Boundary Effects in Developing Macro-level CPMs: A Case Study of City of Ottawa

Feng Wei
PhD student
School of Engineering
University of British Columbia, Okanagan
Email: fwei@ubc.ca

Supervisor: Dr. Gordon Lovegrove

4 tables and 3 figures: 7*250=1750
Total words: 3297+1750=5047

Submit at 1st, April 2010

Table of Contents

ABSTRACT **3**
1. INTRODUCTION **4**
2. PREVIOUS WORK **5**
3. METHODOLOGY **6**
4. CASE STUDY **8**
5. RESULT ANALYSES **10**
5. CONCLUSION **14**
ACKNOWLEDGE **15**
REFERENCES **15**

ABSTRACT

The well-known Modifiable Areal Unit Problem shows that geo-analytical results may differ substantially according to how data are aggregated. In the process of developing macro-level Collision Prediction Models (CPMs), aggregating geo-coded data at the zonal level is a key step. Generally, Traffic Analysis Zones (TAZs) are seen as the areal units of a region and the boundaries of TAZs are mostly the major roads. Related collision history indicates that major roads are the location on which collisions occur most frequently. Therefore, how to assign the collisions and other geo-coded data along the TAZ boundaries into adjacent zones is a considerable research as it associates to the reliability of macro-level CPMs.

Based on Ottawa city data, we use General Linear Regression techniques to develop modelled and measured macro-level CPMs. Following 5 different assignment methods for geo-coded data on TAZ boundaries, 8 groups of CPMs (4 are measured, the other 4 are modelled) are built. Then, apply these CPMs in black spot program to identify the boundary effects of different aggregating methods. One result shows that the similarity rate of CPZ identification and ranking in different aggregating methods is 61% with modelled CPMs, and 63% with measured CPMs. This result not only demonstrates Fotheringham's viewpoint that collisions located near zone boundaries have an inter-zonal influence, but also illustrates that the different assignment methods for boundary data impact the developed CPM results significantly.

1. INTRODUCTION

The Moveable Areal Unit Problem (MAUP) is a statistical bias problem resulting from various ways of dividing one area into non-overlapping areal units (Openshaw, 1984). It exists in any aspect of large-scaled, spatial data. Macro-level Collision Prediction Models (CPMs) is an empirical tool used for evaluating regional road safety. It can be applied in both reactive and proactive Road Safety Improvement Programs (RSIP). To develop macro-level CPMs, extracting proxy variables as explanatory variables and collecting community-based collisions as responsible variables is a previous important work. This work ineluctably refers to the MAUP as it relates to how to aggregate zonal geo-coded data according to areal units.

The division of Traffic Analysis Zones (TAZs) takes three elements into accounts: (1) they are based on trip assignments; (2) they keep population and employment densities at a roughly uniform level; and (3) they consider the census tracts and municipalities. In recent studies (Hadayeghi & Shalaby, 2003; Ladron de Guevara & Washington, 2004; Lovegrove, 2007; Sun, 2009; Schalkwyk, 2009), TAZs are also often used as areal units because they are reasonable for purposes of macro-level CPM development and helpful to reduce the ecological fallacy. However, based on observation, most TAZ boundaries of a region are the major roads (or arterial roads), on which more traffic volumes and collisions are generated. This situation leads to that the collisions involved on boundaries may account for a relatively large proportion of total collisions. Since developing Macro-level CPMs needs aggregated data at the TAZ level, how to aggregate collisions and other geo-coded data on TAZ boundaries into adjacent zones will impact the macro-level CPM results definitely. Moreover, the more boundary geo-coded data there are, the greater effects on CPM results the boundary data assignment ways will make. This paper particularly focuses on the boundary issue of MAUP in developing macro-level CPMs under the condition that TAZs are areal units.

In the research of spatial models and GIS, Fotheringham (2000) observed that spatial data located near zone boundaries might have an inter-zonal influence. Ladron de Guevara et al (2004) examined the issue when he developed macro-level CPMs of Tucson, Arizona. However, he found that the number of collisions involved on TAZ boundaries was about 5%, which do not significantly impact their CPM results. Similar with Ladron de Guevara's boundary distribution pattern of collisions, Lovegrove (2006) assigned the boundary data in an automatic geo-spatial precision way in developing macro-level CPMs for Great Vancouver Regional District, Canada. He supposed a potential boundary effect in aggregating geo-coded data but did not research in this area. Sun (2009) revised the assignment way for boundary data when developing the CPMs for Victoria and Ottawa, Canada. He split the data overlaying boundaries according to a half-to-half proportion and assigned them into adjacent TAZs equally.

Although fewer studies focus on how to aggregate the boundary geo-coded data and the impacts of aggregation ways on CPM results, in order to make macro-level CPM results more reliable, the boundary effect is worthy to be researched. The objectives of this study are to identify the influence of collisions/geo-coded data along TAZ boundaries on macro-level CPM results by aggregating collisions/geo-coded data on/near zone boundaries in different ways and to recommend the most appropriate aggregating method.

