

مجلة

كلية التراث الجامعة



رقم الايداع في دار الكتب والوثائق 719 لسنة 2011

مجلة كلية التراث الجامعة معترف بها من قبل وزارة التعليم العالي والبحث العلمي بكتابها المرقم
(ب 3059/4) والمؤرخ في (2014/ 4/7)



Analysis of Traffic Accidents in Iraq using Geographically Weighted Poisson Regression

Saba A. Jabbar

Eman M. Abdullah

University of Baghdad

Abstract

Injuries resulting from traffic accidents are the first cause of death for persons aged 15 to 29 years, according to the World Health Organization report on traffic accidents issued on 20-June-2022. For the study and analysis of traffic accidents, it is preferable to use spatial analysis to know the influencing causes and according to each region. Therefore, geographically weighted Poisson regression models (which is an updated model of the traditional regression, so that the diagnosis is more accurate for the studied phenomenon by adding the spatial dimension to the model, so the new model takes into account this relationship (the spatial dimension), that is, the geographically weighted regression (GWR) performs a per-site regression rather than a single regression for the entire study area) were used for this purpose. As the fixed relationship may hide some important spatial factors that affect the relationship between the response variables and the explanatory variables, and then may negatively affect the accuracy of the models. The geographically weighted Poisson regression (GWPR) model was used in this study on a sample of traffic accidents for the period from 2006 to 2022 for all governorates of Iraq except for the Kurdistan region for the purpose of carrying out prior work to determine the locations in which traffic accidents may increase. Two Kernel functions (gaussian and tri-cube) were also used. The study showed that the tri-cube function was the best and according to the corrected Akaike information criterion (AICc) .

Paper type : Research paper

Keywords : Traffic Accidents , Geographically Weighted Poisson Regression , Gaussian, Tri-cube , corrected Akaike information criterion (AICc) .

1. Introduction

A traffic accident is an unplanned event that occurs when one or more cars (vehicles) collide with other cars (vehicles), pedestrians, animals or objects on a public or private road. A traffic accident usually results in damages and injuries ranging from minor to property and vehicles. serious cause of death or permanent disability (Dhahad , 2015) . According to the report of the World Health Organization on traffic accidents issued on June 20, 2022, injuries resulting from traffic accidents were the first cause of death for people aged 15 to 29 years, and about 1.3 million people die annually as a result of traffic accidents. Another million people suffer non-fatal injuries, many of whom become incapacitated as a result of their injuries (World Health Organization , 2022) . Ewadh & Habieb (2007) studied the basic rates of traffic accident data that follow the binomial distribution, as there is a way to estimate the basic rate in the event that the data distribution is a Poisson distribution, while this case is not available in other distributions, as the research presents a new way to estimate higher limits And the lowest baseline accident rate based on binomial distribution data with different degrees of confidence. The method included examining data with regard to randomness and appropriate distribution.

