

In recent years various studies were conducted to explore the topological properties of complex networks, and eventually indices were developed to analyze their structural properties. The concept of centrality, a crucial structural attribute of a network, was first developed by Bavelas in the late 1940s. Since then, several centrality measures, including Degree Centrality (12), Betweenness Centrality (12; 22), Closeness Centrality (12), Straightness Centrality (23), and Information Centrality (24), have been proposed. Each of these indices depicts the centrality of a network from a different perspective. Clustering, another important property of networks, quantifies the structure of network cycles (25). Two indices, the Clustering Coefficient (26) and the \( k \) Clustering Coefficient (27), are often used for depicting network structures. In 2004, Buhl et al., introduced the Meshedness Coefficient to measure network cycles in a more general way (10). In recent years, both the Meshedness Coefficient and Centrality measures of a network have been introduced into urban street network analyses. In 2006, Crucitti et al. analyzed the spatial and statistical distribution of Centrality based on a 1-square-mile street network sample in 18 different cities (28). Using the same data, Cardillo et al. investigated the basic properties of street networks among different street patterns (11). These network measures can mirror the roadway network structures to some extent. However, whether the indices they used can be used to link the effects of network structure to safety needs further investigation.

**Macro-level Modeling Strategies**

Some studies on macro-level safety prediction models were focused on the ward level (similar to precinct) (29; 30) and county level (5; 31; 32; 33). Being government jurisdictions, designing models at these levels allows the use of readily available analysis data. However, for the conventionally applied four-step travel demand forecasting method, the study area should be divided into Traffic Analysis Zones (TAZs), because these are required to be homogeneous in land use and in travel patterns. Therefore, to accommodate the transportation planning process, it is advantageous to use the TAZ as the analysis unit, rather than a municipal jurisdiction.

For models at the TAZ level, the total crashes of each TAZ are typically used as the dependent variable, and then are modeled as a function of a set of explanatory variables (4; 6; 7; 34). However, two concerns still need to be addressed. First, major arterials usually serve as TAZ boundaries, with the result that crashes occurring on these boundaries may not have a strong relationship with the street network pattern of the TAZ. Second, crashes occurring on arterial roads serving long distance travel may not have a strong relationship to the demographic or land use characteristics of the corresponding TAZ. In this paper, the modeling strategy used will address these concerns in an effort to more accurately model crashes at the macro level.
Statistical Analysis Approach

Researchers traditionally used Poisson or Negative Binomial (NB) regression models to analyze the relationship between crash frequencies and explanatory variables at the TAZ level (6; 7; 8; 16). Both Poisson and NB models are special cases of the Generalized Linear Model (GLM). Under the GLM framework, the selected TAZs are assumed to be spatially independent across TAZs in the study area. However, those assumptions are sometimes violated because the TAZs near to each other are similar in nature as well as in safety performance. To address this spatial correlation problem, a Geographically Weighted Poisson Regression (GWPR) approach was developed by Fotheringham (35). This approach allows coefficient estimates to vary across spaces. Hadayeghi et al. used the GWPR to model TAZ level crash frequencies and compared it with the traditional GLM. His study showed that the GWPR model, by capturing the spatial variation, generally outperforms the GLM (36).

To address the spatial dependency problem, Bayesian models that can account for the spatial correlation were also applied in many traffic safety studies. Aguero-Valverde and Jovanis developed full Bayesian hierarchical models to predict county level crash frequencies (32). Hadayeghi et al. used the full Bayesian semi-parametric additive technique to accommodate the spatial correlation problem (37). Guo et al. developed conditional autoregressive (CAR) Bayesian count data models to account for the spatial correlation among signalized intersections along corridors (38). The flexibility of the Bayesian method allows for modeling data with spatially correlated complex data structures. The Bayesian statistical methods’ ability to incorporate spatial correlations is superior to the ordinary GLM, and in part justifies for its increased use in modeling.

3. DATA PREPARATION

Orange County is located in the Central Florida area in the U.S. For this study, the required data are the TAZ boundaries, the roadway, and crash data. The TAZ boundary shape file was extracted from the Central Florida Regional Planning Model (CFRPM). The base year of this model is 2005, and its horizon year is 2035. For this planning model, there are 662 TAZs. The crash data, the area land use, the state-maintained (on-system), and the non-state maintained (off-system), road data were provided by the Florida Department of Transportation (FDOT). These data were available in Geographic Information System (GIS) format. The on-system road data provided by FDOT includes intersection location, signal location, and Average Annual Daily Traffic (AADT) data. Both on and off-system data provided by FDOT do not contain community extent roads, so it was necessary to use the US Census Bureau’s TIGER file to extract the complete road information.
**TAZ and Data Matching Problems**

The first step is to allocate roads and crashes into their corresponding TAZs. The TAZ boundaries from the CFRPM and the state-maintained arterial roads from FDOT do not match perfectly; this can be seen below in Figure 1. These mismatches would lead to some incorrect allocations of collisions, signals, intersections, and roads, and would result in inaccurate calculation of road lengths in the TAZs. As stated earlier, these GIS files are from different sources. This leads to an imperfect data match - a common problem when assembling data from different sources. Most previous studies in the field did not address this problem when preparing the data. However, Jonsson et al. did consider the TAZ boundary and road matching problem. They tested a set of different values to draw buffers for the TAZ polygons when connecting the TAZs with the roadways (40). Once that was done, they found the best way to solve the matching problem was to apply a 200-foot buffer outside of each TAZ.