2. PREVIOUS WORK

In this study, the macro-level CPMs use Generalized Linear Regression Model (GLM) techniques with Negative Binomial error distribution. This non-linear regression method is promising and was used more widely in recent macro-level CPM studies (Hadayeghi & Shalaby, 2003; Ladron de Guevara & Washington, 2004; Lovegrove, 2007; Jones et. al., 2008). GLM method can not only overcomes the limitation of linear regression in modelling discrete, rare and non-negative collisions; but also fits the observed data much better than the linear regression. The form of macro-level CPMs in this study is extracted from Lovegrove's (2007) general model form, presented as following:

$$E(\Lambda) = a_0 Z^{b_0} \quad (1)$$

where,

$E(\Lambda)$: predicted collision frequency for a specific collision type at macro-level.

Z: zonally lead exposure variables. (e.g. VKT, TLKM)

a_o, b_0: model parameters derived via GLM process.

This general model form meets two points: (1) zero risk logic (zero VKT or TLKM means zero collisions); and (2) no negative collisions. Usually, the resulting prediction models are stratified into 4 categories according to land use (urban and rural) and data derivations (modelled and measured). One exposure variable, VKT, is derived from modelled data, and the other exposure variable, TLKM is derived from measured data. The development of macro-level CPMs follows the Lovegrove's GLM process (2007). Pearson χ^2 and Scaled Deviance are quantitative measures used to test the model goodness of fit. These two measures are originally described in McCullagh and Nelder (1989); updated by Sawalha & Sayed (2006) in micro-level CPM development; and then drawn by Lovegrove (2007) in developing macro-level CPMs. Pearson χ^2 and Scaled Deviance are defined in Equations (2) and (3), respectively.

$$Pearson\chi^2 = \sum_{i=1}^{n} \frac{\left[y_i - E(\Lambda_i)\right]}{Var(y_i)} \quad (2)$$

$$SD = 2\sum_{i=1}^{n}\left[y_i \ln\left(\frac{y_i}{E(\Lambda_i)}\right) - (y+\kappa)\ln\left(\frac{y_i+\kappa}{E(\Lambda_i)+\kappa}\right)\right] \quad (3)$$

where,

y_i is zonal observed collisions for TAZ i

$E(\Lambda_i)$ is the predicted collision for TAZ i

κ is the over-dispersion parameter of Negative Binomial error distribution, which is derived from GLM process.

Var(y_i) is variance of the observed mean collision frequency at TAZ *i*. For NB error distribution, it is formulated as $Var(y_i) = E(y_i) + \frac{E(y_i)}{\kappa} = E(\Lambda_i) + \frac{E(\Lambda_i)}{\kappa}$

In the goodness-of-fit analysis, the values of SD or Pearson χ^2 should be less than the value of Chi Square distribution with *n-p* degrees of freedom at a desired level of confidence (e.g. 95%). *n* is the number of TAZs in the region and *p* is the number of parameters in the model. If a model does not meet goodness-of-fit measures, outlier analysis for model refinement will be taken. Outlier analysis addresses on data quality problem and aims to remove the response or explanatory data points unusual or not typical with the rest of data points. The Cook' Distance (CD) (Sawalha & Sayed, 2001) is used for removing these non-qualified points from samples. It is formulated as the Equation (4). By sorting the observations in descending order of CD values, the points with the largest CD values are removed in a stepwise progression. As each point is removed, regression parameter estimation is re-run with κ at its previous value to observe the changes in SD value. If the SD change is greater than $\chi^2_{0.05,1} = 3.84$, GLM process is re-run to provide new parameter estimates, new κ, and new CD values. The outlier analysis will be stopped until the refined CPM meets the goodness-of-fit measures: Pearson χ^2 and Scaled Deviance.

$$CD = \frac{h_i}{p(1-h_i)}(r_i^{ps'})^2 \tag{4}$$

where,

h_i is the leverage value, gotten in GLM process;

$r_i^{ps'}$ is the standardized residual of TAZ i, calculated as $\frac{E(\Lambda_i) - y_i}{\sqrt{(1-h_i)Var(y_i)}}$;

p is the number of model parameters.

3. METHODOLOGY

Three softwares are needed for macro-level CPM development. The transportation planning software (e.g. VISSUM, Emme) is used to generate a transportation planning model to provide modelled data, such as VKT; the geographic information system software (e.g. ArcGIS, Mapinfo) is applied to aggregate zonal geo-coded data; and the statistics software (e.g. GLIM, GenStat) is used for GLM process to estimate all parameters in the model.

