STOCHASTIC CUSP CATASTROPHE MODELS WITH TRAFFIC AND WEATHER DATA FOR CRASH SEVERITY ANALYSIS ON URBAN ARTERIALS

Athanasios Theofilatos*, PhD
Research Associate
National Technical University of Athens,
5 Iroon Polytechniou St., GR-15773, Athens, Greece
Tel: +30 210 772 1575; Fax: +30 210 772 1454
E-mail: atheofil@central.ntua.gr

George Yannis, PhD
Professor
National Technical University of Athens
Department of Transportation Planning and Engineering
5 Iroon Polytechniou St., GR-15773, Athens, Greece
Tel: +302107721326, Fax: +302107721454
E-mail: geyannis@central.ntua.gr

Eleni I. Vlahogianni, PhD
Assistant Professor
National Technical University of Athens,
5 Iroon Polytechniou Str, Zografou Campus, 157 73 Athens Greece
Tel: +30 210 772 1369; Fax: +30 210 772 1454
E-mail: elenivl@central.ntua.gr

John C. Golias, PhD
Professor
National Technical University of Athens,
5 Iroon Polytechniou Str, Zografou Campus, 157 73 Athens Greece
Tel: +30 210 772 1276; Fax: +30 210 772 1454
E-mail: igolias@central.ntua.gr

*Corresponding Author

Word Count: 7180 **(**Tables: 5, Figures: 2)

ABSTRACT

The investigation of crash severity 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 injury severity enhanced with traffic and weather data collected from urban arterials in Athens, Greece. Cusp catastrophe models are applied and compared with traditional statistical models. The results of crash severity models support the potential applicability of the cusp catastrophe theory to road safety, at least when crash severity is expressed as the number of severely and fatally injured by total number of persons involved in a crash. Variations in speed, average flow upstream of the location of interest, crash type and wind speed, were found to have a potential effect on the system dynamics. However, findings do not always confirm the strong presence of nonlinearity. When crash severity is expressed as the number of injured persons by the total number of vehicles involved in a crash, linear models could also be used to describe the underlying phenomenon.

Keywords: Crash, injury severity, cusp catastrophe, macroscopic traffic data, weather information, urban arterials

INTRODUCTION

The effective treatment of crashes and the proactive transportation safety is a major concern to societies due to the losses in human lives and the economic and social costs. According to World Health Organization (1), the total number of road fatalities worldwide remains at 1.24 million per year.

Over the last decade, much research that utilized real-time collected traffic and weather data in freeways has been carried out. Specifically, a large number of studies have investigated the effect of short-term traffic and weather parameters prior to a crash in order to explore crash likelihood (2, 3, 4, 5, 6,) and crash severity (7, 8, 9). The methodology of these studies is to perform a matched case-control approach by considering crash cases, but also a random sample of non-crash cases.

However, relevant research for crash injury severity is relatively limited. Fewer studies utilizing real-time traffic and weather data were found in international literature (6, 7, 10, 8, 9). Findings are diverse; Christoforou et al. (7) investigated injury severity by applying fixed and random parameters ordered probit models and found that increased traffic volume leads to less severe injuries. Other studies indicate that traffic parameters have limited influence (11, 12, 13) or even reduce severity of crashes (6). Other findings indicate that large speed variations and low visibility increase crash severity (8). Another study (11) examined single-vehicle crashes in Wisconsin and found that increased rainfall intensity increases severity of crashes. Low visibility conditions and fog were found to be positively correlated with crash severity (8, 14, 15).

From a deeper look at the related literature it becomes evident that very limited research on both crash severity and likelihood focuses on urban arterials (16, 17). Nevertheless, a number of studies examine urban expressways (12, 13, 18, 19, 20, 21). Another issue relates to the methodological part, as alternative methods should be sought to better explain crash severity (7, 22). Cusp catastrophe theory can be considered an alternative and promising methodological approach and it has been applied in traffic flow theory (23, 24, 25), but very rarely in transportation safety (26). This method of analysis is different than the existing classical statistical methods, due to the fact that cusp catastrophe theory investigates the existence of potential non-linearity in the "dynamic system" that causes sudden transitions between states (e.g. "safe" and "unsafe state"), due to small changes in the input parameters (e.g. independent variables).

The knowledge of the transitions between safe and unsafe traffic conditions is critical to road safety. The conceptual gain of introducing cusp catastrophe theory in modeling road safety is that it is theoretically possible to identify which risk factors may cause sudden deterioration of road safety levels. For example, specific road sites which are considered safe (are in a safe state), could be easily turned into high crash risk sites (transition to unsafe state), if even slight changes in specific risk factors take place. On the other hand, researchers may detect which are the most effective road safety measures that can easily cause sudden improvements in road safety (e.g. sudden drop in crash counts) without the need of dramatic changes. This is also discussed by Park and Abdel-Aty (26). Summing up, the better understanding of cusp catastrophe model applicability in crash analyses is considered promising, since it can contribute to develop proactive safety approaches.