In this study, crashes were geocoded on both state-maintained and non-state-maintained roads. To ensure accurate assignment of the crash to the correct TAZs, each of the TAZ boundaries was checked, and for those TAZs with a mismatch problem, the TAZs’ boundaries were manually drawn based on the route of the state-maintained and non-state-maintained roads. That is, we used the road to construct the new TAZ boundaries.
FIGURE 1  Area TAZ boundaries and data match problem
**Demographic Data and Land Use**

Demographic data includes population, age, and household size. These data were extracted from the 2000 U.S. Census data at the block level for Orange County. Then the population for the study year of each block was adjusted upward from the year 2000 by multiplying a growth factor over 6 years. The growth factor was calculated as the ratio of the entire county’s population of two years using data from the Orange County Planning Division (41). The demographic data of 14,582 census blocks were then aggregated into 662 TAZs through the spatial join function provided in ArcMap (42). The land use categories in Orange County include urban, suburban, and rural. By using the spatial join function, the land use feature for each TAZ was defined using the urban land use layer provided by FDOT.

**Road Network Features**

The state-maintained (on-system) and non-state-maintained (off-system) roads consist of the major roads of the study area. This is supplied by the TIGER file to form the complete road network. Aggregation related features (e.g., total number of 3-leg/4-leg intersections, total number of arterials and freeways, total number of collectors) as well as road network structure related features were calculated for each TAZ.

For the quantitative features of the road network, the GIS Spatial Join function was used to extract signals, intersections, roads, and AADT data at the TAZ level. In this study, a method proposed by Sun and Lovegrove (8) was applied to treat roads on the TAZ boundary. Polyline or point features on the TAZ boundaries were shared by corresponding TAZs by pre-rating them with a weight equal to the reciprocal of related TAZs. Using this method, an intersection on the boundaries of three TAZs, for example, were assigned a weight of 0.333. The roads on the TAZ boundary were assigned by the number of adjacent TAZs.

For road network structures, street networks were represented by transferring segments into links, and intersections into nodes. In a TAZ of roads composed of K links and N nodes, the Meshedness Coefficient is calculated as:

\[
M = \frac{(K-N+1)}{(2N-5)}
\]

where M can vary from 0 to 1. If M equals 0, the graph refers to a tree structure. If M equals to 1, the graph is a complete planar graph. The Meshedness Coefficient measures the structure of circles in a graph. In graph theory, the term path refers to a sequence of vertices of a graph, and a circle is defined as a closed path. As illustrated in Figure 2, a-b-d-c is a path, and a-b-d-a is a circle.
The TIGER file contains all the roads in the Orange County including community-extent streets, and therefore was used for determining Closeness Centrality. Closeness Centrality measures the decentralization of points in the graph and its value grows bigger as the points become farther away from each other. Closeness Centrality is defined as:

$$C_i^c = \frac{(N - 1)}{\sum_{j \in G, j \neq i} d_{ij}}$$  \hspace{1cm} (2)

Given that a roadway network is represented by a unidirectional graph $G$, of $N$ nodes and $K$ links, the adjacent matrix was defined as a $N \times N$ matrix $A$, whose entry $a_{ij}$ was set to 1 if the node $i$ and $j$ is connected by a link. If not connected by a link it was set to 0. The Closeness Centrality was calculated as follows: split the roads where they intersect, and calculate the adjacent roadway network matrix for each TAZ using a customized functionality. UCINET is a social network analytical program that can calculate some network structural indices (43). The adjacent matrix for each TAZ was imported into UCINET, and was used to calculate the Closeness Centrality of each TAZ.