It is noted that the collision locations in the following case study have been geo-coded claimed as either mid-block or intersection. Therefore, the reported collision locations are assumed as reasonably accurate. However, we still note that many collisions occur on the boundaries of two adjacent TAZ or in the intersection points of several TAZs (as shows in Figure 1). In this case, 5 different assigning methods for these boundary collisions are proposed:

Method 1: assign collisions using geo-spatial precision available.
Method 2: assign boundary collisions into adjacent zones according to the Half-to-Half ratio.
Method 3: assign boundary collisions into adjacent zones according to the One-to-One ratio.
Method 4: assign boundary collisions into adjacent zones according to the ratio of $VKT_{TAZi}:VKT_{TAZj}$ *(this is for modelled CPMs).*
Method 5: assign boundary collisions into adjacent zones according to the ratio of $TLKM_{TAZi}:TLKM_{TAZi}$ *(this is for measured CPMs).*

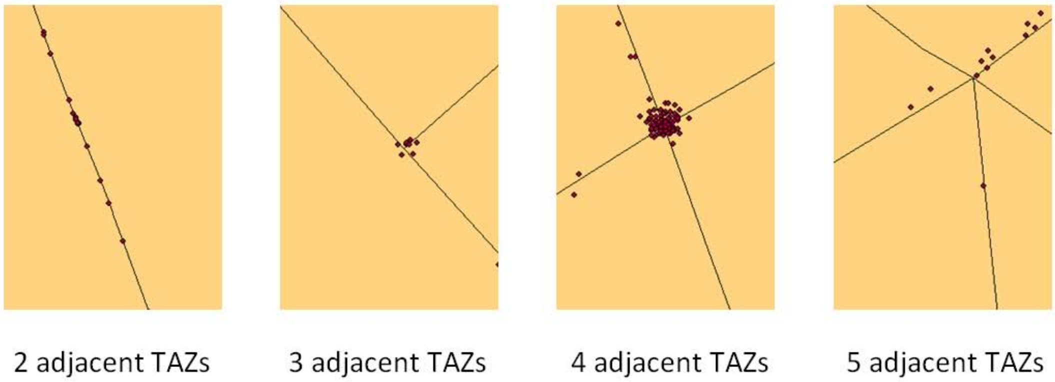

FIGURE1 TAZ boundary collisions

Table 1 presented these 5 assignment methods in a quantitative way. The first assignment method aggregates boundary spatial data in an automatic geo-process way. This method ignores how to deal with the collisions on the major roads that are seen as TAZ boundaries. Its aggregating precision depends on the accuracy degree of reported spatial data. The second and third methods are considered as to be logic since any collision happening on the boundary of two adjacent zones can be seen belong to both sides. Additionally, previous studies (Lovegrove, 2007) show that two exposure variables, VKT and TLKM, both have significant relationships with collisions. Thus, it is reasonable to assign boundary collision with VKT or TLKM ratio. In order to coordinate with the collision assigning methods, other boundary geo-coded data (e.g. VKT and TLKM) are assigned in the geo-automatic, half-to-half ratio and one-to-one ratio ways as well.

TABLE1 Boundary collision assignments in 5 different methods

	(1) GEO	**(2) HALF** $(\sum \frac{1}{n} = 1)$	**(3) ONE** $(\sum_n 1 = n)$	**(4) VKT** $(\sum_{i=1}^{n} vkt_i = 1)$	**(5) TLKM** $(\sum_{i=1}^{n} tlkm_i = 1)$
2 TAZs		0.5:05	1:1	vkt_1:vkt_2	$tlkm_1$:$tlkm_2$
3 TAZs		0.33:0.33:0.33	1:1:1	vkt_1:vkt_2:vkt_3	$tlkm_1$:$tlkm_2$:$tlkm_3$
4 TAZs		0.25:0.25:….	1:1:1:1	vkt_1:vkt_2:vkt_i….	$tlkm_1$:$tlkm_2$:$tlkm_i$….
n TAZs	……				

According to these assignment methods, various groups of CPMs are developed. By applying these models to in black spot studies, the boundary effects on macro-level CPMs can be identified. Following Lovegrove's (2007) methodology of identifying and ranking black spots, the Collision Prone Zone (CPZ) is the zone meeting the following formulation:

$$1 - \int_0^{E(\Lambda)} \frac{[\kappa / E(\Lambda) + 1]^{(\kappa + count)} \lambda^{(\kappa + count - 1)} e^{-[\kappa / E(\Lambda) + 1]\lambda}}{\Gamma(\kappa + count)} d\lambda \geq \delta \qquad (5)$$

where, δ is a 95% confidence level; $E(\Lambda)$ is the prediction collision of TAZ *i* from CPMs; and *count* is zonal observed collisions of TAZ *i*.

The twin-ranking criteria, Potential Collision Reduction (PCR) and Collision Risk Ratio

(CRR), are used to rank CPZs. EB is the zonal empirical Bayes safety estimate combined CPM results ($E(\Lambda)$) and observed results (*count*). They are formulated in Equations (6), (7) and (8). The larger the values of PCR and CRR are, the more hazardous the CPZ's safety problem is.

$$PCR = EB - E(\Lambda) \tag{6}$$

$$CRR = \frac{EB}{E(\Lambda)} \tag{7}$$

$$EB_i = \left[\frac{E(\Lambda_i)}{\kappa + E(\Lambda_i)}\right](\kappa + count) \tag{8}$$

Different from the regional road system, the limited access highway system is less related to the regional proxy variables. Therefore, the highway system is excluded from the macro-level CPM estimates and the non-highway macro-level CPM results still generally reflect reality in a region. In the aggregation process, all associated geo-coded data to highways are removed from the regional data.

