Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

mobil.TUM 2018 "Urban Mobility – Shaping the Future Together" - International Scientific Conference on Mobility and Transport

Utilizing Real-time Traffic and Weather Data to Explore Crash Frequency on Urban Motorways: A Cusp Catastrophe Approach

Athanasios Theofilatos[a,*]

[a]*National Technical University of Athens, Irroon Polytechneiou str 8, Athens, GR15773, Greece*

Abstract

The investigation of crash frequency with freeway traffic and weather data has recently received significant attention by researchers. This paper extends previous research by proposing nonlinear models for modeling crash frequencies which incorporate real-time traffic and weather data collected from an urban motorway in Athens, Greece. Cusp catastrophe theory is applied and compared with traditional statistical models such as the negative binomial model. The results of crash frequency models provide evidence of the potential applicability of the cusp catastrophe theory to road safety, however it seems that linearity of the system is preserved. Hence, traditional models such as the negative binomial model are proved equally capable of describing the underlying phenomenon, even though the goodness-of-fit is not as good as that of the cusp model. Therefore, the explanation of crash frequency phenomenon only with nonlinear model can be supported. A number of interesting findings have also been disclosed. Firstly, is that rainfall intensity has a strong linear impact on crashes (high rainfall intensity causes more crashes). On the other hand, average flow is indicated to have a strong non-linear relationship with crash frequency. Finally, more research is needed to further understand the applicability of cusp catastrophe theory in road safety.

Peer-review under responsibility of the scientific committee of the mobil.TUM 2018 conference.

Keywords: Crash frequency; cusp catastrophe theory; urban motorway; real-time data

1. Introduction and background

There have been numerous studies examining the impact of various factors (geometrical, traffic etc.) on the frequency of crashes (that is the number of expected crashes on a specific segment) over some period of time (Poch

* Corresponding author. Tel.: +30-210-7721575; fax: +30-210-7721454.
E-mail address: atheofil@central.ntua.gr

2352-1465 © 2018 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the scientific committee of the mobil.TUM 2018 conference.

and Mannering, 1996; Abdel-Aty and Radwan, 2000; Savolainen and Tarko, 2005; Anastasopoulos et al. 2008, Anastasopoulos and Mannering, 2009; Anastasopoulos et al., 2012a and 2012b).

Considerable efforts have been made recently to investigate road safety by incorporating weather and traffic characteristics measured and recorded on a real-time basis and also to address the shortcomings of past literature which mainly used aggregated measures of traffic flow and speed. In that context, most of recent relevant studies aim to assess real-time safety of major freeways by utilizing real-time data (Ahmed et al., 2011; Yu et al., 2013; Yu and Abdel-Aty, 2013a and 2013b). This is achieved through the identification of crash precursors, such as traffic volumes, speed variation between lanes, variation in the traffic volume etc., in order to predict a crash occurrence or to predict how severe it would be given that it occurs.

While several studies in the literature predict the impact of congestion on accident frequencies and rates (Lee et al 2002; Noland and Quddus, 2005) the impact of congestion on other incident types is barely explored. A typical approach in the accident literature is to use very high aggregated traffic flow parameters such as average annual daily traffic (AADT), posted speed limits, etc. (Park and Abdel-Aty, 2011) as proxies for traffic congestion.

However, studies having a focus on crash frequency with real-time data has received far less attention from researchers and scholars. Another gap of knowledge often observed in crash frequency studies is the use of statistical methods. More specifically, authors rely on more traditional and "safe" methods such as the well-known Poisson or Negative binomial regression, either fixed-or –random effects. Other methods such as the Tobit model have also been applied (Anastasopoulos et al., 2008; Anastasopoulos et al., 2012b). Bayesian methods have utilized as well (Yu et al., 2013). Even though these methods are generally able to provide reasonable statistical fit, it is not certain whether the underline phenomena are adequately explained. This is because in such type of models the logarithm of the outcome is predicted with a linear combination of the predictors.

Consequently, more complex relationships (e.g. non-linear) cannot be capture by such statistical models and other modeling approaches may have to be sought. One example is the cusp catastrophe theory. This method is entirely different from the existing classical statistical methodology, as cusp catastrophe theory investigates the potential existence of strong non-linearity in the system and explains sudden transitions between states of a dynamic system due to small changes in the input parameters.

