GENERALIZED LINEAR AND GENERALIZED ADDITIVE MODELS IN STUDIES OF MOTORCYCLE ACCIDENT PREDICTION MODELS FOR THE NORTH-SOUTH ROAD CORRIDOR IN SURABAYA

Machsus
Doctor Candidate in Civil Engineering Dept, University of Brawijaya, Malang, and Lecture at Dept. of Civil Engineering, Institut Teknologi Sepuluh Nopember Surabaya,
Jln. Menur 127, Kampus ITS Manyar, Surabaya, 65116,
Telp: (031) 5947637
machsus@ce.its.ac.id

Harnen Sulistio
Professor on Civil Engineering Department, University of Brawijaya, Malang, Indonesia
Jln. MT. Haryono 16, Kampus UB Malang, 65145,
Telp: (0341) 580120
harnen@ub.ac.id

Achmad Wicaksono
Associate Professor on Civil Engineering Department, University of Brawijaya, Malang, Indonesia
Jln. MT. Haryono 16, Kampus UB Malang, 65145,
Telp: (0341) 580120
wicaksono@ub.ac.id

Ludfi Djakfar
Associate Professor on Civil Engineering Department, University of Brawijaya, Malang, Indonesia
Jln. MT. Haryono 16, Kampus UB Malang, 65145,
Telp: (0341) 580120
ldjakfar@ub.ac.id

Abstract

One of the advances in the development of a model of traffic accidents is indicated by the availability of generalized linear models (GLMs) and generalized additive models (GAMs) in the regression analysis. This paper will discuss the motorcycle accident prediction models using GLMs and GAMs on the north-south road corridor in Surabaya. The first part will discuss the model prediction of traffic accidents, as well as providing a brief review related to the use GLMs and GAMs in building models of accidents. Furthermore, application examples of GLMs and GAMs will be presented. To determine the effect of non-linear in each explanatory variable, smoothing the GAMs will be conducted in each variable gradually. Model diagnostic and intrepretations will be done in the final part. The results of the application of GLMs and GAMs indicates that the development of predictive models of motorcycle accidents with statistical methods can be used to diagnose problems in road safety. GAMs produces better models than the GLMs in which its condition without using the Poisson distribution, as shown in the difference in the value of the model parameter R-sq.(Adj), deviance explained, and the GCV score. By using the Poisson distribution with a log link-function, it appears that GLMs and GAMs produce the same model parameter values.

Key Words: *Regression Analysis, Generalized linear models, Generalized additive models, Motorcycle accident prediction models, Urban roads, Surabaya.*

Abstrak

Salah satu kemajuan dalam pengembangan model kecelakaan lalu lintas ditunjukkan dengan tersedianya generalized linear models (GLMs) dan generalized additive models (GAMs) dalam analisis regresi. **Objective:** Pada paper ini akan dibahas pembuatan model prediksi kecelakaan sepeda motor dengan menggunakan GLMs dan GAMs pada ruas jalan koridor utara-selatan di Surabaya. Pada bagian awal akan dibahas model prediksi kecelakaan lalu lintas, serta memberikan review singkat terkait penggunaan GLMs dan GAMs dalam upaya pemodelan kecelakaan. Selanjutnya akan disajikan contoh aplikasi GLMs dan GAMs. Untuk mengetahui pengaruh non-linier pada tiap-tiap variable penjelas, maka pemulusan pada GAMs dilakukan secara bertahap dan dilakukan pada setiap variable. Pada bagian akhir dilakukan diagnostik model dan interpretasi. Hasil dari aplikasi GLMs and GAMs pada model prediksi kecelakaan sepeda motor menunjukkan bahwa pengembangan model prediksi kecelakaan sepeda motor dengan metode statistik dapat digunakan untuk mendiagnosis masalah keselamatan di jalan. GAMs menghasilkan model lebih baik dibandingkan dengan GLMs pada kondisi tanpa menggunakan distribusi poisson, seperti yang tampak pada perbedaan nilai parameter model R-sq.(adj), deviance explained, dan GCV score. Dengan menggunakan distribusi poisson dengan link-function log, ternyata GLMs dan GAMs menghasilkan nilai parameter model yang sama.