The optimal time for the duration of traffic accident data provides a relatively accurate estimate of the rate of that data. Most of them show a "relatively little" lack of uncertainty in estimating the basic rates of traffic accidents (Ewadh & Habieb , 2007). The two researchers, Zalzal and Mahmoud (2013) published a paper entitled "A Statistical Study of Traffic Accidents in Iraq for the Year (2006)" by studying a set of variables to determine the relationships between them and to know the extent to which they are related to each other, such as (the causes of the accident and the class of the road) or in terms of the nature of the accident. being (collision, rollover, run-over, other accidents), and the two researchers concluded that the driver is the main cause, and the (main) road category is the highest among the road categories and drivers of the two age groups (24-29) and (30-35) as they constituted the highest percentage of among the rest of the categories (Zalzal and Mahmoud ,2013) . In (2015) the researcher AL-Obaedi studied the rates of traffic accidents in the city of Diwaniyah, using real data from the Diwaniyah Traffic Directorate for recorded accidents for the period from 2004 to 2013, as well as creating a questionnaire to find out the number of traffic accidents not registered with the Traffic Directorate and The researcher presented some proposals to reduce traffic accidents in the mentioned city, including the use of barriers to regulate pedestrian crossing and the use of bridges and tunnels at traffic intersections to reduce vehicle collisions (AL-Obaedi , 2015) . Al-Saadi published in (2017) a study entitled "Forecasting Fatal and Non-Fatal Traffic Accidents, Numbers of Deaths and Numbers of Injuries in Iraq Using Optimization Models" aimed at predicting the period from (2014-2026) in addition to indicating the general trend line for the numbers of fatal and non-fatal accidents, the number of dead and the number of injuries in Iraq for the time period (2002-2013) (Al-Saadi , 2017) . In the year (2019), the researcher Abbas discussed through his research, which aims to analyze the time series of the number of traffic accidents in the southern governorates of Iraq (Basra, Dhi Qar, Muthanna, Maysan, and Qadisiyah) for the period from (1979-2017) to choose the best prediction model for the years (2018 -2022) using the (Box Jenkins) method and comparing it with the traditional method of general trend models (Abbas , 2019). As for using the geographically weighted Poisson regression" in the case of traffic accidents, in 2010, (Hedayeghi et al) used the geographically weighted Poisson regression model (GWPR) to study the effect of local spatial variables on the relationship between the number of collisions occurring in a region, and the potential influences of transportation planning, such as traffic volume, roads, the economic and social level, demographic characteristics and other influences, explaining that the main advantage of GWPR models is that the parameters of the explanatory variables can vary spatially, which can add importance to the spatial location of the data, such as the location of the collision using The data for this study was based on the data available for City of Toronto's 481 traffic analysis zones in year 2001 (Hedayeghi et al , 2010) . Hezaveh et al (2019) conducted a study to estimate the geographically weighted Poisson regression model, using the (Gaussian) and (Bi-square) functions, and several criteria for comparison to reduce the number of traffic accidents in the state of Tennessee in the United States of America. They found that many reasons affect the number of traffic accidents. Therefore, the geographically weighted Poisson regression model was the most appropriate (Hezaveh et al ,2019) . In the year 2021 (Al-hasani et al) conducted a study for the comparison between some exponential functions of the geographically weighted Poisson regression model. They applied the model formula to traffic accident data in the

Sultanate of Oman and concluded that the exponential function provided the best fit for the data (Al-hasani et al , 2021) .

1.1 The Problem of Research paper

Traffic accidents lead to loss of life and cost to societies in general. These losses may be minor to vehicles and property, or they may be serious, leading to death or permanent disability. According to the World Health Organization report dated 6/21/2022, “1.3 million people die annually in the world due to road accidents, and road accidents are the main cause of death for children and young people between the ages of 5-29 years.” Therefore, this phenomenon has become a problem in all societies, especially the Iraqi society, where the governorates and cities of Iraq are witnessing a continuous population increase. This increase leads to an increase in the number of vehicles (wheels), and thus an increase in traffic accidents on the roads, in addition to fast driving, bad roads, lack of maintenance and care for them, and other reasons.

1.2 Research paper Objective

Choosing the Geographically Weighted Poisson Regression (GWPR) model to predict the number of traffic accidents in Iraq to help the concerned authorities develop appropriate strategies and take the necessary measures to eliminate or reduce this dangerous phenomenon.

2. Materials and Methods

In this paper, we will estimate the geographically weighted Poisson regression model (GWPR) using the Kernel functions to analyze traffic accidents in Iraq .