For that reason, the aim of the present research is to add to current knowledge by utilizing real-time traffic and weather data on an urban motorway for modeling crash frequency. The urban motorway Attica Tollway ('Attiki odos') in Greater Athens Area in Greece was considered. Moreover, the cusp catastrophe model is applied in order to investigate whether strong non-linear relationships exists between the number of crashes and the independent variables and more specifically if small changes in traffic and weather parameters could dramatically increase or decrease crash numbers. For reasons of comparison, traditional statistical models such as the Negative Binomial model are also considered and applied.

2. Data preparation

Three datasets were used in this study; one dataset with crash data, one with traffic data and one with weather data. The required crash data for Attica Tollway was extracted from the Greek crash database SANTRA provided by the Department of Transportation Planning and Engineering of the National Technical University of Athens. The crash dataset consists of 107 crash cases for the period 2006 to 2011 and for the route "Eleutherios Venizelos" airport to Elefsina. Traffic data for the Attica Tollway were collected from the Traffic Management and Motorway Maintenance. Inductive loops (sensors), placed every 500 meters inside the asphalt pavement of the open sections of the motorway and every 60 meters inside tunnels, are used to provide information regarding the various traffic parameters measured in 5-minute intervals.

Traffic flow (5-min veh per lane), occupancy (%) and speed (km/h) were considered for this study and each crash was assigned to the closest upstream loop detector. The raw 5-min intervals were aggregated to 15-min, 30-min and 1-hour prior to the crash in order to obtain averages and standard deviations. Time intervals prior to the crash occurrence were considered. As for the weather data, the weather records for each meteorological station covered the whole period from 2006-2011. Consequently, each crash case had to be assigned to the closest meteorological station and then the relevant weather data had to be extracted. The parameters that were considered were the temperature and rainfall. The 10-min raw data were aggregated in order to obtain averages and standard deviations for 1-hour, 2-hours, 6-hours and 12-hours prior to the time of each crash occurrence.

3. Methods of analysis

3.1. Cusp catastrophe model

This section is dedicated to a brief description of the methodological background of cusp catastrophe theory. For a more detailed theoretical description, the reader is encouraged to refer to (Park and Abdel-Aty, 2011; Theofilatos et al., 2017). The cusp catastrophe model is capable of handling complex linear and nonlinear relationships simultaneously by applying a high-order probability density function. This density function can replicate sudden behavior jumps and transitions. The cusp equilibrium surface may be considered as a response surface, where depending on the values of variables α and β, its height predicts the value of the dependent variable. Moreover, the dependent variable y have not necessarily to be observed (i.e. being an observed quantity), but it is rather a canonical variable depending on a number of measured dependent variables. The control variables α and β are canonical (latent) as well and depend on a number of actual measured independent variables.

Catastrophe theory examines the qualitative changes in the behaviour of systems when the control factors that influence their behavioural state face smooth and gradual changes. In other words, the catastrophe theory assumes the existence of a dynamic system and explains the sudden transition between the system states, when small changes in the parameters of the system (known as α and β) take place. The term “catastrophe” may be confusing but as it has nothing to do with the consequences of the event. In mathematical sciences, the term catastrophe implies a nonlinear transition from one state to another.

The variable β is the bifurcation factor and α is the asymmetry factor. The asymmetry factor governs how close the system is to a sudden discontinuous change of events, while the bifurcation factor governs how large a change will take place. Now let’s assume a set of measured dependent variables $Y_1, Y_2, \dots, Y_n$, then:

$$y = w_0 + w_1Y_1 + w_2Y_2 + \cdots + w_nY_n. \tag{1}$$

Similarly, if a set of measured independent variables $X_1, X_2, \dots, X_n$ is considered the control factors α and β can be estimated as:

$$\alpha = \alpha_0 + a_1X_1 + \alpha_2X_2 + \cdots + \alpha_nX_n, \tag{2}$$

$$\beta = b_0 + b_1X_1 + b_2X_2 + \cdots + b_nX_n. \tag{3}$$

The cusp catastrophe model is capable of handling complex linear and nonlinear relationships simultaneously by applying a high-order probability density function. This density function can replicate sudden behavior jumps and transitions. Overall, the cusp equilibrium surface may be considered as a response surface, where depending on the values of variables α and β, its height predicts the value of the dependent variable. Moreover, the dependent variable y have not necessarily to be observed (i.e. being an observed quantity), but it is rather a canonical variable depending on a number of measured dependent variables.