Kata Kunci: *Analisis regresi, Generalized linear models, Generalized additive models, Model prediksi kecelakaan sepeda motor, Ruas jalan perkotaan, Surabaya.*

INTRODUCTION

The study of traffic accidents prediction model continues to develope. Several previous studies that investigated the traffic accident on the road, including: Xie and Zhang (2008) who researched the roads in Toronto, Canada; Sobri (2010) researched a motorcycle accident on the road in Malang; Ackaah and Salifu (2011) who researched the roads in Ghana; and Polus A. and Cohen M. (2011) who researched the roads in Israel; Li, XG, Lord, D., & Zhang, Y. (2011) who evaluated the frontage road in Texas, as well as Machsus et al. (2013) who researched the arterial roads in Surabaya.

Generalized Linear Models (GLMs) is used in the formation of an accident prediction model with data that are not assumed to be normally distributed. GLMs is able to describe the characteristics of traffic accidents, but this method is still considered to have limitations on the assumption of a linear relationship in the explanatory variables used (Machsus, et al., 2013). Generalized Additive Models (GAMs) is introduced to overcome these limitations. GAMs has the ability related to the non-linear relationship in the explanatory variables, and the result of statistical interpretation. So, GAMs offer form of functionality which is more flexible than GLMs in the regression analysis.

The aim of this research to examine the application of GLMs and GAMs in studies of motorcycle accident prediction models on urban roads. This research is to develop a predictive model study of traffic accidents, especially studies related to motorcycle accidents on the north-south road corridor in Surabaya.

A LITERATURE REVIEW ON ACCIDENT PREDICTION MODELS

Accident prediction models are used to estimate the frequency of traffic accidents. In addition, it can also be used to identify and determine the relationship among the factors affect such as: geometric, environmental, and operational from the road segments (BB Nambuusi., 2008; Chengye P., and P. Ranjitkar, 2013).

Originally, conventional linear regression approach is often used in the accident prediction modeling. In this approach, the accident data is assumed to be normally distributed with constant variants. But this theoretical distribution assumption cannot represent the distribution of traffic accident data well. Characteristics of traffic accident data are not able to be represented by a normal distribution, especially time distribution and place of occurrence (Taylor et al., 2002; Harnen S. et al., 2003, 2004, 2005, 2006, 2010).

Finally, the conventional linear regression approach is rarely used in the modeling of traffic accidents. Furthermore, GLMs and GAMs approach is used in the establishment of an accident prediction modeling. Both of these approaches are considered to have advantages compared with conventional linear regression.

Accident Prediction Models Using GLMs

GLMs method is the development of a linear model, which contains linear predictor component, exponential family distribution and Link-functions. GLMs development was pioneered by Nelder and Wedderburn (1972). GLMs model was developed to resolve existence of irregularities in the variance, or the response does not follow a normal distribution. This linear model uses the assumption that the response has exponential family distributions. Exponential family distribution is the distribution that is more general, such as: the Poisson, negative binomial, and the others (Nelder and Wedderburn, 1972; Mc Cullagh and Nelder, 1989).

GLMs approach used in the modeling of traffic accidents in which in this approach, the data of traffic accidents was no longer assumed to be normally distributed. Distribution which is often used in the modeling of traffic accidents in the previous studies is the Poisson and the negative binomial distribution. Besides the two types of distribution, other distributions are also used: the geometric and logarithmic distribution (Harnen S. et al., 2003, 2004, 2006, 2010; Polus and Cohen, 2011).

The GLMs method with the assumed distribution is able to describe events randomly, discretely, and non-negatively, which are the characteristic of a traffic accident, but this quite popular method is still considered to have limitations. GLMs limitation lies in the assumption of linear relationships among the variables used in the modeling. If the predictive variable relationships with several explanatory variables on the data of traffic accident are non-linear pattern, but assumed to be linear, then the value of the significance of the results of the regression analysis would be reduced. Consequently, the resulting prediction model becomes less realistic (Xie and Zhang, 2008; Li, Lord, & Zhang, 2011).