4. CASE STUDY

City of Ottawa is used as a case study to identify the boundary effects on macro-level CPM results. In this case study, the collision geo-coded data is from 2005 to 2007. A 2007 EMME/2 transportation planning model is provided by City of Ottawa for generating VKT data. And the TLKM data is collected from the road network shapefile (2007) published by Statistics Canada. Within 400 TAZs of City of Ottawa, 275 zones are urban and 125 are rural. In the predicting progress, total and sever collisions over 3 years are chosen as responsible variables. Total collisions mean all classes of collisions including property damage only, non-fatality injuries, fatalities, and non-reported collisions; sever collisions only involve non-fatality injuries and fatalities. After making geo-processing analysis in ArcGIS (e.g. create a buffer of 20 meters for the boundary centroid in this case), the results show that collisions along TAZ boundaries accounts for more than 50% of total collisions in City of Ottawa. Other geo-coded data, Vehicle Kilometres Travelled (VKT) and Total Lane Kilometres (TLKM) along boundaries account for 26% and 13%, respectively. Figure 2 displays the spatial distributions of collisions, VKT, and TLKM along TAZ boundaries. The red collisions in map (a) are collision located on boundaries, and the highlight blue lines in maps (b) and (c) present the road lanes and traffic links overlay boundaries.

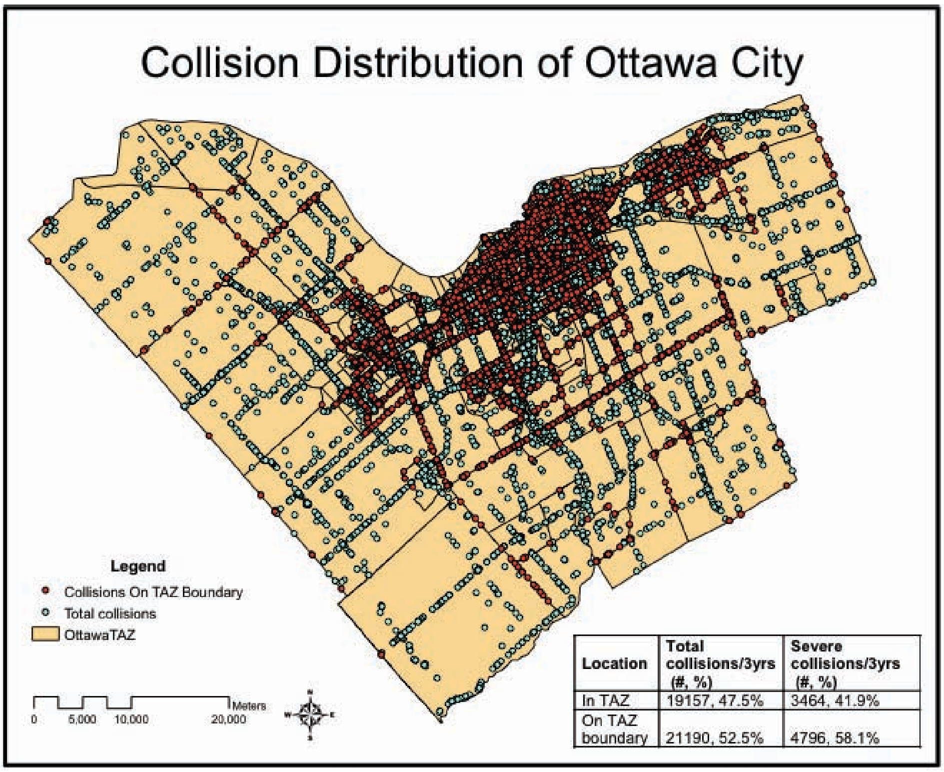