In order to assess the fit of the cusp model, a set of diagnostic tools have been suggested such as the pseudo-R2 (Cobb, 1998), the well-known AIC (Akaike, 1974) and BIC (Schwarz, 1978). One alternative diagnostic measure is the comparison of the cusp model with a nonlinear logistic model (Hartelman, 1997; Van der Maas et al., 2003):

$$y = \frac{1}{1+\exp(-\frac{\alpha}{\beta})} + \varepsilon \tag{4}$$

where the parameters (x, α, β) were defined previously and ε is the random disturbance.

3.2. Negative binomial model

Negative binomial regression links the dependent variable with various independent predictor variables, as shown in equation (5). If Y is the dependent variable, predicted by the independent predictor variables x_i, then the variables are connected through the following relationship (Washington et al., 2010):

$$Y = exp(b_i x_i + \varepsilon_i) \tag{5}$$

Where ε_i is the error term. Results are given by the coefficients b_i of the independent predictor variables x_i and the constant term a as shown in equation (6).

$$log(Y) = a + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n \tag{6}$$

4. Results

Table 1 illustrates the descriptive statistics of the dependent variable (total number of crashes per road segments) along with all candidate independent real-time traffic and variables.

Table 1. Summary statistics of crash related variables (cusp model).

Variable	Mean	Std. Deviation
Total crashes per segment	3.567	1.8134
Q_avg_15m_upstream	56.110	24.514
Q_stdev_15m_upstream	4.644	2.505
Q_avg_30m_upstream	55.678	24.433
Q_stdev_30m_upstream	5.107	1.937
Q_avg_1h_upstream	55.160	23.885
Q_stdev_1h_upstream	5.815	2.265
V_avg_15m_upstream	104.469	10.822
V_stdev_15m_upstream	3.075	1.738
V_avg_30m_upstream	104.973	10.856
V_stdev_30m_upstream	3.117	1.240
V_avg_1h_upstream	105.441	10.221
V_stdev_1h_upstream	3.358	1.254
Occ_avg_15m_upstream	0.036	0.020
Occ_stdev_15m_upstream	0.004	0.002
Occ_avg_30m_upstream	0.035	0.020
Occ_stdev_30m_upstream	0.004	0.002
Occ_avg_1h_upstream	0.034	0.019
Occ_stdev_1h_upstream	0.004	0.003
T_1h_avg	20.207	5.293
T_1h_stdev	0.378	0.203
T_2h_avg	20.137	5.191

T_2h_stdev	0.647	0.395
T_6h_avg	19.885	5.060
T_6h_stdev	1.369	0.572
T_12h_avg	19.364	4.971
T_12h_stdev	2.124	0.590
Rain_1h_sum	0.145	0.516
Rain_1h_stdev	0.027	0.094
Rain_2h_sum	0.221	0.637
Rain_2h_stdev	0.028	0.080
Rain_6h_sum	0.241	0.649
Rain_6h_stdev	0.020	0.053
Rain_12h_sum	0.686	1.779
Rain_12h_stdev	0.030	0.068

All independent variables were initially checked for potential multicollinearity. Figure 1 that follows, illustrates the correlation plot among variables. Values of correlation coefficients vary between -1 (perfect negative correlation) and 1 (perfect positive correlation). Values close to 0 (zero) indicate no correlation. Only non-correlated variables can be inserted simultaneously in the statistical models.