Accident Prediction Models Using GAMs

GAMs first developed by Hastie and Tibshirani in 1986. GAMs is the new types of modelling which was introduced in the statistics community to create a model of the observed data (Hastie and Tibshirani, 1990; Wood, 2006, 2003). Accident prediction models are mostly based on the assumption of a linear relationship between the predicted frequency of accidents and some of the explanatory variables, though not all of them are linear. If the predictive variable relationships with several explanatory variables on the data of traffic accident are non-linear pattern but assumed to be linear, then the value of the significance of the results of the regression analysis would be reduced, so that the resulting prediction model becomes less realistic.

To resolve the problem of limited GLMs the usage of neural network and support vector machine, which has a strong ability in a non-linear approach, and does not require a specific form of the function, are proposed. However, both methods are criticized for not presenting an explicit functional relationship and the results of statistical interpretation.

Therefore, GAMs was introduced as a new modeling method which has a strong ability to the non-linear approach and it can also show the results of the statistical interpretation. This method offers a more flexible functional form than generalized linear models (GLMs) and it allows to be more adaptive to variable interactions (Xie and Zhang, 2008; Li, Lord, & Zhang, 2011).

GAMs is a statistical method to determine the non-linear relationship between the response variable and the prediction variable (Hastie and Tibshirani, 1990). GAMs is an extension of the additive model by modeling the response variable, as an additive combination of univariate functions of the independent variables. This method can directly accommodate the existence the non-linear influence of the independent variables without having to know the explicit form of influence well (Hastie and Tibshirani, 1990). Thus, in a predictive modeling GAMs is better than the linear regression model when there is evidence of non-linear effect of the independent variables.

GAMs method has also been used in predictive modeling of traffic accidents. Xie and Zhang (2008), was the first to introduce GAMs in predicting the frequency of road accidents. His research results show that the GAMS provide a better ability to the non-linear approach than GLMs that retains the basic framework of GLMs.

According to Li, Lord, & Zhang (2011), although GAMs is more flexible than GLMs but it still has some limitations. Firstly, GAMs covers more parameters, therefore the estimated coefficients can be very complex, especially when the default values in the statistical software package is not used. Secondly, GAMs using smoothing Spline functions, so that it is possible that prediction coefficients cannot clearly presented or defined. Thirdly, the modeling results between GAMs and GLMs tend to be similar if the covariates are completely independent and dependent variables have a linear or exponential relationship with covariates.

GAMs use a smoothing function on each explanatory variable and it is very flexible in modeling the non-linear relationship (Xie and Zhang, 2008). Smoothing technique was first suggested by Ezekiel in 1941. Smoothing is basically a process which systematically can eliminate rough data pattern or fluctuate and can take data pattern that can be described in general. Non-parametric smoothing technique is used to model the relationship among the variables without assigning specific form of the regression function (Hastie and Tibshirani, 1990).

Smoothing is one of important steps in the estimation of GAMs. This process is necessary to predict the function of the independent variable. Smoother is a tool for summarizing the trend of the response as a function of one or more independent variables (Hestie and Tibshirani, 1990). Smoother does not require strong assumptions about the relationship form of the response variable with independent variables. Therefore, smoother is better known as a tool in a non-parametric regression (Hastie and Tibshirani, 1990).

Smoother has two functions: (1) to description, that a smoother can be used to improve the visual appearance of a scatter diagram of the variable response to the independent variable, so the tendency of the plot can be more precisely determined; and; (2) to estimate the dependence of a response variable to the independent variables (Hastie & Tibshirani, 1990).

MATERIALS AND METHODS

Motorcycle accident data are reviewed are that occurred on the north-south road corridor in Surabaya, from Jl. Tanjung Perak to Jl. Achmad Yani. The road corridor is planned to be used as a route for the development of mass transit in Surabaya. The data collected include: number of motorcycle accidents, traffic volume, length of roads, and traffic speed, as in depicted in Table 1 below.

These motorcycle accident data were obtained from the Traffic Accident Unit, Surabaya Police, Indonesian National Police for 2009 to 2012. The traffic volume; vehicle speed, and geometric road were acquired from Surabaya Government Agencies, including: the Transportation Department, Department of Highways, Planning and Urban Development.