FIGURE2 Spatial distributions of collisions

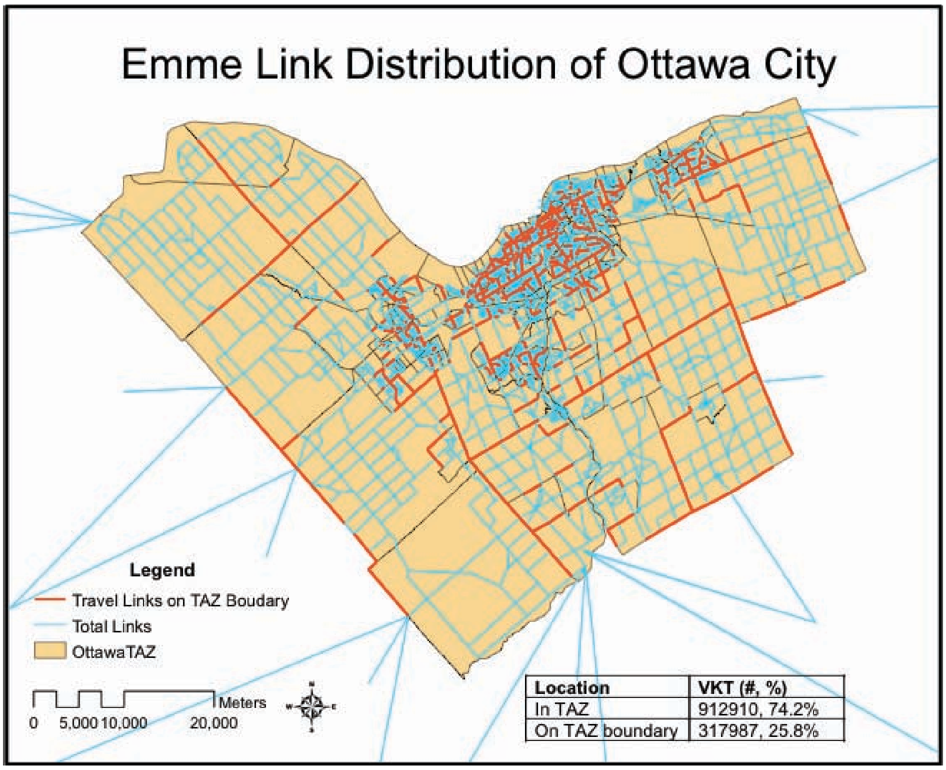

FIGURE3 Spatial distribution of VKT

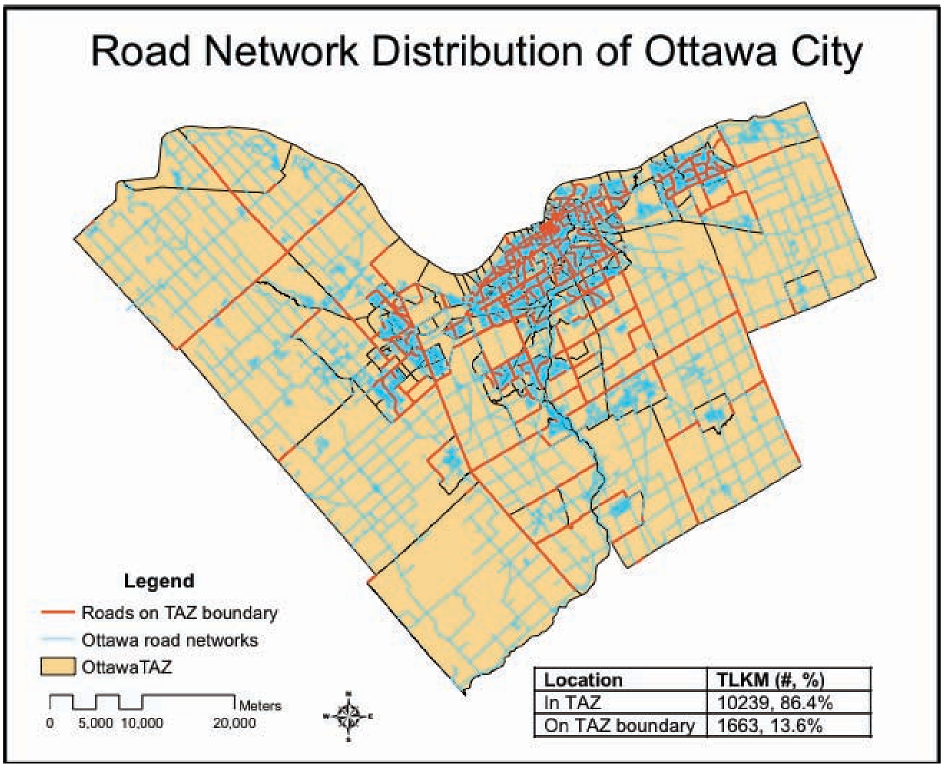

FIGURE4 Spatial distribution of TLKM

5. RESULT ANALYSES

The modelled and measured general CPMs for 275 urban TAZs in City of Ottawa are provided in this study. Four groups of modelled CPMs are developed according to the assignment methods (1), (2), (3), and (4) mentioned in Section 3; and the other four groups of measured CPMs are developed based on assignment methods (1), (2), (3), and (5). Tables 2 and 3 present the urban CPM results and their related statistics information.

TABLE2 Urban, modelled exposure CPMs-Total/Severe (City of Ottawa, 2007)

Model Form	κ	DoF	Pearson χ^2	SD	$\chi^2_{0.05,dof}$	t-Statistics
Total collisions/3yrs (T3)						
(1) Geo-spatial automatically						Const: 2.91
$T3=2.226VKT^{0.5112}$	4.17	273	266	291	313	vkt: 13.87
(2) Half to half assignment						Const: 2.87
$T3=2.201VKT^{0.5117}$	4.82	273	277	289	313	vkt: 13.96
(3) One to one assignment						Const: 5.47
$T3=5.173VKT^{0.4512}$	4.28	273	276	289	313	vkt: 11.71
(4) VKT ratio assignment						Const: -4.38
$T3=0.2284VKT^{0.8046}$	3.46	273	277	299	313	vkt: 17.91
Severe collisions/3yrs(S3)						
(1) Geo-spatial automatically						
$S3=0.2595VKT^{0.5820}$	4.20	273	262	300	313	Const: -4.12
(2) Half to half assignment						vkt: 13.38
$S3=0.2859VKT^{0.5698}$	5.40	273	277	298	313	Const: -3.99