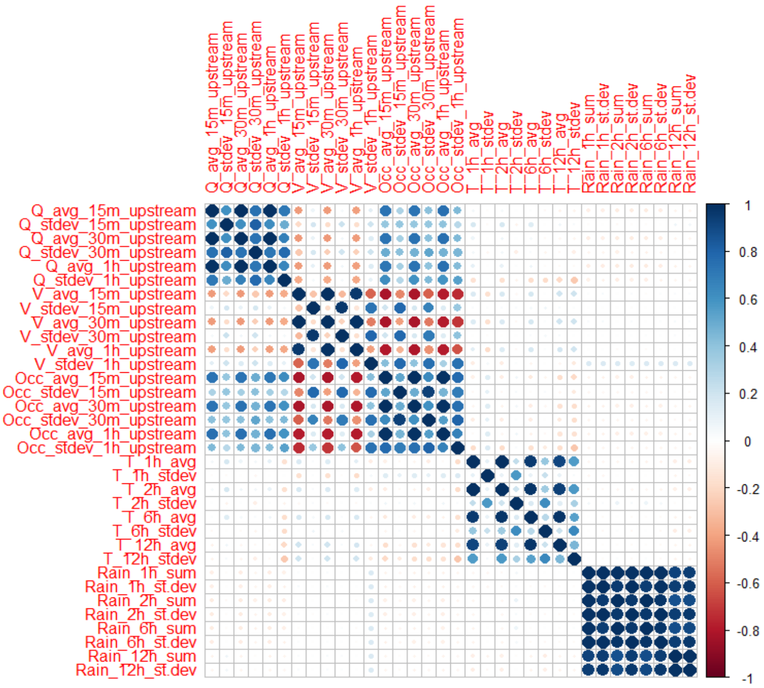

Fig. 1. Correlation plot of independent variables.

One cusp catastrophe model was developed in this study for the total number of crashes per road segment. The dependent variable is related to the state variable y, while other variables including the average 30-min flow (Q_avg_30m_upstream) constitute the control factors α. It is noted that there is no priori selection of the variables to be assigned to each control factor. The results of the best model in terms of goodness-of-fit are presented in Table 2.

Overall, it can be seen that traffic flow has a strong non-linear effect on crash numbers. Thus, small changes in traffic flow can cause high number of crashes. It is interesting that no variables were found to significantly impact the bifurcation factor β.

Figures 2 and 3 visualize the 2D and 3D projections of the cusp catastrophe surfaces for crash rate and incident rate respectively. The x-axis represents the asymmetry factor α, and the y-axis represents the bifurcation or splitting factor β. Each dot represents a single motorway location in the dataset. The cases that lie inside the bifurcation areas are the cases where a sudden or dramatic change in crash rates can occur when there is a small change in parameters α and β. A lot of cases however, lie outside the "instable" area. Hence, this cusp model does not guarantee that the system is non-linear as there is a lot space for improvement, even though the pseudo-R-squared of the cusp model is high.

Table 2. Summary statistics of crash related variables (cusp model).

Assymetry factor a	**Related Variable**	**Coefficient**	**Std.error**	**p-value**
ao	constant term	-1.521	0.708	0.0318 ***
a1	Q_avg_30m_upstream	0.017	0.009	0.0748**
Bifurcation factor β				
βo	constant term	1.180	0.712	0.0976 **
Dependent variable y				
wo	constant term	-2.591	0.304	< 0.001 ***
w1	Number of total crashes per segment	0.563	0.077	< 0.001 ***
pseudo-R^2	0.667			
Logistic model-R^2	0.204			

***: Significant at 95% level
**: Significant at 90% level

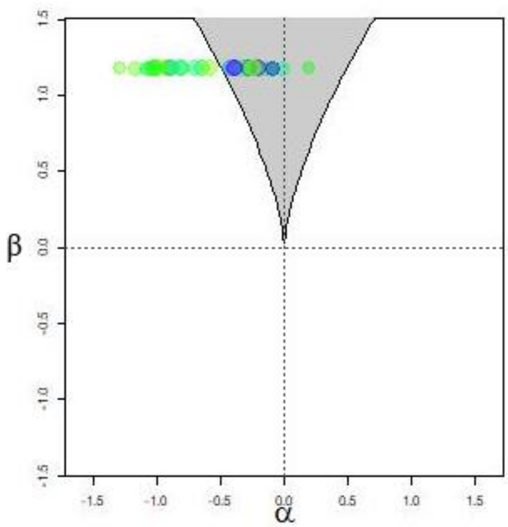

Fig. 2. 2-Dimension projection of cusp surface.