Table 1: Motorcycle accidents on the north-south road corridor in Surabaya

No	Road Names	Motorcycle Accident (McA)	Length of Road (LR)	Flow (FLOW)	Speed (SPEED)
1	Jl. Tanjung Perak	10.00	3,918.00	4480	44
2	Jl. Rajawali	3.50	1,180.00	4574	41
3	Jl. Jembatan Merah	0.50	700.00	2225	36
4	Jl. Veteran	1.25	738.00	2720	39
5	Jl. Pahlawan	1.50	1,220.00	5219	39
6	Jl. Kramat Gantung	0.25	600.00	1403	30
7	Jl. Gemblongan	1.25	400.00	3181	45.5
8	Jl. Tunjungan	2.00	910.00	4596	41
9	Jl. Gubernur Suryo	2.50	563.70	5226	45
10	Jl. PB. Sudirman	2.50	2,100.00	5723	36
11	Jl. Urip Sumoharjo	3.00	968.00	8595	38.5
12	Jl. Raya Darmo	23.75	4,056.00	8749	53
13	Jl. Wonokromo	11.25	1,160.00	10338	51
14	Jl. Achmad Yani	54.25	5,835.00	12565	55
15	Jl. Basuki Rahmat	8.50	1,229.00	6980	52.5
16	Jl. Embong Malang	3.75	770.00	6003	42.5
17	Jl. Blauran	2.25	276.00	4022	42
18	Jl. Bubutan	2.50	2,496.00	3482	41
19	Jl. Indrapura	8.75	2,847.00	3448	52

Source: *Traffic Accident Unit, Surabaya Police, Indonesian National Police*

In this study, two approaches are used: GLMs and GAMs approach. Algorithm flowchart of GLMS and GAMs application can be seen in Figure 1 below. The results of the application of the two approaches are compared, to be known the advantages and disadvantages. For the establishment of accident prediction models, the R software package is used. The R software, is software that is distributed as open source software so that it can be obtained and used for free and it is opens to be modified and developed continuously.