2.1 Geographically Weighted Poisson Regression Model

The geographically weighted regression (GWR) model is used, which is an updated model of the traditional regression, so that the diagnosis is more accurate for the studied phenomenon by adding the spatial dimension to the model, so the new model takes into account this relationship (the spatial dimension), that is, the geographically weighted regression (GWR) performs a per-site regression rather than a single regression for the entire study area (Jasim , 2014) . The regular regression models give one equation for whole study area, while the spatial regression models give a different equation for each part (spot) of the study area, that is, for each part there is an equation that differs from the equation of a second part, and so on, in order to find out which variables are effective and which are not, or their effect is less (Dawod , 2018). Therefore, spatial modeling techniques provide significantly improved estimates and predictions compared to non-spatial models(Al-hasani et al , 2021). As the fixed relationship may hide some important spatial factors that affect the relationship between the response variables and the explanatory variables, and then may negatively affect the accuracy of the models(Hezaveh et al ,2019). When the variables are continuous (non-countable), the geographical weighted regression (GWR) is used, and when the variables are discontinuous (countable), it is preferable to use the geographically weighted Poisson regression (GWPR) as predicting the number of patients who visit the emergency room in the hospital or the death rate during a certain period or number of Traffic accidents for a specific area during a period of time and so on (Al-hasani et al , 2021). Spatial Poisson model, a special form of spatial count model, allows parameter values to vary with spatial unites u_i which is a vector describing the location i and it can be written as(Al-hasani et al , 2021) (Nakaya et al , 2005) :-

$$y_i = e^{\sum_j \beta_j(u_i)x_{ij} + e_{ij}} \quad (1)$$

where \mathbf{x}_{ij} is j th explanatory variable in location i , β_j is the parameter for the j th explanatory variable and $\mathbf{u}_i = (\mathbf{u}_{li}, \mathbf{u}_{mi})$ is the vector describing the latitude and longitude at location i . The geographically weighted log-likelihood at location \mathbf{u}_i can be given by :-

$$\mathbf{L}(\mathbf{u}_i) = \sum_{k=1}^A (-\hat{\mathbf{Y}}_k(\beta(\mathbf{u}_i)) + \mathbf{Y}_k \log \hat{\mathbf{Y}}_k(\beta(\mathbf{u}_i)) * \mathbf{W}_{ik}(|\mathbf{u}_i - \mathbf{u}_k|)) \quad (2)$$

Where :-

$$\hat{\mathbf{Y}}_k(\beta(\mathbf{u}_i)) = e^{\sum_j \hat{\beta}_j(\mathbf{u}_i) \mathbf{x}_{kj}}$$

And \mathbf{W}_{ik} is Kernel weighting functions.

To estimate the parameter vector β , we find the first derivative of the equation (2).

$$\frac{\partial \mathbf{L}(\mathbf{u}_i)}{\partial \hat{\beta}(\mathbf{u}_i)} = 0$$

$$\frac{\partial}{\partial \hat{\beta}} = \left[\sum_{k=1}^A (-\hat{\mathbf{Y}}_k(\beta(\mathbf{u}_i)) + \mathbf{Y}_k \log \hat{\mathbf{Y}}_k(\beta(\mathbf{u}_i)) * \mathbf{W}_{ik}(|\mathbf{u}_i - \mathbf{u}_k|)) \right] \quad (3)$$

To solve the above equation, we use Fisher's scoring method, as:-

$$\beta^{(\ell+1)}(\mathbf{u}_i) = (\mathbf{X}' \boldsymbol{\varphi}(\mathbf{u}_i) \mathbf{V}^{(\ell)}(\mathbf{u}_i) \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\varphi}(\mathbf{u}_i) \mathbf{V}^{(\ell)}(\mathbf{u}_i) \mathbf{y}^{*(\ell)}(\mathbf{u}_i) \quad (4)$$

Where $\beta^{(\ell+1)}(\mathbf{u}_i)$ The vector of the estimated parameters for location i , $(\ell + 1)$ the number of iterations, the parameter vector :-

$$\beta^{(\ell)}(\mathbf{u}_i) = [\beta_0^{(\ell)}(\mathbf{u}_i), \beta_1^{(\ell)}(\mathbf{u}_i), \dots, \beta_j^{(\ell)}(\mathbf{u}_i)]' \quad (5)$$

\mathbf{X}' is the transpose of the matrix \mathbf{X} , $\boldsymbol{\varphi}(\mathbf{u}_i)$ is the spatial weight matrix, $\mathbf{V}^{(\ell)}(\mathbf{u}_i)$ is the variance weights matrix for the location i , $\mathbf{y}^{*(\ell)}(\mathbf{u}_i)$ is the vector of adjusted dependent variables which is given by :-

$$\mathbf{y}^{*(\ell)}(\mathbf{u}_i) = (\mathbf{y}_1^{*(\ell)}(\mathbf{u}_i), \mathbf{y}_2^{*(\ell)}(\mathbf{u}_i), \dots, \mathbf{y}_A^{*(\ell)}(\mathbf{u}_i))' \quad (6)$$