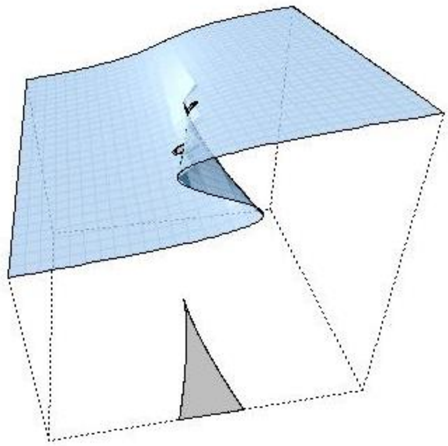

Fig. 3. 3-Dimension projection of cusp surface

For comparison purposes, a separate negative binomial model was developed. The model has an adequate R-squared (0.223), but shows worse fit than the cusp catastrophe model. This however, does not reject the hypothesis that the system is linear due to a number of problems of the cusp model as discussed earlier. Random effect models were also tested but results showed no evidence and hence the fixed effect model is adequate for this study.

The binomial model indicates that average flow upstream (Q_avg_30m_upstream) increases crash frequency. This finding can be considered in line with previous relevant studies in the field (Yu et al., 2013; Yu and Abdel-Aty, 2013a and 2013b).

Additionally, the positive beta coefficient (1.903) indicates that the standard deviation of rainfall in the 0-12h time interval prior to crash (Rain_12h_st.dev) is associated with high number of crashes. It is noted that similar to previous studies in the field, the standard deviation of rainfall also expresses rain intensity. However, opposite findings are suggested by Yu et al. (2013), who state that under intense rainfall drivers are more cautious and therefore fewer crashes occur. The positive association in our studies may indicate that drivers are surprised and more crashes occur. Finally, rainfall intensity is suggested to have a linear relationship with crash numbers. That is the reason why it is not significant in the cusp model.

Table 3. Summary statistics of crash related variables (negative binomial model).

Variable	**Coefficient**	**Std. error**	**p-value**
constant term	0.754	0.270	0.0053***
Q_avg_30m_upstream	0.008	0.004	0.0568**
Rain_12h_st.dev	1.903	1.142	0.0957**
pseudo-R^2	0.223		

***: Significant at 95% level

**: Significant at 90% level

5. Conclusions

This study has presented the analysis of crash frequencies in road segments of an urban freeway in Athens, Greece by applying cusp catastrophe models. The aim was to examine the assumption that safety of the system as expressed by the crash frequency, could be considered as a nonlinear dynamic system, where the transitions from safe to unsafe conditions and vice versa, can occur due to smooth or small changes to some control factors. Real-time traffic and weather information were considered as potentially critical to the construction of the control factors.

A number of interesting findings have been disclosed. Firstly, is that rainfall intensity has a strong linear impact on crashes (high rainfall intensity causes more crashes). On the other hand, average flow is indicated to have a strong non-linear relationship with crash frequency. The obtained results provide some evidence that road safety in urban freeways could be treated as nonlinear dynamic system, when high resolution traffic and weather traffic data are considered. In other words, the findings may imply that the dynamic change in urban road safety levels expressed by crash severity is likely to be nonlinear in nature. Unlike the traditional negative binomial approach, the results indicate the possible existence of a catastrophic influence of medium-term changes in traffic flow on the system, as sudden changes between different states of the system take place. As a consequence, this theory could be seen as a useful tool for developing indicators of a catastrophe, although the actual points at which the catastrophic changes occur cannot be easily predicted.

However, findings do not strongly confirm the strong presence of a dynamic system, due to the fact that the diagnostic tests of the cusp models did not produce very robust conclusions. Hence, traditional models such as the negative binomial model are proved equally capable of describing the underlying phenomenon, even though the goodness-of-fit is not as good as that of the cusp model. Overall, one can conclude that in such cases the linearity of the safety system is preserved.