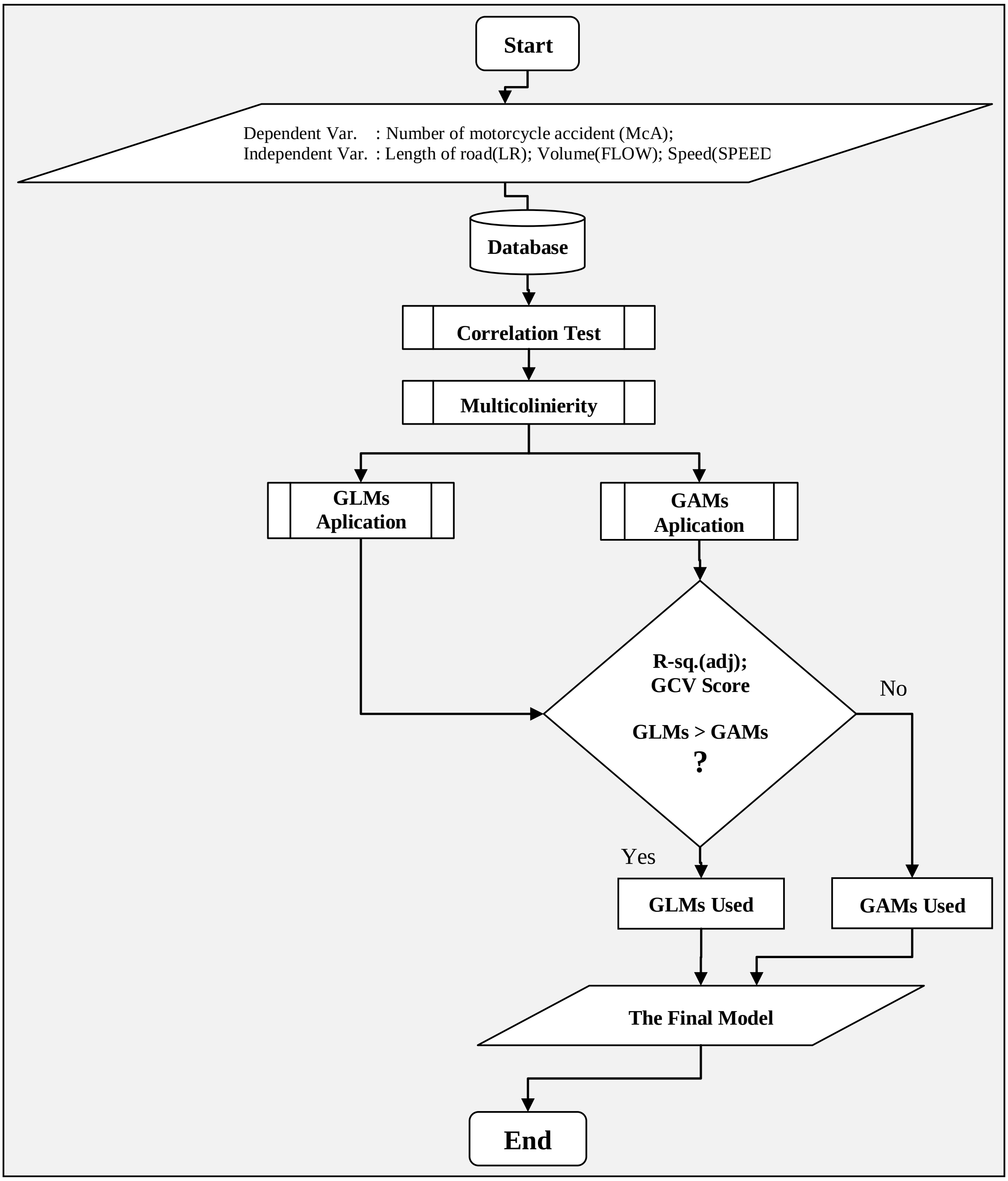

Figure 1. Algorithm flowchart of GLMS and GAMs applications

RESULTS AND DISCUSSION

An Example of GLM Output

Below is an example of GLM output using the application of the R software package on motorcycle accident prediction models.

```
## Call:
## glm(formula = McA ~ Ln_LR + Ln_FLOW + SPEED, family = poisson(link = log), data = data)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -0.76108  -0.31188  -0.08503   0.20571   1.05877
##
## Coefficients:
##              Estimate  Std. Error  z value   Pr(>|z|)
## (Intercept) -13.69966    1.71994   -7.965  1.65e-15 ***
## Ln_LR         0.65763    0.14909    4.411  1.03e-05 ***
## Ln_FLOW       0.81236    0.25929    3.133  0.001730 **
## SPEED         0.07792    0.02319    3.361  0.000778 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 226.2281  on 18  degrees of freedom
## Residual deviance:   4.1514  on 15  degrees of freedom
## AIC: Inf
##
## Number of Fisher Scoring iterations: 4
##
## Family: poisson
## Link function: log
##
## Formula:
## McA ~ Ln_LR + Ln_FLOW + SPEED
##
## Parametric coefficients:
##              Estimate  Std. Error  z value   Pr(>|z|)
## (Intercept) -13.69966    1.71994   -7.965  1.65e-15 ***
## Ln_LR         0.65763    0.14909    4.411  1.03e-05 ***
## Ln_FLOW       0.81236    0.25929    3.133  0.001730 **
## SPEED         0.07792    0.02319    3.361  0.000778 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =   0.99   Deviance explained = 98.2%
## UBRE score = -0.36046  Scale est. = 1        n = 19
```

Based on the GLM output, motorcycle accident prediction models can be formulated, as follows:

$$\mathrm{Ln(McA)} = \mathrm{Ln(k)} + \boldsymbol{\alpha_1} \mathbf{Ln}(\mathbf{LR}) + \boldsymbol{\alpha_2} \mathbf{Ln}(\mathbf{FLOW}) + \boldsymbol{\beta_1}(\mathbf{SPEED}) \quad (1)$$

k = antilog(-13.69966)
antilog in formula =power(2.718281828459, number)

$$Ln(McA) = ln(k)+0.65763\ ln(LR)+0.81236\ ln(FLOW)+0.07792(SPEED) \quad (2)$$

$$McA = k\ \mathbf{LR}^{\alpha_1}\mathbf{FLOW}^{\alpha_2}\ e^{\beta_1 \mathbf{SPEED}} \quad (3)$$

$$McA = 0{,}00000112\ \mathbf{LR}^{\mathbf{0.65763}}\mathbf{FLOW}^{\mathbf{0.81236}}e^{\mathbf{0.07792\ SPEED}} \quad (4)$$

remark:
MCA = the number of motorcycle accidents per year
FLOW = the traffic volume (pcu / hour)
LR = the length of roads (meters)
SPEED = the 85 percentile vehicle speed (km / hour)