Where :-

$$\mathbf{y}_k^{*(\ell)}(\mathbf{u}_i) = (\beta_0^{(\ell)}(\mathbf{u}_i) + \sum_j \beta_j^{(\ell)}(\mathbf{u}_i) \mathbf{x}_{kj}) + \frac{\mathbf{y}_k - \hat{\mathbf{Y}}_k(\beta^{(\ell)}(\mathbf{u}_i))}{\hat{\mathbf{Y}}_k(\beta^{(\ell)}(\mathbf{u}_i))} \quad (7)$$

$$= \eta_k(\beta^{(\ell)}(\mathbf{u}_i)) + \frac{\mathbf{Y}_k - \hat{\mathbf{Y}}_k(\beta^{(\ell)}(\mathbf{u}_i))}{\hat{\mathbf{Y}}_k(\beta^{(\ell)}(\mathbf{u}_i))} \quad (8)$$

And η_k is a linear predictor of k th observation. The sets of the local parameter can be estimated by repeating the iterative procedure for each location i . The final step in the estimation process is written as follows:-

$$\beta(\mathbf{u}_i) = (\mathbf{X}' \boldsymbol{\varphi}(\mathbf{u}_i) \mathbf{V}(\mathbf{u}_i) \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\varphi}(\mathbf{u}_i) \mathbf{V}(\mathbf{u}_i) \mathbf{y}^*(\mathbf{u}_i) \quad (9)$$

The prediction of each observation k at each regression points i would be :-

$$\hat{\mathbf{Y}}_{ik} = \mathbf{C} \mathbf{y}^*(\mathbf{u}) \quad (10)$$

where the i th row of the matrix \mathbf{C} is:-

$$\mathbf{c}_i = \mathbf{x}_i (\mathbf{X}' \boldsymbol{\varphi}(\mathbf{u}_i) \mathbf{V}(\mathbf{u}_i) \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\varphi}(\mathbf{u}_i) \mathbf{V}(\mathbf{u}_i) \quad (11)$$

2.2 Kernel functions

The kernel function has several names, such as the weight function, the window function, or the shape function, etc., and they are of two types (Wasserman , 2006) :-

1. Optimal kernel functions, which work to reduce the Mean Integrated Square Error (MISE) .

2. Minimum Variance Kernels, which work to reduce the alignment variance .

There are several kernel functions used for weighting in GWPR model including Gaussian and Tri-cube:-

Table 1. kernel functions

Kernel	Equation	
Gaussian	$\frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}}$	$(v < \infty), I$
tri-cube	$\frac{70}{81} (1 - v ^3)^3$	$(v \leq 1), I$

Where :-

$$v = \frac{x - x_i}{h}, h \text{ is bandwidth}$$

3. Discussion of Results

Two geographically weighted Poisson regression models will be evaluated and compared using two kernel functions including gaussian and tri-cube kernel weighting function, to find the best model for representing traffic accident data in Iraq. The traffic accident data in the governorates of Iraq (except the Kurdistan Region) was described, which was obtained from the Ministry of Planning / Central Bureau of Statistics / Directorate of Transport and Communications Statistics for the period (2006-2022) and the program used is R

Table 2. Governorates of Iraq (except for the Kurdistan region) and the number of traffic accidents and their causes for each governorate for the period (2006-2022).