Acknowledgements

This research is implemented through IKY scholarships programme and co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the action entitled "Reinforcement of Postdoctoral Researchers", in the framework of the Operational Programme "Human Resources Development Program, Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) 2014 – 2020. The authors would like to explicitly thank George Yannis, Professor of the National Technical University of Athens, for his guidance and advice for improving the work carried out in this research. The authors would also like to thank the Attica Tollway Operations Authority ("Attiki Odos") for providing the data, especially officers Pantelis Kopelias and Fanis Papadimitriou.

References

Abdel-Aty, M.A., Radwan, A.E., 2000. Modeling traffic accident occurrence and involvement. Accident Analysis and Prevention 32.5, 633–642.

Ahmed, M., Huang, H., Abdel-Aty, M., Guevara, B., 2011. Exploring a Bayesian hierarchical approach for developing safety performance functions for a mountainous freeway. Accident Analysis and Prevention 43, 1581–1589.

Akaike, H., 1974. A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control 19.6, 716–723.

Anastasopoulos, P., Mannering, F., 2009. Tobit analysis of vehicle accident rates on interstate highways. Accident Analysis and Prevention 41, 153–159.

Anastasopoulos, P., Mannering, F., Shankar, V., Haddock, J., 2012a. A study of factors affecting highway accident rates using the random-parameters tobit model. Accident Analysis and Prevention 45, 628–633.

Anastasopoulos, P., Shankar, V., Haddock, J., Mannering, F., 2012b. A multivariate tobit analysis of highway accident-injury-severity rates. Accident Analysis and Prevention 45, 110–119.

Anastasopoulos, P., Tarko, A., Mannering, F., 2008. Tobit analysis of vehicle accident rates on interstate highways. Accident Analysis and Prevention 40, 768–775.

Cobb, L., 1998. An Introduction to the Cusp Surface Analysis. Aetheling Consultants, Louisville, CO, USA.

Hartelman, P., 1997. Stochastic Catastrophe Theory. Ph.D. thesis, University of Amsterdam, Amsterdam, the Netherlands.

Lee, C., Saccomanno, F., Hellinga B., Lee C., Hellinga, B., 2002. Analysis of Crash Precursors on Instrumented Freeways. Transportation Research Record: Journal of Transportation Research Board, 1–8.

Noland, R.B., Quddus, M.A., 2005. Congestion and Safety: A Spatial Analysis of London. Transportation Research Part A: Policy and Practice 39.7, 737–754.

Park, P.Y., Abdel-Aty, M., 2011. A Stochastic Catastrophe Model Using Two-Fluid Model Parameters to Investigate Traffic Safety on Urban Arterials. Accident Analysis and Prevention 43.3, 1267–1278.

Poch, M., Mannering, F., 1996. Negative Binomial analysis of intersection accident frequencies. Journal of Transportation Engineering 122.2, 105–113.

Savolainen, P.T., Tarko, A.P., 2005. Safety impacts at intersections on curved segments. Transportation Research Record 1908, 130–140.

Schwarz, G., 1978. Estimating the Dimension of a Model. The Annals of Statistics 6.2, 461–464.

Theofilatos, A., G. Yannis, E. I. Vlahogianni, and J. C. Golias. Stochastic Cusp Catastrophe Models with Traffic and Weather Data for Crash Severity Analysis on Urban Arterials. Presented at the Transportation Research Board 96th Annual Meeting Transportation Research Board, 2017.

Van der Maas H., Kolstein, R., van der Pligt, J., 2003. Sudden transitions in attitudes. Sociological Methods and Research 23.2, 125–152.

Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2010. Statistical and Econometric Methods for Transportation Data Analysis, second ed. Chapman Hall/CRC, Boca Raton, FL.

Yu, R., Abdel-Aty, M., 2013a. Multi-level Bayesian analyses for single- and multi-vehicle freeway crashes Accident Analysis and Prevention 58, 97–105.

Yu, R., Abdel-Aty, M., 2013b. Investigating the different characteristics of weekday and weekend crashes. Journal of Safety Research 46, 91–97.

Yu, R., Abdel-Aty, M., Ahmed, M., 2013. Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors. Accident Analysis and Prevention 50, 371–376.