An Example of GAMs Output

Below is an example of GAMs output using the application of the R software package on motorcycle accident prediction models.

```
## Family: gaussian
## Link function: identity
##
## Formula:
## McA ~ Ln_LR + Ln_FLOW + s(SPEED)
##
## Parametric coefficients:
##                Estimate  Std. Error  t value  Pr(>|t|)
## (Intercept)    -32.426     16.448   -1.971   0.0814 .
## Ln_LR            2.355      1.121    2.100   0.0664 .
## Ln_FLOW          2.745      1.824    1.505   0.1680
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                edf   Ref.df      F    p-value
## s(SPEED)  7.343    8.086 20.77  2.07e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.958   Deviance explained =   98%
## GCV score = 14.532  Scale est. = 6.6214    n = 19
```

Based on the GAMs output, motorcycle accident prediction models can be formulated as follows:

$$Ln(McA) = Ln(k)+\boldsymbol{\alpha_1}\mathbf{Ln}(\mathbf{LR})+\boldsymbol{\alpha_2}\mathbf{Ln}(\mathbf{FLOW})+s(SPEED) \quad (5)$$

k = antilog(-32.426)
antilog in formula =power(2.718281828459, number)

$$Ln(McA) = ln(k) + 2.355\ ln(LR) + 2.745\ ln(FLOW) + s(SPEED) \quad (6)$$

$$McA = k\ \mathbf{LR}^{\alpha_1}\mathbf{FLOW}^{\alpha_2}\ e^{s(\mathbf{SPEED})} \quad (7)$$

$$\text{McA} = 0{,}00000827\ \mathbf{LR}^{2.355}\mathbf{FLOW}^{2.745}\boldsymbol{e}^{s(\mathbf{SPEED})} \quad (8)$$

remark:
MCA = the number of motorcycle accidents per year
FLOW = the traffic volume (pcu / hour)
LR = the length of roads (meters)
SPEED = the 85 percentile vehicle speed (km / hour)

To determine the effect of non-linear in each explanatory variable, then the smoothing is done in stages and performed on each variable. Furthermore, smoothing is done on more than one independent variable, or on all the variables, especially the independent variable that has a quite large a non-linear effect.

Model Comparison

The diagnosis of accident prediction models resulting from the application of GLMs and GAMs can be done by observing the changes in the model parameter values. Table 2 shows the values of R-sq.(Adj), deviance explained, GCV score and UBRE score, both in the condition of without and with using the Poisson distribution.

In conditions without using the Poisson distribution or following the default distribution on R software, different model parameter values between GLMs and GAMs are obtained. For example, in the parameters of R-sq.(Adj) the value obtained is 0.554 for GLM, and the value range of 0925-0964 for GAMs. The same thing also applies to the parameters of deviance explained and the GCV score. This means that the motorcycle accident prediction models generated using GAMs is better than using GLMs.

Table 2: Model Comparison between Using GLM and GAMs

Model Parameters	**GLM**	**GAMs with Spline Smoothing**		
		SPEED	FLOW	LR
Without Using The Poisson Distribution				
R-sq.(adj)	0.554	0.958	0.925	0.964
Deviance explained	62.90%	98.00%	96.60%	98.10%
GCV score	90.051	14.532	28.323	11.271
By Using The Poisson Distribution				
R-sq.(adj)	0.99	0.99	0.99	0.99
Deviance explained	98.20%	98.20%	98.20%	98.20%
UBRE score	-0.3605	-0.3605	-0.3605	-0.3605

Meanwhile for the condition by using the Poisson distribution on the execution of the R software the same model parameters value is obtained, both for the GLM and GAMs. For instance, in the parameters of the R-sq.(Adj) the value obtained is 0.99 for the GLM and GAMs. The same thing also applies to the parameters of deviance explained and UBRE score. After using the Poisson distribution with a log link-function, motorcycle accident prediction models generated using GAMs turned out to be no better than using GLM, or the result is the same. The result of this study confirmed the findings in the previous studies (Xie and Zhang, 2008; Li, Lord, & Zhang, 2011).