Therefore, the aim of this study is twofold; a) firstly to add to the current knowledge by applying cusp catastrophe models in order to investigate crash injury severity, and b) to consider real-time traffic and weather data in urban arterials. The developed nonlinear models are contrasted to well-established traditional statistical approaches. Data come from urban arterials located in the Athens metropolitan area (Greece).

The remainder of the paper is organized as follows. The data preparation is demonstrated, followed by the description of methodology applied in order to explore crash severity. Next, the application of the models is illustrated and the results are presented and discussed. The last section provide the conclusions.

DATA PREPARATION

The available dataset refers to the period 2006-2011 and come from two high demand urban arterials in the center of Athens (Greece). These two arterials have similar geometrical and traffic characteristics. The dataset contains safety, traffic and weather data. More specifically, crash data were collected from the Greek accident database, SANTRA, which is provided by the National Technical University of Athens. It provides access to road crash in Greece since 1985 including all relevant information about each crash (persons injured, severity of injuries, location, weather, accident type etc.).

Traffic data were extracted from the Traffic Management Centre (TMC) of Athens, which has been in operation since 2004 and covers several major roads in Athens. The TMC data included traffic flow, traffic occupancy and time speed every 1 minute. Traffic data from the adjacent upstream loop detector were considered. Data were further aggregated to 1-hour traffic information to obtain averages, standard deviations and so on, prior to a crash occurrence. It was anticipated that the 60-min traffic data before crash occurrence would cover the hazardous traffic conditions, consequently only the traffic data 1-hour prior to crash occurrence were initially considered.

Weather data were collected from the Hydrological Observatory of Athens (27), which is an online open-access database, covering more than 10 meteorological stations located in the greater Athens area and providing measurements about rainfall, temperature, relative humidity, solar radiation, wind direction, wind speed etc. Each crash case was assigned to the closest meteorological station and then the relevant weather data had to be extracted. Then the 10-min raw data were aggregated over hour in order to obtain maxima, averages and standard deviations, in the time-slice of 1-hour, 2-hours, 6-hours and 12-hours prior to the time of the crash occurrence.

For the analysis to follow, a time lag of 20 minutes was used. This means that 20 minutes were subtracted from the time of each crash case in order to avoid the impact of the accident itself on the traffic variables and also to compensate for any potential inaccuracies in the precise time of the accident. This approach has been followed by previous relevant studies (7, 28). The following example illustrates the approach. If a crash occurred on 14 February at 12:00 at the loop detector “MS258”, then the relevant traffic and weather data are extracted for the time period 10:40 to 11:40 from the closest upstream loop detector and from the closest meteorological station respectively.

The final available dataset included 353 crash cases (not including crashes for intersections) for Kifisias and Mesogeion avenues from 2006 to 2011. As in various crash severity studies (8, 9) crash severity consisted of two levels, namely, fatal/severe injury (KSI) and slight injury (SI). A percentage of 11% of crashes were classified as severe (KSI), while 89% were classified as slight (SI). Powered-Two-Wheelers (PTWs) were involved in 225 of those crashes (63.7% of crashes). In order to explore crash severity, the dependent variable “severity” was decided to be re-defined and re-coded. Two types of severity were used. First, severity was defined as a percentage of the total severely or killed persons involved in each crash by the total number of persons involved in a crash (Severity_1). The other type of severity is the total number of persons involved in a crash divided by the total number of vehicles involved in a crash (Severity_2). The two types of severity are defined as follows:

$$\text{Severity\_1} = \frac{\textit{Number of severely injured and killed}}{\textit{Total number of persons invovled}} \quad (1)$$

$$\text{Severity\_2} = \frac{\textit{Total number of persons invovled}}{\textit{Total number of vehicles involved}} \quad (2)$$

Table 1 shows the descriptive statistics for crash severity.

METHODOLOGY

The proposed methodology is based on the cusp catastrophe models. The theoretical background is illustrated in this section of the study providing also a brief description of the catastrophe theory in general. For a detailed description about the cusp catastrophe, the reader is encouraged to refer to (29). A further comparative analysis will also take place using censored regression to compare the performance and explanatory power of the two methods.

Catastrophe Theory

Catastrophe theory examines the qualitative changes in the behavior of systems when the control factors that influence their behavioral state face smooth and gradual changes (30). 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, 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. Catastrophe theory became popular in the 1970's and since then its applications range from economics to psychology. However, this approach had a few major drawbacks. The major reason of criticism stems from the qualitative methodology used in the aforementioned applications, due to the fact that catastrophe theory concerned deterministic dynamical systems (29). Another issue is the ad hoc nature of the selection of the variables that would be used as control factors. (31), comprehensively summarizes the critiques of catastrophe theory and the reader may refer to this study. Consequently, there was a deep need to make a stochastic formulation in the catastrophe theory. Indeed, several stochastic formulations have been found along with statistical methods, so that the quantitative comparison of catastrophe models with data is enabled (32, 33, 34).