(3) One to one assignment						vkt: 13.72
$S3=0.7038VKT^{0.5082}$	4.29	273	277	302	313	Const: -1.05
(4) VKT ratio assignment						vkt: 11.88
$S3=0.03192VKT^{0.8503}$	4.24	273	270	301	313	Const: -9.26
						vkt: 17.35

TABLE3 Urban, measured exposure CPMs-Total/Severe (City of Ottawa, 2007)

Model Form	κ	**DoF**	**Pearson** χ^2	**SD**	$\chi^2_{0.05,dof}$	**t-Statistics**
Total collisions/3yrs (T3)						
(1) Geo-spatial automatically						Const: 41.51
$T3=55.41TLKM^{0.2436}$	2.95	273	262	293	313	tlkm: 6.98
(2) Half to half assignment						Const: 41.88
$T3=56.94TLKM^{0.2325}$	3.28	273	264	291	313	tlkm: 6.69
(3) One to one assignment						Const: 39.36
$T3=96.21TLKM^{0.2212}$	3.14	273	253	290	313	tlkm: 5.60
(5) TLKM ratio assignment						Const: 31.06
$T3=38.42TLKM^{0.3756}$	2.28	273	258	300	313	tlkm: 8.89
Severe collisions/3yrs(S3)						
(1) Geo-spatial automatically						
$S3=10.14TLKM^{0.2782}$	2.71	273	252	301	313	Const: 20.96
(2) Half to half assignment						tlkm: 6.26
$S3=11.02TLKM^{0.2501}$	3.23	273	251	296	313	Const: 22.55
(3) One to one assignment						tlkm: 6.58
$S3=18.22TLKM^{0.2632}$	2.94	273	233	297	313	Const: 23.00
(5) TLKM ratio assignment						tlkm: 6.16
$S3=6.493TLKM^{0.4392}$	2.32	273	251	305	313	Const: 14.58
						tlkm: 9.63

By comparing different groups of CPMs, we get the following results:

(1) For both modelled and measured CPMs, the CPM parameters gotten from automatic geo-spatial precision and the assignment method of half-to-half ratio are most close to each other.

(2) For both modelled and measured CPMs, the parameters of exposure variables (i.e. VKT or TLKM) gotten from assignment methods (1), (2), and (3) are similar, but the CPM constants in assignment method (3) are larger than constants in other two methods. This is explained by that the method (3) repeatedly aggregates boundary collisions, which means one boundary collision is assigned more than one adjacent TAZ.

(3) The modelled CPMs gotten from VKT ratio assignment and the measured CPMs gotten from TLKM ratio assignment have distinct parameters, when comparing to the CPM parameters derived from other 3 assignment methods. In these two groups of CPMs, the collision prediction results are more significant to parameters of exposure variables. For example, the exponent of VKT in modelled CPMs in groups (1), (2), and (3) is around 0.5, but becomes 0.8 in the group (4). This result is reasonable because boundary collisions are aggregated into adjacent TAZs according to zonal exposure ratios.

(4) All developed CPMs in different groups have the same degree of freedom. This means the assignment methods do not impact the results of goodness-of-fit and outlier analyses significantly in this case.

8 groups of macro-level CPMs for total collisions/3years are applied to identify and rank the CPZs among 275 urban TAZs. Table 4 shows the top ten urban CPZs and the number of CPZs in different assignment methods. Figure 3 displays the maps of urban CPZ ranks. In Figure 3, CPZs are symbolized by five colours; one colour represents one hazardous level.

TABLE4 Black spots in different assignments, Top 20

Urban	**Total collisions/Modelled**				**Total collisions/Measured**			
	GEO	HALF	ONE	VKT	GEO	HALF	ONE	TLKM
CPZ rank	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(5)
1	222	222	222	222	523	222	222	222
2	221	210	210	210	1070	523	1523	523
3	523	523	1523	523	1523	2151	210	2151
4	1070	2151	202	533	222	1041	1070	1523
5	2151	202	221	513	2151	1523	523	201
6	210	743	1070	230	2152	1640	2272	210
7	743	221	523	1070	1041	210	2621	721
8	202	1640	741	621	221	1070	1023	1100
9	1041	1041	3012	743	721	2152	1100	533
10	3012	513	1640	521	210	201	1640	1041
11	1523	1070	2151	2151	2101	1461	202	1070
12	602	533	2272	3121	602	602	2101	2152
13	513	3121	743	512	1461	513	2152	513
14	533	602	230	221	201	2101	2151	2621
15	1521	820	1023	1640	513	202	1041	1842
16	3121	621	1041	3012	2272	1023	2611	1023
17	820	3012	533	1041	743	1100	741	1461
18	2152	230	3121	741	1023	721	221	602
19	621	2732	513	2732	820	820	533	651
20	230	651	2621	1523	2720	533	513	2720
CPZ #	87	88	103	93	87	90	98	88