The initial indices for the road network structures are continuous variables. In order to evaluate whether the calculated values can represent the information of the road network structure, the initial indices were classified into several levels. Southworth and Ben-Joseph have categorized street patterns into five classes: grid, fragmented parallel, warped parallel, loops and lollipops, and lollipops on a stick (44). For the Orange County road network, this way of categorizing streets into several patterns was adopted with the modification that two special types, the mixed pattern and sparse pattern, were added. The sparse pattern is used for TAZs with sparse roadway layouts, and the mixed pattern was used where no single pattern predominated. The four typical street patterns for Orange County are shown in Figure 3.
Crash Data

In order to link crashes with each TAZ, crashes needed to be geocoded on the GIS road map. Guevara et al. showed that random errors are small when assigning point data (e.g., crashes) and polyline data (e.g., road segments) to TAZs, and therefore they assumed these errors would not impact the analysis significantly (7). Lovegrove also assumed that errors in the allocation of point and polyline data were reasonably acceptable and would not significantly influence the aggregation of data into the TAZ level (20). In 2010, Sun pre-rated the crashes on corners at 0.25 and on boundaries at 0.5 to minimize the bias in allocating crashes that intersect with TAZ boundaries (8). However, studies that focused on intersection crash spatial distributions found crashes at intersections were not equally distributed among four approaches (45; 46; 47). In this study, the vehicle traveling directions recorded in the crash data were used to decide the appropriate zone of crash occurrence.

A total of 17,249 crashes occurred in 2006 in Orange County. In this study, crashes occurring at intersections were first geocoded in the intersecting point of the GIS road network, while those on the road segments were geocoded on the center line, as shown below in Figure 4a. However, if the intersecting roads formed TAZ boundaries, knowing the accurate location for each crash became important (as shown below in Figure 4b). In this study, data on the direction of travel (north, east, south or west), and the first harmful event for each conflicting vehicle was...
derived from the Florida Department of Highway and Motor Vehicle and FDOT Crash Analysis Reporting System. These data were used to determine the accurate crash location so it could be assigned to the proper TAZ.

FIGURE 4 Allocation of intersection crashes to the proper TAZ

Similarly, for crashes occurring on road segments serving as the part of TAZ boundaries, the corresponding TAZs can be determined by using the side of road information recorded in the crash data file (e.g., east vs. west bound). Those crashes for which travel directions of involved
vehicles were unknown were weighted as weight=1/number of TAZs to which the collision corresponded. In Orange County, around 40% of crashes in 2006 occurred on the TAZ boundaries. This shows how important it is to ensure proper matching of the crash location to its TAZ.

**Data Assembly**

Demographic, land use, road network, and crash information were assembled for each TAZ. Descriptive statistics for these variables are summarized in Table 1. For road features, on and off-system, and total road information were extracted. For the Meshedness Coefficient, both its original form and categorized values were summarized.