TABLE 1 Description and Descriptive Statistics of Sample Variables

Variable	Description	Unit	Mean	Std. deviation
Severity_1	Number of severely injured and killed divided by Total number of persons involved	unitless	0.086	0.259
Severity_2	Total number of persons involved divided by Total number of vehicles involved	unitless	0.885	0.500
	Acc.Type1 (Off road/Fixed object/Other)	unitless		155*
	Acc.Type2 (Head-on)	unitless		36*
Acc.Type	Acc.Type3 (Rear-end)	unitless		73*
	Acc.Type4 (Side)	unitless		53*
	Acc.Type5 (Sideswipe)	unitless		36*
Q_avg_1h_up	1h average flow per lane upstream	veh/hour/lane	810.450	301.719
Q_stdev_1h_up	1h st.deviation of flow per lane upstream	veh/hour/lane	264.330	339.374
Q_median_1h_up	1h median of flow per lane upstream	veh/hour/lane	628.600	437.337
Q_cv_1h_up	1h coefficient of variation of flow per lane upstream	unitless	0.109	0.085
V_avg_1h_up	1h average speed upstream	km/h	47.340	18.959
V_stdev_1h_up	1h st.deviation of speed upstream	km/h	5.333	5.591
V_cv_1h_up	1h coefficient of variation of speed upstream	unitless	0.154	0.175
Occ_avg_1h_up	1h average occupancy upstream	percentage %	15.730	11.143
Occ_stdev_1h_up	1h st.deviation of occupancy upstream	percentage %	4.097	4.917
Occ_cv_1h_up	1h coefficient of variation of occupancy upstream	unitless	0.248	0.216
T_1h_max	1h maximum temperature	^oC	19.240	7.710
T_1h_avg	1h average temperature	^oC	18.700	7.714
T_1h_stdev	1h st.deviation of temperature	^oC	0.397	0.335
Rain_1h_sum	1h sum of rainfall	mm	0.030	0.265
Rain_1h_st.dev	1h st.deviation of rainfall	mm	0.004	0.031
Rain_2h_sum	2h sum of rainfall	mm	0.068	0.618
Rain_2h_st.dev	2h sum of rainfall	mm	0.010	0.094
Rain_6h_sum	6h sum of rainfall	mm	0.152	0.921
Rain_6h_st.dev	6h st.deviation of rainfall	mm	0.013	0.083
Rain_12h_sum	12h sum of rainfall	mm	0.252	1.142
Rain_12h_st.dev	12h st.deviation of rainfall	mm	0.014	0.068
W.Sp_1h_max	1h maximum wind speed	m/sec	2.759	1.836
W.Sp_1h_avg	1h average wind speed	m/sec	2.204	1.688
W.Sp_1h_stdev	1h st.deviation of wind speed	m/sec	0.387	0.223
Sol_1h_max	1h maximum solar radiation	W/m^2	377.410	362.890
Sol_1h_avg	1h average solar radiation	W/m^2	307.550	321.884

* Distribution of crash types

Cusp Catastrophe

Of the seven elementary types of catastrophe models, perhaps the most popular and easy is the cusp catastrophe. 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. Let a deterministic dynamical system by defined as:

$$\frac{\partial y}{\partial t} = -\frac{\partial V(y;\alpha,\beta)}{\partial y} \quad (3)$$

where y represents the state variable (can be considered as the dependent variable) and α, β are the two control parameters that determine the behaviour of the system. The canonical form of the cusp catastrophe function is:

$$-V(y;\alpha,\beta) = \alpha y + \frac{1}{2}\beta y^3 - \frac{1}{4}\beta y^4 \quad (4)$$

This system moves towards equilibrium and will reach one when:

$$-\frac{\partial V(y;\alpha,\beta)}{\partial y} = 0 = \alpha + \beta y - y^3 \quad (5)$$

There is one solution to this equation if $\delta > 0$, and three solutions if $\delta < 0$. The term δ is also called Cardan's discriminant and is defined as

$$\delta = 27\alpha - 4\beta^3 \quad (6)$$

The set of values α and β for which $\delta = 0$, determines the bifurcation set. A few studies (29, 35) have well-explained the cusp equilibrium surface for nonlinear deterministic systems. Statistically speaking, the cusp equilibrium surface may be considered as a response surface, where depending on the values of α and β, its height predicts the value of the dependent variable. Moreover, the dependent variable y cannot be necessarily observed (and thus being an observed quantity), but it is rather a canonical variable depending on a number of measured dependent variables. In that context, the control variables α and β are canonical as well and depend on a number of actual measured independent variables.