	Governorates	Traffic Accidents	The Road	The Car	The driver	Pedestr -ians	Passeng -ers	Other reasons
1	Baghdad	18148	1704	2964	11628	1367	284	201
2	Al-Basrah	17689	927	1465	12819	1880	583	15
3	Al-Najaf	17223	1809	2517	12130	616	42	109
4	Babil	16079	227	882	14133	745	31	61
5	Al-Qadisiyah	12861	307	1604	9761	776	92	321
6	Dhi Qar	11428	305	699	9909	345	111	59
7	Wasit	11040	1638	2134	5688	1110	210	260
8	Diyala	8774	506	1274	6394	178	135	287
9	Karbala	6769	401	318	5527	412	37	74
10	Al-Muthanna	5627	129	306	5050	110	22	10
11	Kirkuk	4941	117	401	4305	81	7	30
12	Maysan	4402	38	247	3628	50	9	430
13	Ninawa	3750	573	806	1787	478	47	59
14	Salah Al-Din	3626	394	601	2027	452	134	18
15	Al-Anbar	2708	109	172	2326	78	4	19

Table 3 The values of the parameters of the geographically weighted Poisson regression model.

	Gaussian	Tri-cube
β_0	5.3480	5.3208
β_1	0.0022	0.0022
β_2	0.0008	0.0013
β_3	0.0015	0.0016
β_4	0.0017	0.0019
β_5	0.0002	0.0012
β_6	0.0034	- 0.0025

To compare and evaluate the GWPR model performance with two kernel weighting functions, we use corrected Akaike Information Criterion (AICc) which is given (Al-hasani et al , 2021) (Nakaya et al , 2005) :-

$$AICc(h) = D(h) + 2 \left(K(h) + \frac{K(h)(K(h)+1)}{A-K(h)-1} \right) \quad (12)$$

Where $D(h)$ is the standard deviation of the model, K is number of parameters , h is the bandwidth , A is the number of spatial observations . As the model that gives the lowest value for this criterion is considered the best in comparison with the rest of the models.

Table 4 . AICc values for the geographically weighted Poisson regression models estimated by kernel functions

Kernel Functions	Value AICc
Gaussian	5481.17
Tri-cube	2914.04

Table 4 shows that the lowest value for this criterion when we estimate the geographically weighted Poisson regression models using **tri-cube** kernel function with (2914.04) , while the largest value was when we estimate the geographically weighted Poisson regression models using gaussian kernel function with (5481.17).

4. Conclusion

This study demonstrates that, In all governorates, the biggest cause of accidents was the driver. The second cause was the car, except for the governorates of Basra and Karbala, the second cause was pedestrians. In Maysan, the second cause was other reasons. The third cause was between the road, the car, and pedestrians alternately, and the fourth, fifth, and sixth cause was between the road, the car, pedestrians, passengers, and other reasons, according to each governorate . According to the corrected Akaike Information Criterion (AICc), the geographically weighted Poisson regression model estimated using the tri-cube function was better than the same model when the gaussian function was used.

5. Recommendations

1. Using spatial analysis instead of general analysis, as it shows more precise details of the relationship between the dependent variable and the independent variables.
2. The necessity of using the Poisson regression model or the geographically weighted Poisson regression model in studying and analyzing rarely occurring phenomena and events or phenomena measured as rates.
3. Increasing awareness campaigns regarding traffic accidents by the competent authorities through the media and educational institutions, as well as establishing strict controls in granting driving licenses, penalties for speeding, limiting the random import of cars, and working to improve and maintain road networks and provide them with lighting and indicative and warning traffic signs.