**TABLE 1  Descriptive Statistics of Analysis Data at the TAZ Level**

<table>
<thead>
<tr>
<th>Category</th>
<th>Descriptions</th>
<th>Mean</th>
<th>S. D.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crash data</td>
<td>Total crashes</td>
<td>26.037</td>
<td>23.731</td>
<td>0.000</td>
<td>140.000</td>
</tr>
<tr>
<td></td>
<td>On-system crashes</td>
<td>14.446</td>
<td>17.247</td>
<td>0.000</td>
<td>116.000</td>
</tr>
<tr>
<td></td>
<td>Off-system crashes</td>
<td>11.591</td>
<td>13.297</td>
<td>0.000</td>
<td>72.000</td>
</tr>
<tr>
<td>Road features in size</td>
<td>Total arterial and expressway (km)</td>
<td>1.172</td>
<td>2.053</td>
<td>0.000</td>
<td>28.807</td>
</tr>
<tr>
<td></td>
<td>Total arterial and expressway as a portion of the total roads</td>
<td>0.108</td>
<td>0.115</td>
<td>0.000</td>
<td>0.644</td>
</tr>
<tr>
<td></td>
<td>Total collector (km)</td>
<td>1.149</td>
<td>1.672</td>
<td>0.000</td>
<td>13.819</td>
</tr>
<tr>
<td></td>
<td>Total collector as a portion of the total roads</td>
<td>0.108</td>
<td>0.119</td>
<td>0.000</td>
<td>0.969</td>
</tr>
<tr>
<td></td>
<td>Total state-maintained roads (km)</td>
<td>0.955</td>
<td>1.740</td>
<td>0.000</td>
<td>24.605</td>
</tr>
<tr>
<td></td>
<td>Total state-maintained roads density (km/km²)</td>
<td>0.753</td>
<td>0.842</td>
<td>0.000</td>
<td>7.543</td>
</tr>
<tr>
<td></td>
<td>Total state-maintained roads as a portion of total roads</td>
<td>0.099</td>
<td>0.104</td>
<td>0.000</td>
<td>0.643</td>
</tr>
<tr>
<td></td>
<td>Total roads (km)</td>
<td>13.082</td>
<td>15.789</td>
<td>0.000</td>
<td>229.985</td>
</tr>
<tr>
<td></td>
<td>Total off-system roads (km)</td>
<td>12.141</td>
<td>14.812</td>
<td>0.000</td>
<td>212.905</td>
</tr>
<tr>
<td></td>
<td>Total roads density (km/km²)</td>
<td>8.280</td>
<td>4.571</td>
<td>0.000</td>
<td>25.883</td>
</tr>
<tr>
<td></td>
<td>Number of 3-leg intersections</td>
<td>27.121</td>
<td>28.793</td>
<td>0.000</td>
<td>279.833</td>
</tr>
<tr>
<td></td>
<td>Number of 3-leg intersections per km² (/km²)</td>
<td>19.598</td>
<td>15.256</td>
<td>0.000</td>
<td>79.392</td>
</tr>
<tr>
<td></td>
<td>Number of 5-leg and 6-leg intersections per km² (/km²)</td>
<td>0.206</td>
<td>0.914</td>
<td>0.000</td>
<td>14.050</td>
</tr>
<tr>
<td></td>
<td>Number of signals on the state-maintained roads</td>
<td>0.643</td>
<td>0.841</td>
<td>0.000</td>
<td>5.080</td>
</tr>
<tr>
<td></td>
<td>Number of intersections on the state-maintained roads</td>
<td>3.933</td>
<td>4.510</td>
<td>0.000</td>
<td>35.750</td>
</tr>
<tr>
<td></td>
<td>Off-system intersections</td>
<td>31.236</td>
<td>36.076</td>
<td>0.000</td>
<td>454.673</td>
</tr>
<tr>
<td></td>
<td>Total intersections</td>
<td>35.146</td>
<td>37.047</td>
<td>0.000</td>
<td>476.333</td>
</tr>
<tr>
<td>Road network structure features</td>
<td>Meshedness Coefficient</td>
<td>0.123</td>
<td>0.085</td>
<td>-0.231</td>
<td>0.333</td>
</tr>
<tr>
<td></td>
<td>Categorized Meshedness Coefficient</td>
<td>2.480</td>
<td>0.706</td>
<td>1.000</td>
<td>4.000</td>
</tr>
<tr>
<td></td>
<td>Closeness Centrality</td>
<td>0.142</td>
<td>0.095</td>
<td>0.022</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Betweenness Centrality</td>
<td>0.248</td>
<td>0.129</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Demographic &amp; Land Use</td>
<td>Population of TAZ (thousand)</td>
<td>1.592</td>
<td>1.934</td>
<td>0.000</td>
<td>12.847</td>
</tr>
<tr>
<td></td>
<td>Population aged 65 years and over (thousand)</td>
<td>0.133</td>
<td>0.243</td>
<td>0.000</td>
<td>4.251</td>
</tr>
<tr>
<td></td>
<td>Percentage of population aged 65 years and over</td>
<td>0.108</td>
<td>0.124</td>
<td>0.000</td>
<td>0.941</td>
</tr>
<tr>
<td></td>
<td>Average Family Size</td>
<td>2.413</td>
<td>1.142</td>
<td>0.000</td>
<td>5.000</td>
</tr>
<tr>
<td></td>
<td>Area of TAZ (/km²)</td>
<td>3.900</td>
<td>14.250</td>
<td>0.052</td>
<td>230.897</td>
</tr>
<tr>
<td></td>
<td>Land use (urban=0, suburban=1 or rural=2)</td>
<td>0.068</td>
<td>0.269</td>
<td>0.000</td>
<td>2.000</td>
</tr>
<tr>
<td></td>
<td>Number of schools</td>
<td>0.816</td>
<td>1.098</td>
<td>0.000</td>
<td>7.000</td>
</tr>
<tr>
<td>Traffic</td>
<td>Daily vehicle kilometers of travel on the state maintained roads (thousand of vehicle-kilometers)</td>
<td>45.522</td>
<td>64.411</td>
<td>0.000</td>
<td>594.863</td>
</tr>
</tbody>
</table>

Paper revised from original submittal.
4. MACRO LEVEL MODELING STRATEGIES

As mentioned, previous studies adopted different levels for modeling area crash frequencies at the macro-level, including the ward (29; 30), county (31; 32), TAZ (4; 6; 7), census block group (13), and census tract levels (48). However, these studies modeled the total crashes including those occurring on the zonal boundary and within the boundary of each zone with features of the corresponding zones. This is referred to as the traditional approach. It is illustrated below in Table 2.

<table>
<thead>
<tr>
<th>Modeling strategy</th>
<th>Models</th>
<th>Figure Illustration</th>
<th>Dependent Variable</th>
<th>Independent Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional approach</td>
<td>Model 1: Total</td>
<td></td>
<td>Total crashes</td>
<td>All roads and zonal features</td>
</tr>
<tr>
<td>New strategy</td>
<td>Model 2a: On-System</td>
<td></td>
<td>Crashes on the on-system roads</td>
<td>On-system roads (state-maintained roads)</td>
</tr>
<tr>
<td></td>
<td>Model 2b: Off-System</td>
<td></td>
<td>Crashes on the off-system roads</td>
<td>Off-system roads (non state-maintained roads)</td>
</tr>
</tbody>
</table>

Note: Off-System includes off-system roads from the FDOT planning office and other lower level roads.