A number of qualitative behaviors of the cusp model were derived by (36). These characteristics are called catastrophe flags. These characteristics are of major importance, because the existence of some (or all) of them indicates a strong presence of a good fit to the data and therefore evidence is gathered for the presence of cusp catastrophe in the system. Some of them are sudden jumps in the value of the canonical state variables, hysteresis and multi-modality. As stated earlier, catastrophe models are applied in deterministic systems. As a consequence, these models cannot be directly applied in stochastic environments. For that reason, a stochastic catastrophe theory was proposed (37, 38, 39, 40) by adding a white noise Wiener process, namely $dW(t)$ to the initial Equation 3. Therefore, Equation 3, is transformed to a stochastic differential equation:

$$dY = \frac{\partial V(Y;\alpha,\beta)}{\partial Y} dt + dW(t) \quad (7)$$

This stochastic differential equation is affiliated with a probability density that describes the allocation of system states at any moment in time. It can be expressed as follows:

$$f(y) = \frac{\sigma}{\psi^2} exp[\frac{\alpha(y-\lambda)+\frac{1}{2}\beta(y-\lambda)^2-\frac{1}{4}(y-\lambda)^4}{\sigma^2}] \quad (8)$$

where ψ is the normalizing constant and λ determines the location parameter. 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. As stated earlier, the variables y, α and β are canonical. Now let's assume a set of measured dependent variables $Y_1, Y_2, \ldots, Y_n$, then:

$$y = w_0 + w_1 Y_1 + w_2 Y_2 + \cdots + w_n Y_n. \tag{9}$$

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

$$\alpha = \alpha_0 + a_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n, \tag{10}$$

and

$$\beta = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_n X_n. \tag{11}$$

In order to assess the fit of the cusp model a set of diagnostic tools have been suggested such as the pseudo-R^2 (41), the well-known AIC (42) and BIC (43). It is noted that the pseudo-R^2 can become negative. In order to further evaluate the cusp model fit, more diagnostics are suggested (41). For example, each one of the coefficients $w_1, w_{2,} \ldots w_n$ should be statistically significant (except w_0) as well at least one of the a's or the b's. Moreover, at least 10% of the pairs (a_i, β_i) should lie inside the bifurcation region. One alternative diagnostic measure according to a number of studies such as this of (44) and (45), is the comparison of the cusp model with a nonlinear logistic model:

$$y = \frac{1}{1+\exp(-\frac{\alpha}{\beta})} + \varepsilon \tag{12}$$

where the parameters (x, y, z) were defined previously in Equations 9, 10 and 11, while ε is the random disturbance. The nonlinear logistic model has the ability to model the sudden changes in the response variable y in a way "similar" to the sudden transition in the cusp model.

Censored Regression

In our study, the dependent variables are censored, as they cannot take all values. More specifically, the severity variables are continuous and take all values between 0 and 1(are censored). Tobit model was introduced by Tobin (46) and is proposed to be applied when the dependent variable is censored in some way. In such case it is suggested that traditional linear regression models are not appropriate (47).

The censored regression model is a generalization of the standard Tobit model. All corresponding equations can be found in (48). The dependent variable can have a lower threshold (left-censored) or an upper threshold (right-censored) or both:

$$y_i^* = x_i^{'} \beta + \varepsilon_i \tag{13}$$

$$y_i = \begin{cases} a & if\ y_i^* \le a \\ y_i^* & if\ a < y_i^* < b \\ b & if\ y_i^* \ge b \end{cases} \quad (14)$$

The censored regression model is usually estimated by the Maximum Likelihood method. The error term ε is assumed to follow a normal distribution (0, σ^2). The likelihood function is the following:

$$logL = \sum_{i=1}^{N}[I_i^a\, log\Phi\left(\frac{a-x_i^{'}\beta}{\sigma}\right) + I_i^b log\Phi\left(\frac{x_i^{'}\beta-b}{\sigma}\right) + (1-I_i^a-I_i^b)(log\varphi\left(\frac{y_i-x_i^{'}\beta}{\sigma}\right) - log\sigma)],$$
(15)

where φ() and Φ() denote the probability density function and the cumulative distribution respectively of the standard normal distribution. I_i^a and I_i^b are indicator functions:

$$I_i^a = \begin{cases} 1 & if\ y_i = a \\ 0 & if\ \ y_i > a \end{cases} \quad (16)$$

$$I_i^b = \begin{cases} 1 & if\ y_i = b \\ 0 & if\ \ y_i < b \end{cases} \quad (17)$$

RESULTS

In this study, a series of cusp catastrophe models are developed by utilizing real-time traffic and weather data, to model the micro-level safety (crash severity) in urban arterials. The dependent variable is related to the state variable y, while the traffic, weather as well as other crash variables constitute the control factors α and β. However, there is no a priori determination of which parameter (traffic, weather, other) would be assigned to each control factor. This occurs because there is a lack of objective criteria to determine whether a predictor variable should be classified as an asymmetry or as a bifurcation variable (29, 49). For interpretation reasons, one measured dependent variable is investigated each time, therefore equation 9 is simplified to:

$$y = w_0 + w_1 Y_1. \quad (18)$$

The intention is to investigate the potential existence of non-linearity in the system and also the potential transition of the “crash severity state” from a lower crash severity state (safe state) to a higher crash severity state (unsafe state), or vice versa, through the changes of the various traffic, weather and other crash predictors. Due the special nature of the cusp modeling approach results should always be interpreted carefully (26). Tables 2 and 3 illustrate the findings of cusp catastrophe and censored regression model respectively for Severity_1 (number of severely injured and killed divided by total number of persons involved).