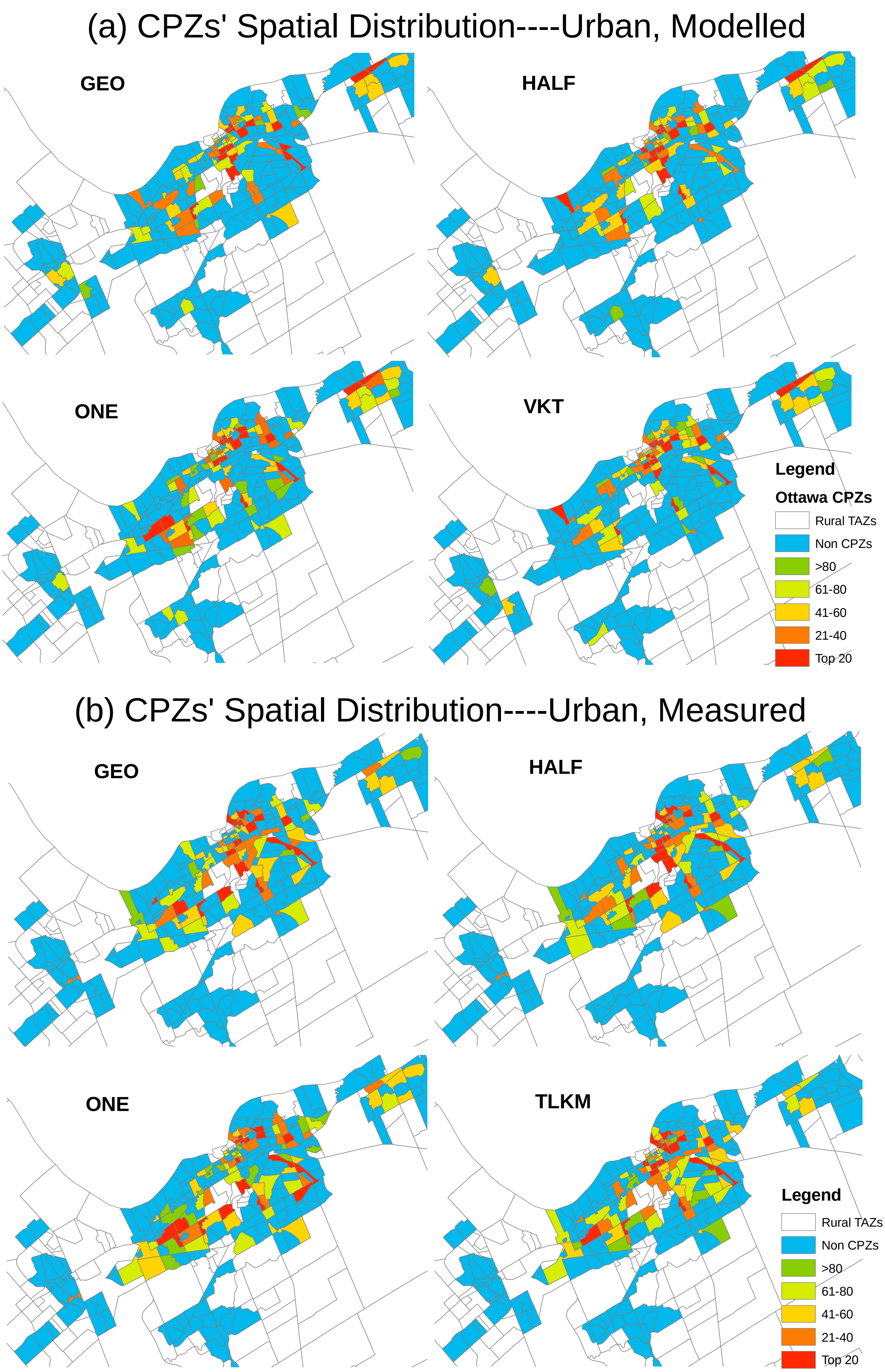

FIGURE3 Urban CPZs and non-CPZs distribution in City of Ottawa

Different black spot results are attributed to the individual CPM form in its associated assignment method. As mentioned previously, the ranking of CPZs combines the analysis results from PCR and CRR measures. According to a descending hazardous ranking order, every 20 CPZs are summed in one level and all CPZs are divided into five levels. By analysing, several results are concluded as follows:

(1) The number of CPZs identified by one-to-one assignment (Method 3) is much larger than CPZ numbers identified by other methods. This illustrates that the assignment method 3 may exaggerate the CPZ identification result because it repeatedly aggregates the same collisions. Therefore, this method is not recommended in assigning boundary geo-data.

(2) Both modelled and measured results show that the CPZ identification and ranking getting from the geo-automatically assignment (Method 1) and half-to-half assignment (Method 2) have the highest similarity. Their similarities of CPZ ranking levels are 74% based on modelled CPMs and 82% based on measured CPMs.