The areas of the TAZs in Florida are usually between 0.65 to 2.59 square kilometers (49). The average diameter of the TAZs in Orange County is about 1.41 kilometers. The crash record contains the postal codes of the involved drivers’ home locations. Using the ZIP GIS file of Orange County, the straight line distance between the site of a collision and the ZIP center of an at-fault driver’s home location was calculated for each crash. About 75% collisions occurred farther than 4.8 kilometers from the at-fault drivers’ home ZIP centers. This indicates a large percentage of people involved in crashes were not living in the TAZs where they were involved in crashes.

The state-maintained roads are usually arterials that provide direct service for relatively longer travel distances. We found that collisions for many at-fault drivers occurred on state-maintained roads far from the driver’s home TAZ. It is therefore reasonable to assume that many collisions on state-maintained roads may not be strongly associated with the network and land use features of the TAZs where the crashes occurred. The non state-maintained roads are
usually local roads, some of which are collectors. These roads collect traffic from the bottom-level system of local roads. Crashes occurring on non state-maintained roads within TAZs are more strongly associated with the TAZ’s features. The new strategy adopted here is designed to capture these different effects of different roadway functional classes by modeling crashes of the state-maintained roads and local roads separately. As shown in Table 2, some on-system roads are located within TAZs. In Orange County, about 80% of the state-maintained roads are on the boundaries of TAZs.

5. STATISTICAL METHODOLOGY

Poisson models are frequently used in traffic safety studies to model random, discrete, nonnegative, and sporadic crash data. If we let $y_i$ represent the number of crashes occurring for the TAZ$_i$, the Poisson model assumes the dependent variable $y_i$ follows the Poisson distribution as:

$$y_i \sim \text{Poisson} (\lambda_i)$$

Where $\lambda_i$ is the expectation of $y_i$.

For Poisson models, the logarithm was used as a link function to link the expectation of $y_i$ with explanatory variables as:

$$\log (\lambda_i) = \psi_i = X\beta$$

Where $X$ is the covariate matrix, and $\beta$ is the vector of regression coefficients.

Because TAZs in the study area cannot be assumed to be spatially independent, and adjacent TAZs may have similar safety characteristics, we proposed using the Conditional Autoregressive (CAR) models to capture this spatial dependence between TAZs. Guo et al. employed a CAR model to estimate crash frequencies of spatially correlated signalized intersections along corridors and found that the Poisson CAR model outperformed all the alternative models (38). The CAR models introduce a random effect variable $\phi_i$ into the model to account for the spatial correlation, which is defined as follows:

$$\psi_i = X\beta + \phi_i$$

A proximity matrix $W$ with entry $w_{ij}$ was used to illustrate the spatial relationship between TAZ$_i$ and TAZ$_j$ as:

$$w_{ij} = \begin{cases} 1, & \text{if TAZ}_i \text{ is adjacent to TAZ}_j \\ 0, & \text{if TAZ}_i \text{ is not adjacent to TAZ}_j \end{cases}$$

In the Bayesian framework, the conditional distribution of CAR prior of random effect variable $\phi_i$ is:

$$\phi_i | \phi_{(-i)} \sim N \left( \sum_j \frac{w_{ij}}{\tau_c w_{i+}^{-1}} \phi_j, \frac{1}{\tau_c w_{i+}} \right)$$

where $\phi_{(-i)}$ is the set of $\phi_j$s, for any $j \neq i$. $\tau_c$ is a precision parameter, and $w_{i+}$ is the collection of $w_{i,j}$ of those TAZs adjacent to TAZ$_i$. 

TRB 2012 Annual Meeting Paper revised from original submittal.
The Deviance Information Criterion (DIC) was used to compare the models (39). DIC is defined as

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

(8)

where $\overline{D}$ is the Bayesian deviance which measures the goodness of fit of the model, and $P_D$ is effective number of parameters which measures the complexity of the model. Models with smaller DIC values are preferred.

6. ROAD NETWORK GRAPHICAL PATTERNS AND TOPOLOGICAL FEATURES

In this section, Closeness Centrality, Betweenness Centrality, and the Meshedness Coefficient were evaluated to determine which index can best quantify the different network patterns. The calculated values for Meshedness Coefficients and Closeness Centrality Coefficients are continuous variables. These continuous variables are based on the topological features of the network structure and were used in the statistical model to assess their relationship to safety. As mentioned, road network structures have been traditionally categorized into different graphical classes, (e.g., grid, parallel, loops & lollipop). It is necessary to explore which type of quantitative indices are the most appropriate measures of network structures for quantitatively distinguishing different graphical pattern types.