Results of Table 2 constitute a promising evidence of presence of cusp and imply strong nonlinear relationships between crash injury severity and independent variables. More specifically, it is found that small changes to crash type (Acc.type1, Acc.type2, Acc.type3, Acc.type4), maximum wind speed values (W.Sp_1h_max), traffic flow (log(Q_avg_1h_up)) or coefficient of variation of speed (V_cv_1h_up), may lead to sudden changes to crash severity. The cusp model has a considerably high value of McFadden R^2 (0.789) and pseudo-R^2 (0.845). However, this diagnostic does not guarantee the presence of cusp and therefore more evidence is needed. It is interesting though that the value of the logistic curve value of R^2 is significantly lower (0.095) showing that the cusp model is superior to the nonlinear logistic model.

TABLE 2 Cusp Catastrophe Model Results For Severity_1 Defined As The Number Of Severely Injured And Killed By Total Number Of Persons Involved

Assymetry factor a	Variable	Coefficient	Std. error	p-value
a_0	Constant term	-0.200	0.093	0.032**
a_1	V_cv_1h_up	-0.720	0.372	0.052*
a_2	Acc.type1	-0.417	0.228	0.067*
a_3	Acc.type2	-0.280	0.144	0.052*
a_4	Acc.type3	-1.021	0.420	0.015**
a_5	Acc.type4	-0.493	0.286	0.084*
a_6	W.Sp_1h_max	-0.068	0.032	0.031**
Bifurcation factor β				
b0	Constant term	-	-	-
b1	log(Q_avg_1h_up)	0.755	0.034	0.000**
Dependent variable y				
w0	Constant term	-2.342	0.047	0.000**
w1	Severity_1	4.732	0.101	0.000**
McFadden R^2	0.789			
pseudo-R^2	0.845			
Logistic model R^2	0.095			

**=95% significance level
*=90% significance level

Figure 1 visualizes the 3D and the 2D projection of the cusp catastrophe surface, where the x-axis is the α and the y-axis is the β. Each dot is a single case and its size varies according to the observed bivariate density of the control factors' values at the location of the point. More specifically, the color of the dots varies depending on the value of the response (state) variable y, with higher values being associated with more intense purple, while lower values with more intense green. It is observed that the 100% of cases fall inside the V-shaped curve which is the bifurcation area (instability area), meaning that all cases are in a very vulnerable condition where the current (low) severity state could be easily turned into a high severity state. Strictly speaking, the cases that lie inside the bifurcation areas are the cases where a sudden or dramatic change in severity level can occur when there is a small change in parameter α. Therefore, there is high possibility of presence of a dynamic nonlinear system.

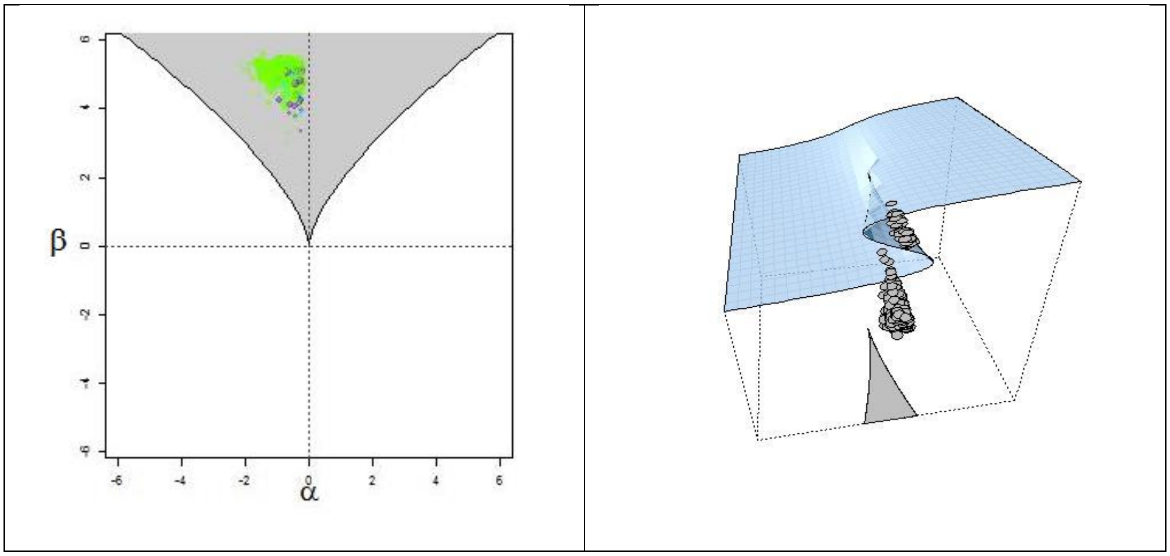

FIGURE 1 2D projection (left panel) and 3D projection (right panel) of cusp surface for Severity_1.