(3) Via the modelled CPMs, the CPZ identification and ranking result in VKT ratio assignment (Method 4) has the least similarity with results from other 3 methods. Similarly, via the measured CPMs, the result in TLKM ratio assignment (Method 5) has the least similarity with results in other methods. This result intuitively coordinates with the fact that the model parameters getting from assignment methods (4) and (5) are significant different from the parameters in other 3 methods.

(4) Both modelled and measured results show that the black spot result in half-to-half assignment (Method 2) has a higher similarity with results from other 3 methods.

(5) By modelled CPM application, the similarity rate of CPZ identification and ranking result for all 4 assigning methods is 61%. By measured CPM application, the similarity rate for all 4 assigning methods is 63%.

5. CONCLUSION

In this paper, the macro-level CPMs described as an empirical tool of evaluating road safety level is designed to demonstrate the magnitude of boundary effects associated the aggregating ways for boundary data. There are five assignment methods for boundary geo-coded data proposed, including: automatic geo-spatial precision way, half-to-half ratio way, one-to-one ratio way, VKT ratio way and TLKM ratio way. The method examining the boundary effects here is fairly simple minded: first, zonal geo-coded data are collected by the five assignment methods; second, 8 groups of CPMs are developed based on previous collected data; third, apply these CPMs to identify and rank black spots; finally, compare the CPZ ranking results of 5 assignment methods.

Different with Ladrón de Guevara's finding (2004) on boundary collisions, this study shows that a large proportion of collisions happened near boundaries. This finding is not only proved truly in Ottawa data, but also in GVRD and Victoria data. The big differences from Ladrón de Guevara probably attribute to different research regions. In addition, the comparison result states that the similarity rate of CPZ identification and ranking for all 4 assigning methods with modelled CPMs is 61% and the similarity rate for all 4 assigning methods with measured CPMs is 63%. This result not only demonstrates Fotheringham's viewpoint that collisions located near zone boundaries have an inter-zonal influence, but also illustrates that the different assignment methods significantly impact the developed CPM results. Despite lacking of a reliable methodology for evaluating which assignment method is the most appropriate one, via a simple statistical comparison, the second assignment method discussed in this study:

half-to-half ratio assignment way is best recommended for application. In an intuitive view, this method assigns the boundary spatial data into adjacent zones equally without increasing the total collision number of a region (different with one-to-one ratio assignment); in a statistic view, the CPZ identification and ranking result from this method has the highest similarity rate with other assignment methods.

The boundary effects on macro-level CPM results were hardly researched before; this study can be a primary one in this area. In future, developing reliable empirical tools for evaluating the methods of aggregating boundary geo-coded data should be a considerable research topic.

Acknowledge

I would like to thank Dr. Gordon Lovegrove for giving me many good suggestions on this study. I also want to appreciate City of Ottawa for providing the collision data and the Emme model data.

REFERENCES

1. Openshaw, S. Ecological Fallacies and the Analysis of Areal Census Data. *Environment and Planning A 16*, Pion Ltd, London, UK, 1984, pp.17–31.
2. Fotheringham, A. S., and M. Wegener. *Spatial Models and GIS: New Potential and New Models*.UK: London, 2000.
3. Hadayeghi, A., A. Shalaby, and B. Persaud Macro-level Accident Prediction Models for Evaluating the Safety of Urban Transportation Systems, *Transportation Research Record: Journal of the Transportation Research Board*, TRB, Washington D.C., Vol. 1840, 2003, pp. 87-95.
4. Ladrón de Guevara. F, S. Washington, and J. Oh. Forecasting Crashes at the Planning Level: Simultaneous Negative Binomial Crash Model Applied in Tucson, Arizona. *Transportation Research Record: Journal of the Transportation Research Board*, TRB, Washington D.C., Vol. 1897, 2004, pp. 191-199.
5. Lovegrove, G. *Road safety Planning: New Tools for Sustainable Road Safety and Community Development*, Verlag Dr. Müller, Germany, 2007.
6. Sun J. *Sustainable Road Safety: Development, Transference, and Application of Community-based Macro-level Collision Prediction Models*. University of British Columbia Okanagan, Canada, 2009.
7. McCullagh, P., & Nelder, J.A. *Generalized Linear Models*. New York: Chapman and Hall, 1989.
8. Sawalha, Z. & T. Sayed “Traffic Accident Modeling: Some Statistical Issues”, in *CSCE Journal*, Vol 33, Issue 9, Canada: CSCE, 2006, pp. 1115-1123.
9. Jones, A.P., Haynes, R., & Kennedy, V. “Geographical Variations in Mortality and Morbidity from Road Traffic Accidents in England and Wales”, *Health & Place*, Vol. 14, 2008, pp. 519–535.
10. Sawalha, Z. & T. Sayed, “Evaluating Safety of Urban Arterial Roadways”, *Journal of Transportation Engineering*, 127(2), ASCE, USA, 2001,pp151-158.