In this study, 662 TAZ road network structures were visually inspected and a graphical street pattern was assigned to each TAZ. The effect of Meshedness Coefficient alone on safety may be more difficult to understand because it is an abstract variable and it is difficult to imagine what it corresponds to. In order to highlight the effect of each pattern on safety, the Meshedness Coefficient was transferred to the commonly classified street patterns. As shown in Table 3, graphical patterns with Meshedness Coefficients ranging from -0.3 to 0 are associated with sparse street patterns; from 0 to 0.11 with loops & lollipops; from 0.11 to 0.17 with the mixed pattern; and from 0.17 to 0.35 with the parallel & grid pattern. This close association is unique to the Meshedness Coefficient. Both the Closeness and Betweenness Centrality coefficients cannot be systematically linked to these street patterns.
TABLE 3 Summary of Street Patterns and Meshedness Coefficients

<table>
<thead>
<tr>
<th>Street Patterns</th>
<th>Sparse</th>
<th>Loops &amp; Lollipops</th>
<th>Mixed</th>
<th>Parallel &amp; Grid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roadway network figures</td>
<td><img src="image1" alt="Sparse Network" /></td>
<td><img src="image2" alt="Loops &amp; Lollipops Network" /></td>
<td><img src="image3" alt="Mixed Network" /></td>
<td><img src="image4" alt="Parallel &amp; Grid Network" /></td>
</tr>
<tr>
<td>Sample size</td>
<td>68</td>
<td>238</td>
<td>171</td>
<td>185</td>
</tr>
<tr>
<td>Meshedness Coefficient</td>
<td>(-0.3, 0]</td>
<td>(0, 0.11]</td>
<td>(0.11, 0.17]</td>
<td>(0.17, 0.35]</td>
</tr>
</tbody>
</table>

The box-plot shown in Figure 5 below illustrates the relationship between total crashes per square kilometer and categorized Meshedness Coefficients for the studied TAZs. Different network structures are associated with the different crash occurrence levels. The TAZs with parallel & grid street patterns are correlated with higher crash frequencies than those with mixed, loops & lollipops, and sparse street patterns.

![FIGURE 5 TAZs road network structures and total crashes (per km²)](image5)

7. BAYESIAN MODELING ASSESSMENT OF ROAD NETWORK STRUCTURES

The Bayesian CAR models were developed for the traditional approach, and the new modeling approach that treats on-system and off-system crashes separately. The safety effect of the categorical variable of network structures based on the calculated topological features and
original continuous form of the network indices were assessed in each of the models. Other relevant roadway and demographic features were added into the models before the final calculations were performed.

**Traditional Approach -- Model 1**

As a test of the traditional approach, a spatial CAR Bayesian model was developed for the total crashes of each TAZ in Orange County for the year 2006. The estimation results are presented in Table 4. The categorized value of Meshedness was found to be closely associated with the properties of the network structures. Some other statistically significant variables were included in the model as control factors. For total crashes, it can be seen that the road network patterns are significantly associated with the level of safety. The model indicates that the proportion of arterial and collector roads of all functional classification roads is positively related to the frequency of crashes. This is similar with previous findings that show a positive correlation between longer mileage of higher classification roads and crash occurrence (6; 16; 36).

**TABLE 4 Traditional Approach Analysis for Total Crashes (Model 1)**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Posterior Means</th>
<th>Posterior S. D.</th>
<th>95% C. I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>2.314</td>
<td>0.075</td>
<td>(2.182, 2.458)</td>
</tr>
<tr>
<td>Population ((10^3))</td>
<td>0.063</td>
<td>0.017</td>
<td>(0.030, 0.098)</td>
</tr>
<tr>
<td>Proportion of population aged 65 years and over</td>
<td>-0.451</td>
<td>0.209</td>
<td>(-0.871, -0.051)</td>
</tr>
<tr>
<td>Number of Schools</td>
<td>0.092</td>
<td>0.022</td>
<td>(0.049, 0.134)</td>
</tr>
<tr>
<td>Street Patterns (Base: Parallel &amp; Grid)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed</td>
<td>-0.082</td>
<td>0.0675</td>
<td>(-0.215, 0.049)</td>
</tr>
<tr>
<td>Loops &amp; Lollipop</td>
<td>-0.145</td>
<td>0.067</td>
<td>(-0.278, -0.014)</td>
</tr>
<tr>
<td>Sparse</td>
<td>-0.772</td>
<td>0.117</td>
<td>(-1.002, -0.542)</td>
</tr>
<tr>
<td>Number of signals on the state-maintained roads</td>
<td>0.363</td>
<td>0.033</td>
<td>(0.297, 0.426)</td>
</tr>
<tr>
<td>Total arterial and expressway as a portion of the total roads</td>
<td>1.385</td>
<td>0.270</td>
<td>(0.844, 1.894)</td>
</tr>
<tr>
<td>Total roads (km)</td>
<td>0.010</td>
<td>0.002</td>
<td>(0.006, 0.014)</td>
</tr>
<tr>
<td>CAR effect</td>
<td>0.704</td>
<td>0.105</td>
<td>(0.547, 0.953)</td>
</tr>
<tr>
<td>DIC</td>
<td></td>
<td></td>
<td>4264.250</td>
</tr>
</tbody>
</table>