One more prerequisite for the confirmation of a good fit of the cusp model is the significance of parameter w_1 and at least one of parameters a and β. The model shows several significant variables and provides evidence for the existence of nonlinearity.

In order to further compare the cusp catastrophe model, censored regression was applied. The goodness of fit was assessed via the Madalla R^2 (47, 50), indicating a relatively good goodness of fit, but inferior to the cusp model. Moreover, the results of the model shows a consistent negative correlation among all independent variables and crash severity. It can be concluded that collisions with fixed object increase crash severity while maximum wind speed, variations in speed and increased traffic flow reduce it.

TABLE 3 Censored Regression Model Results For Severity_1 Defined As Number Of Severely Injured and Killed By Total Number Of Persons Involved

Variable	Coefficient	Std.error	p-value
Constant term	7.078	2.943	0.016**
V_cv_1h_up	-1.975	1.627	0.225
Acc.Type1	-1.731	1.049	0.098*
Acc.Type2	-1.225	0.712	0.085*
Acc.Type3	-12.729	563.132	0.982
Acc.Type4	-1.763	1.075	0.101
W.Sp_1h_max	-0.302	0.169	0.074*
log(Q_avg_1h_up)	-1.312	0.498	0.008**
Madalla R^2	0.119		

**=95% significance level

*=90% significance level

The interpretation of models for Severity_2 (total number of persons involved by total number of vehicles involved) is similar. Table 4 shows the results of the cusp catastrophe model. The cusp R^2 (0.303) is lower than the logistic curve R^2 (0.313), meaning that both models equally perform. As it can be observed, the significant variables introduced to the description of α and β in Severity_2 model differ from those in the Severity_1 model. One main difference is that crash type (Acc.type) variable in Severity_2 model is used as bifurcation factor, while in the first model is used as asymmetry factor. In the former model, the logarithm of traffic flow was used as a bifurcation factor, while in the latter model is used as asymmetry factor. These two findings indicate a different effect of these variables on each type of crash severity. It is also observed that maximum solar radiation (Sol_1h_max) has an impact only on Severity_2 model.

TABLE 4 Cusp Catastrophe Model Results for Severity_ 2 (Total Number Of Persons Involved By Total Number Of Vehicles Involved)

Assymetry factor a	Variable	Coefficient	Std. error	p-value
a_0	Constant term	-0.251	0.226	0.267
a_1	V_cv_1h_up	0.172	0.037	0.004**
a_2	log(Q_avg_1h_up)	-0.101	0.047	0.031**
a_3	Sol_1h_max	-0.001	0.000	0.006**
Bifurcation factor β				
b_0	Constant term	1.048	0.082	0.000**
b_1	Acc.Type1	0.783	0.280	0.005**
b_2	Acc.Type2	2.115	0.077	0.000**
b_3	Acc.Type3	2.350	0.242	0.000**
b_4	Acc.Type4	2.392	0.276	0.000**
Dependent variable y				
w0	Constant term	-2.847	0.041	0.000**
w1	Severity_2	1.689	0.056	0.000**
McFadden R^2	0.251			
pseudo-R^2	0.303			
Logistic model R^2	0.313			

**=95% significance level

*=90% significance level

Table 5 shows the censored regression analysis. In this case however, the cusp model is not confirmed to be better than the traditional statistical methods, as the value of Madalla R^2 of the censored model (0.3), is similar to those of the cusp model. Consequently, in this case, the existence of non-linearity is not clear. Results may also imply that the linearity in the system is preserved or that both linear and non-linear relationships explain the phenomenon.

TABLE 5 Censored Regression Model Results For Severity_ 2 (Total Number Of Persons Involved By Total Number Of Vehicles Involved)

Variable	Coefficient	Std.error	p-value
Constant term	1.484	0.291	0.000**
V_cv_1h_up	0.057	0.132	0.663
log(Q_avg_1h_up)	-0.041	0.000	0.367
Sol_1h_max	0.000	0.000	0.009**
Acc.Type1	-0.259	0.078	0.000**
Acc.Type2	-0.500	0.060	0.000**
Acc.Type3	-0.571	0.068	0.000**
Acc.Type4	-0.543	0.078	0.000**
Madalla R^2	0.300		

**=95% significance level

*=90% significance level

Figure 2 shows the graphical assessment of the 2D projection of the cusp catastrophe surface does not provide strong evidence of nonlinear relationships, although more than 10% of cases lie within the instability area.