References

1. Abbas, H.A. (2019) , "Selecting the best model to predict the number of traffic accidents in the southern governorates of Iraq" Research's Higher Diploma, University of Baghdad - College of Administration and Economics.
2. Al-Hasani , G. , Asaduzzaman , M. & Soliman , A. (2021) , "Geographically Weighted Poisson Regression Models with different Kernel : Application to road traffic accident data " , Journal : Communications in Statistics , Vol. 7 , No. 2 , pp. 166 - 181 .
3. AL-Obaedi , J. T. S. (2015) , " Evaluation of Traffic Accidents Rates in Al -Diwaniya City"Journal of Babylon University - Engineering Sciences , Vol. 23 , No. 3 .
4. Al-Saadi, N.H.J (2017) , " Algorithm predicting seasonal and cyclical rates of change with a practical application of traffic accidents in Iraq",Journal of University of Babylon,Vol. 25,No. 6, pp. 2035 – 2046.
5. Dawod, G.M. (2018) , " Advanced statistical and spatial applications " , Cairo , Egypt .
6. Dhahad , S.N. (2015) " Traffic Accidents in Thi Kar Government Causes and Solutions" , Basic Education College Magazine For Educational and Humanities Sciences, Babylon University, Vol. 20, pp. 639 - 655 .
7. Ewadh , H. A. (2017) , "A Sustainable Strategy To Upgrade Safety At Traffic System " , Journal University Of Kerbala , Vol. 15 , pp. 124 -130.
8. Hadayeghi , A. , Shalaby , A.S. & Persaud , B. (2010) , "Development of Planning Level Transportation Safety Tools Using Geographically Weighted Poisson Regression " , Elsevier, Accident Analysis and prevention. Vol. 42, pp. 676 – 688 .
9. Hezaveh , A.M. , Arvin , R. & Cherry , C.R. (2019) , "Ageographically Weighted Regression to estimate the comprehensive cost of traffic crashes at a zonal level " , Elsevier , Accident Analysis & Prevention , Vol. 131 , pp. 15 – 24 .



10. Jasim , I.A. (2014) , " Evaluation of spatial distribution of educational services using geographical weighted regression , Case study : City of Kut " , The Seventh International Conference of the Faculty of Education - University of Wasit.
11. Nakaya , T. , Fotheringham , A.S. , Brunsdon , C. & Charlton , M. (2005) , "Geographically Weighted Poisson Regression for disease association mapping " , Statistics in Medicine , Vol. 24 , pp. 2695 – 2717 .
12. Wasserman , L. (2006) , " All of Nonparametric Statistics " , Springer Science + Business Media Inc.
13. Zalzal , S.H. Mahmoud, K.K. (2013) " Statistical study of traffic accidents in Iraq for the year (2006)" , Journal of Administration & Economics, University of Babylon, Vol. 36, No. 96, PP. 213-221 .
14. <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>

مستخلص البحث

تعد الإصابات الناجمة عن حوادث المرور السبب الأول للوفاة للأشخاص الذين تتراوح أعمارهم بين 15 إلى 29 عامًا، وذلك وفقًا لتقرير منظمة الصحة العالمية حول الحوادث المرورية الصادر بتاريخ 20-يونيو-2022. ولدراسة وتحليل الحوادث المرورية يفضل استخدام التحليل المكاني لمعرفة الأسباب المؤثرة وحسب كل منطقة. ولذلك تم استخدام نماذج انحدار بواسون الموزون جغرافياً (وهو نموذج محدث للانحدار التقليدي، بحيث يكون التشخيص أكثر دقة للظاهرة المدروسة من خلال إضافة البعد المكاني للنموذج، لذا فإن النموذج الجديد يأخذ في الاعتبار هذه العلاقة (العلاقة المكانية البعد)، أي أن الانحدار الموزون جغرافياً (GWR) ينفذ انحداراً لكل موقع بدلاً من انحدار واحد لمنطقة الدراسة بأكملها) تم استخدامها لهذا الغرض. حيث أن العلاقة الثابتة قد تخفي بعض العوامل المكانية المهمة التي تؤثر على العلاقة بين متغيرات الاستجابة والمتغيرات التفسيرية، ومن ثم قد تؤثر سلباً على دقة النماذج. تم استخدام نموذج انحدار بواسون الموزون جغرافياً (GWPR) في هذه الدراسة على عينة من الحوادث المرورية للمدة من 2006 إلى 2022 لجميع محافظات العراق باستثناء إقليم كردستان لغرض القيام بأعمال مسبقة لتحديد المواقع التي قد تزيد من حوادث المرور. كما تم استخدام دالتين (kernel (gaussian and tri-cube). أظهرت الدراسة أن دالة (tri-cube) كانت الأفضل وفق معيار المعلومات أكايكي المصحح (AICc).

نوع البحث : ورقة بحثية

المصطلحات الرئيسية للبحث : الحوادث المرورية ، انحدار بواسون الموزون جغرافياً" ، Gaussian ، Tri-cube ،
corrected Akaike information criterion (AICc) .

• البحث مستل من رسالة ماجستير