**On-system/Off-System Approach -- Models 2a and 2b**

The spatial CAR Bayesian models developed for on-system and off-system crashes separately are shown in Tables 5 and 6, respectively. For on-system collisions, the road and traffic characteristic variables (i.e., VKT, number of signals, number of intersections, etc.) all showed a significant positive relationship to safety as shown in Table 5. However, the street patterns of TAZs were not significant for on-system crashes and were not included in the model. This could be attributed to the structure of the on-system roads which tend to be straight and direct (higher hierarchy arterial roads). They rarely fit in the four aforementioned patterns.
TABLE 5  New Modeling Approach for On-system Crashes (Model 2a)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Posterior Means</th>
<th>Posterior S. D.</th>
<th>95% C. I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.648</td>
<td>0.078</td>
<td>(0.504,0.793)</td>
</tr>
<tr>
<td>Daily Vehicle kilometers of travel (thousands of vehicle-kilometers--VKT)</td>
<td>0.007</td>
<td>0.001</td>
<td>(0.005,0.009)</td>
</tr>
<tr>
<td>Number of signals on the state-maintained roads</td>
<td>0.427</td>
<td>0.068</td>
<td>(0.295,0.559)</td>
</tr>
<tr>
<td>Number of intersections on the on-system roads</td>
<td>0.050</td>
<td>0.015</td>
<td>(0.020,0.084)</td>
</tr>
<tr>
<td>Total state-maintained roads density (km/km²)</td>
<td>0.206</td>
<td>0.063</td>
<td>(0.087,0.330)</td>
</tr>
<tr>
<td>CAR effect</td>
<td>0.239</td>
<td>0.027</td>
<td>(0.195,0.301)</td>
</tr>
<tr>
<td>DIC</td>
<td></td>
<td>3404.130</td>
<td></td>
</tr>
</tbody>
</table>

The four road network patterns of sparse, loops & lollipops, mixed, and parallel and grid were examined in this study for off-system crashes as shown in Table 6. Using the parallel & grid street pattern as the base case, estimation results showed that mixed, loops & lollipop, and sparse street patterns were significantly associated with fewer crashes than parallel & grid patterns. The mixed, loops & lollipop, and sparse street patterns are associated with lower level of Meshedness compared to the parallel & grid patterns (Table 3). Therefore, we found that as the level of Meshedness decreases, the level of safety consistently increases. The sparse street pattern was found to be the safest among four patterns, probably because the sparse pattern indicates less roadway density per square km. This in turn could mean less exposure. A disadvantage of the sparse pattern, however, is its lack of sufficient connectivity. The second safest pattern, the loops and lollipops pattern, also has a low crash risk associated with it. This may be attributable to its extensive use of curved roadways that slow speeds, and a resulting possible improvement in drivers’ concentration. However, this street pattern also lacks sufficient connectivity to major roads within the TAZ. The mixed pattern’s middle range performance is attributable its incorporation of both safer and more dangerous patterns. These patterns offset each other, and lead to a safety benefit that is not as good as either sparse, or loops and lollipops, but is still better than grid and parallel patterns.
TABLE 6  New Modeling Approach for Off-system Crashes (Model 2b)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Posterior means</th>
<th>Posterior S. D.</th>
<th>95% C. I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1.617</td>
<td>0.095</td>
<td>(1.432,1.807)</td>
</tr>
<tr>
<td>Population (thousand)</td>
<td>0.142</td>
<td>0.023</td>
<td>(0.101,0.190)</td>
</tr>
<tr>
<td>Proportion of population aged 65 years and over</td>
<td>-1.602</td>
<td>0.304</td>
<td>(-1.652,-0.478)</td>
</tr>
<tr>
<td>Number of Schools</td>
<td>0.153</td>
<td>0.029</td>
<td>(0.094,0.208)</td>
</tr>
<tr>
<td>Street Patterns (Base: Parallel &amp; Grid)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed</td>
<td>-0.220</td>
<td>0.090</td>
<td>(-0.398,-0.045)</td>
</tr>
<tr>
<td>Loops &amp; Lollipop</td>
<td>-0.216</td>
<td>0.091</td>
<td>(-0.394,-0.039)</td>
</tr>
<tr>
<td>Sparse</td>
<td>-0.883</td>
<td>0.153</td>
<td>(-1.188,-0.590)</td>
</tr>
<tr>
<td>Total arterial and expressway as a portion of the total roads</td>
<td>0.838</td>
<td>0.345</td>
<td>(0.183,1.530)</td>
</tr>
<tr>
<td>Total roads (km)</td>
<td>0.010</td>
<td>0.002</td>
<td>(0.004,0.015)</td>
</tr>
<tr>
<td>CAR effect</td>
<td>0.552</td>
<td>0.123</td>
<td>(0.372,0.836)</td>
</tr>
<tr>
<td>DIC</td>
<td></td>
<td></td>
<td>3539.750</td>
</tr>
</tbody>
</table>