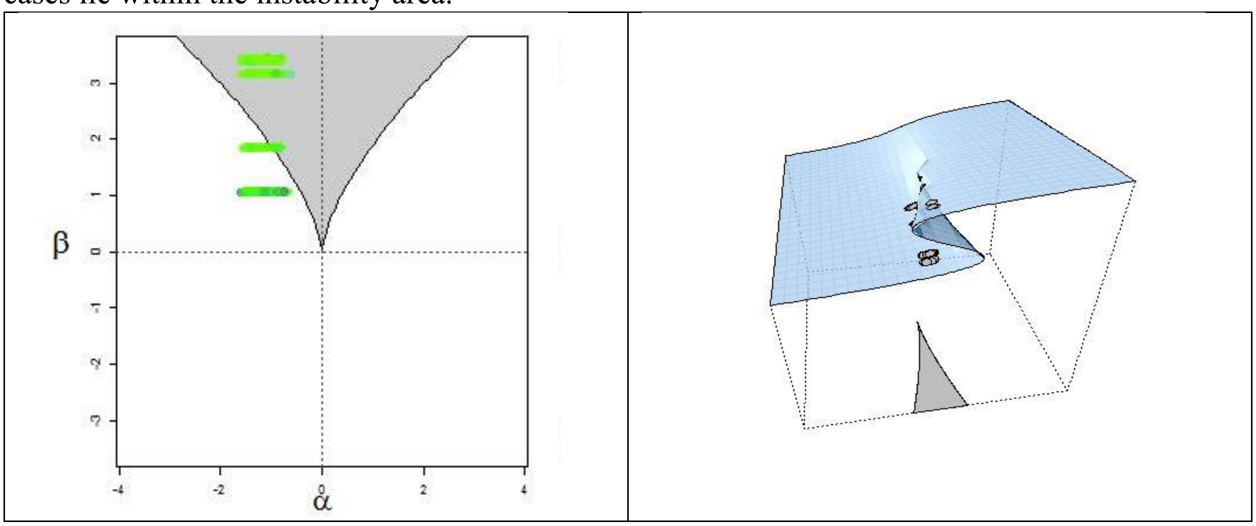

FIGURE 2 2D projection (left panel) and 3D projection (right panel) of cusp surface for Severity_2.

CONCLUSIONS

This study has presented the analysis of crash severity in urban arterials by applying cusp catastrophe models. The aim was to examine the assumption that safety of the system as expressed by the crash severity types, 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. Traffic, weather and traditional crash information were considered as potentially critical to the construction of the control factors.

The results of crash severity models justify the potential applicability of the cusp catastrophe, at least when crash severity is expressed as the number of severely and fatally injured by total number of persons involved in a crash. It is suggested that variations in speed, average flow upstream, crash type and wind speed, were found to have a potential effect on the system dynamics. However, findings do not always confirm the strong presence of a dynamic system. When crash severity is expressed as the number of injured persons by the total number of vehicles involved in a crash, linear models are proved equally capable of describing the underlying phenomenon. One can conclude that in such cases the linearity of the safety system is preserved.

The obtained results confirm in general that road safety in urban roads could be treated as nonlinear dynamic system, when high resolution traffic and weather traffic data are considered. Moreover, some other crash characteristics such as the type of crash, consistently influence the system dynamics. In other words, the findings indicate that the dynamic change in urban road safety levels expressed by crash severity is likely to be nonlinear in nature. Unlike the traditional linear modelling approach, the results indicate the possible existence of a catastrophic influence of medium-term changes in traffic and weather factors 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. Although there is definitely much room for additional research, this paper clearly demonstrates the possibility of using high resolution traffic and weather data to estimate crash severity and probability, through the development of an advanced stochastic differential equation (i.e., cusp catastrophe model).

It is to note that results of have to be treated with care, as the statistically satisfactory fit of the majority of the proposed models, by no means gives us definitive evidence for the presence of dynamical phase transition. In that context, if more research is done towards this direction, the prediction and the qualitative assessment of the catastrophe points would definitely have an outstanding contribution to road safety, due to the enhancement of the proactive safety management system.

REFERENCES

1. World Health Organization, WHO. *Global status report on road safety. Supporting a decade of action*, 2013.
2. Oh C., J-S. Oh, S.G. Ritchie and M. Chang. Real-time estimation of freeway accident likelihood. In *80th Annual Meeting of the Transportation Research Board*, Washington D. C., 8–12 January 2001.
3. Lee C., B. Hellinga and F. Saccomanno. Real-time crash prediction model for the application to crash prevention in freeway traffic. In *82nd Annual meeting of the Transportation Research Board*, Washington D.C., 12–16 January 2003.
4. Zheng Z., S. Ahn and C.M. Monsere. Impact of traffic oscillations on freeway crash occurrences. *Accident Analysis and Prevention*, Vol. 42, 2010, pp. 626–636.
5. Abdel-Aty M., H.M. Hassan, M. Ahmed and A.S. Al-Ghamdi. Real-time prediction of visibility related crashes. *Transportation Research Part C*, Vo. 24, 2012, pp. 288–298.