The results of the GLM and GAMs application are interesting to be discussed because of both approaches yield the same values of model parameters when Poisson distribution with a log link-function is used. This occurs because the use of the Poisson distribution with a log link-function is expected to reduce the influence of non-linear independent variables. Moreover, it also shows that the distribution of motorcycle accident data in this case follows the Poisson distribution. Clearly, non-linear effect of independent variables is reduced or gone, because the choice of distribution used is in accordance with the data distribution. The findings of this study contradict the findings of previous studies (Xie and Zhang, 2008; Li, Lord, & Zhang, 2011).

CONCLUSION

Based on the above discussion, it can be concluded that:

1. Motorcycle accident prediction models generated by using GAMs is better than using GLM in which the conditions is without using the Poisson distribution as indicated by the differences in the model parameter values for the R-sq.(Adj), deviance explained, and the GCV score.
2. After a Poisson distribution with a log link-function is used in the model building process motorcycle accident using GLM and GAMs, it appears that both produce the same model parameter values.
3. In the model building process, non-linear effect on the independent variables can be reduced or gone because of the selection of the Poisson distribution with a log link-function is compatible with the distribution of motorcycle accident data in this case.

REFERENCES

Ackaah W. and Salifu. M. 2011. Crash prediction model for two-lane rural highways in the Ashanti region of Ghana. IATSS Research, 35 (2011) pp 34-40.

Chengye P., and Ranjitkar P., 2013. Modelling Motorway Accidents using Negative Binomial Regression, Proceedings of the Eastern Asia Society for Transportation Studies, Vol. 9.

Harnen, S. Radin Umar R.S. Wong S.V. and Wan Hasim W.I, 2006. Motorcycle Accident Prediction Models for Junctions on Urban Road in Malaysia. Journal of Advances in Transportation Studies, Section A 8. page 31-40

Hastie, T. J., & Tibshirani, R. J. 1990. Generalized Additive Models. New York: Champman and Hall.

J.A. Nelder and R.W.M. Wedderburn. 1972. Generalized linear models. J.R.Statist.Soc., 57:359–407.

Li, X. G., Lord, D., & Zhang, Y. 2011. Development of Accident Modification Factors for Rural Frontage Road Segments in Texas Using Results from Generalized Additive Models. ASCE Journal of Transportation Engineering, 137(1), 74–83.

Li, X. G., Lord, D., Zhang, Y., & Xie, Y. C. 2008. Predicting Motor Vehicle Crashes Using Support Vector Machine Models. Accident Analysis and Prevention, 40(4), 1611–1618.

Machsus, Harnen S., Wicaksono A., and Djakfar L., 2013. The Prediction Models of Motorcycle Accidents on Surabaya Arterial Roads Using Generalized Linear Models. Middle-East J. Sci. Res., 18 (12): 1859-1866, 2013. ISSN 1990-9233, © IDOSI Publications.

McCullagh, P. and J.A. Nelder. 1989. Generalized Linear Models, 2nd edition. London, England: Chapman and Hall.

Nambuusi et.al. 2008. A review of accident prediction models for road intersections. Steunpunt Mobiliteit & Openbare Werken – Spoor Verkeersveiligheid, mei 2008. RA-MOW-2008-004.

Polus A. and Cohen M. 2011. A new, non-canonical Poisson regression model for the prediction of crashes on low-volume rural roads. IATSS Research, Volume 35, Issue 2, March 2012, pages 98-103.

Sobri, A., dkk. 2010. Model Kecelakaan Sepeda Motor pada Ruas Jalan dengan Menggunakan Pendekatan GLM, Jurnal Transportasi FSTPT, ISSN 1411-2442, Vol. 10 No. 2 Agustus 2010.

Taylor, M.C., A. Baruya, and J.V. Kennedy. 2002. The Relationship between Speed and Accidents on Rural Single-Carriageway Roads, Report TRL 511. Berkshire, England: TRL Limited.

Wood, S. N. 2006. Generalized Additive Models: An Introduction with R. Boca Raton, FL: Chapman and Hall/CRC.

Xie, Y., & Zhang, Y. 2008. Crash Frequency Analysis Using Generalized Additive Models. Transportation Research Record, 2061,39–45.