Considering demographic features, the population level within a TAZ was positively associated with the crash frequencies in the model, and this is consistent with previous research (15). The number of elderly persons residing within a TAZ has a significant negative association with total crashes. This might be explained by their lower trip frequency compared to younger people. The number of schools within a TAZ was positively associated with the crash frequencies for total (Model 1) and off-system (Model 2b) crashes. This finding is also consistent with previous research (15; 37) and may be attributed to the greater concentration of a middle aged population in the TAZ—a characteristic generally associated with more trip making.

There are some major differences between the results of on-system crashes and off-system crashes. Firstly, population of a certain TAZ was not significant for on-system crashes while it is positively relative with off-system crashes. The reason could be that drivers involved in on-system crashes often do not live in the specific TAZ to which the crash is assigned. Therefore, on-system crashes do not correlate with the population of that TAZ. The population characteristics are more relevant to the off-system roads because people living in a TAZ are more active on off-system roads. Secondly, the street pattern is significant for off-system collisions while it is not significant for on-system collisions. The street patterns within TAZs affect less the passersby drivers on on-system roads, and the on-system crashes have less association with the road network patterns. Drivers route choice and driving characteristic within TAZs are highly influenced by the road network patterns. This indicates that both the roadway type and the structure of the road network should be considered when developing TAZ level safety models.

The CAR effect is significant for all the above models. This means that spatial dependence between TAZs exists. These dependencies may be attributable to similar characteristics shared by the adjacent TAZs. These similarities (economic status, travel
8. CONCLUSIONS AND RECOMMENDATIONS

The study proposed a new approach to TAZ level crash modeling. Crashes were divided into on-system and off-system roads whereas previous modeling approaches simply used the total crashes within each TAZ as the independent variable (6). The traditional approach does not work for Orange County (and other counties) because the distance between drivers’ homes and collision locations, (for those on-system collisions), averages about 13.0 kilometers, while the average diameter of a TAZ in Orange County is only about 1.41 kilometers. Therefore it is not reasonable to treat the crashes within each TAZ as a whole, and then try to fit a single model. The modeling approach proposed in this paper of modeling on-system and off-system crashes separately overcomes the deficiency of the traditional approach.

This study’s findings are consistent with previous research that has shown the significance of the street pattern to the overall safety (8; 9; 20). The use of objective and quantitative variables to overcome the deficiency of earlier methods that used visual scanning to classify the street network improved the prediction accuracy of the models. Of the three road network related variables, Meshedness Coefficient, Closeness Centrality, and Betweenness Centrality, that were introduced in TAZ safety modeling, the Meshedness Coefficient performed best for off-system crashes. The Meshedness Coefficient has the additional advantage of being easily computed using widely available software packages. This paper also presents a more precise method of crash allocation by incorporating the direction of travel, movement, and first harmful event of each colliding vehicle into the model. Compared to a previous research (8), these changes improved the accuracy of the models’ predictions by considering these variables as they relate to the uneven spatial distribution of crashes. In addition to street patterns, the population, schools, roadway lengths and older population affected crash occurrence on off-system. As would be expected, the daily VKT, intersections and roadway density were the major factors that influenced crashes on system.

It is recognized that the development of models predictive of a network’s safety, while important, still leaves unanswered questions regarding what is needed to optimally design a road network for a new development. These questions are at the core of the urban planning process, and go beyond what we have attempted in this paper. However, a question can be fairly raised about how these models can be used to support the urban planning process. It seems clear that once the most overarching urban planning decisions regarding population size and density, industrial, commercial, residential, and recreational land use, and primary and alternative transportation modes for the community have been settled on, then the models developed in this paper will be useful. Apart from the ability of these models to predict the safety of a given TAZ, and to compare the safety of different road network designs, the Bayesian Hierarchical modeling
approach used here is capable of predicting the safety of networks that incorporate multiple travel modes with differing distributions of service levels among these modes. At this point, however, it suggested that application of the quantitative and objective methods of network modeling methods we proposed will result in better designs and safer road networks. For an existing road network, knowing which pattern is hazardous may help managers to proactively incorporate various safety management options.

ACKNOWLEDGEMENT

This study was jointly sponsored by Shanghai Pujiang Program (No.09PJ1409900) and the Chinese National Science Foundation (51008230). The authors would like to thank the FHWA, MetroPlan Orlando and Florida Department of Transportation for funding the third author and providing the data.

REFERENCES


8. Sun, J. and Lovegrove, G. R., Using Community-Based Macrolevel Collision Prediction


42. ArcMap™, Version 9.3, Environmental Systems Research Institute (ESRI) Inc.