6. Xu C., A.P. Tarko, W. Wang and P. Liu. Predicting crash likelihood and severity on freeways with real-time loop detector data. *Accident Analysis and Prevention*, Vol. 57, 2013, pp. 30–39.
7. Christoforou Z., S. Cohen and M. Karlaftis. Vehicle occupant injury severity on highways: An empirical investigation. *Accident Analysis and Prevention*, Vol. 42, 2010, pp. 1606–1620.
8. Yu R. and M. Abdel-Aty. Using hierarchical Bayesian binary probit models to analyze crash injury severity on high speed facilities with real-time traffic data. *Accident Analysis and Prevention*, Vol. 62, 2014, pp. 161–167.
9. Yu R. and M. Abdel-Aty. Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. *Safety Science*, Vol. 63, 2014, pp. 50–56.
10. Jung S., X. Qin and D.A. Noyce. Rainfall effect on single-vehicle crash severities using polychotomous response models. *Accident Analysis and Prevention*, Vol. 42, 2010, pp. 213–224.
11. Golob T.F., W.W. Recker and Y. Pavlis. Probabilistic models of freeway safety performance using traffic flow data as predictors. *Safety Science*, Vol. 46, 2008, pp. 1306–1333.
12. Orfanou F.P., E.I. Vlahogianni and M.G. Karlaftis. Associating driving behavior with hysteretic phenomena of freeway traffic flow. *IFAC Proceedings Volumes*, 2012, pp. 209–214.
13. Vlahogianni E.I. and M.G. Karlaftis. Comparing traffic flow time-series under fine and adverse weather conditions using recurrence-based complexity measures. *Nonlinear Dynamics*, Vol. 69, Issue 4, 2012, pp. 1949–1963.
14. Al-Ghamdi A.S. Experimental evaluation of fog warning system. *Accident Analysis and Prevention*, Vol. 39, 2007, pp. 1065–1072.
15. Abdel-Aty M., A.-A. Ekrama, H. Huang and K. Choi. A study on crashes related to visibility obstruction due to fog and smoke. *Accident Analysis and Prevention*, Vol. 43, 2011, pp. 1730–1737.
16. Yannis G., A. Theofilatos, A. Ziakopoulos and A. Chaziris. Investigation of road accident severity and likelihood in urban areas with real-time traffic data. *Traffic Engineering and Control*, Vol. 55, Issue 1, 2014, pp. 31–35.
17. Theofilatos A. and G. Yannis. Investigation of Powered-Two-Wheeler accident involvement in urban arterials by considering real-time traffic and weather data. *Traffic Injury Prevention*, forthcoming.
18. Hossain M. and Y. Muromachi. A Bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways. *Accident Analysis and Prevention*, Vol. 45, 2012, pp. 373–381.
19. Hossain M. and Y. Muromachi. Understanding crash mechanism on urban expressways using high-resolution traffic data. *Accident Analysis and Prevention*, Vol. 57, 2013, pp. 17–29.
20. Shi Q., M. Abdel-Aty and J. Lee. A Bayesian ridge regression analysis of congestion's impact on urban expressway safety. *Accident Analysis and Prevention*, Vol. 88, 2016, pp. 124–137.
21. Shi Q., M. Abdel-Aty and R. Yu. Multi-level Bayesian safety analysis with unprocessed Automatic Vehicle Identification data for an urban expressway. *Accident Analysis and Prevention*, Vol. 88, 2016, pp. 68–76.

22. Theofilatos A. and G. Yannis A review of powered-two-wheeler behaviour and safety. *Injury Control and Safety Promotion*, Vol. 22, Issue 4, 2015, pp. 284–307.
23. Dendrinos D. Operating speeds and Vol. to capacity ratios: the observed relationship and the fold catastrophe. *Transportation Research*, Vol. 12, Issue 3, 1978, 191–194.
24. Navin F. Traffic congestion catastrophe. *Transportation Planning and Technology*, Vol. 11, 1986, pp. 19–25.
25. Papacharalampous A. and E.I. Vlahogianni. Modeling Microscopic Freeway Traffic Using Cusp Catastrophe Theory. *IEEE Intelligent Transportation*, Vol. 6, Issue 1, 2014, pp. 6–16.
26. Park P.Y. and M. Abdel-Aty. A stochastic catastrophe model using two-fluid model parameters to investigate traffic safety on urban arterials. *Accident Analysis and Prevention*, Vol. 43, 2011, pp. 1267–1278.
27. Hydrological Observatory of Athens, HOA. www.hoa.ntua.gr. Accessed Feb. 24 2014.
28. Quddus M.A., C. Wang and S.G. Ison. Road traffic congestion and crash severity: econometric analysis using ordered response models. *Journal of Transportation Engineering*, 2010, Vol.136, Issue 5, pp. 424–435.
29. Grasman R., H. van der Maas and E.J. Wagenmakers. Fitting the cusp catastrophe in R: A cusp Package Primer. *Journal of Statistical Software*, Vol. 32, Issue 8, 2009, pp. 1–28.
30. Poston T. and I. Stewart. *Catastrophe Theory and Its Applications*. Pitman, London, 1978.
31. Rosser J. The rise and fall of catastrophe theory applications in economics: Was the baby thrown out with the bathwater? *Journal of Economic Dynamics and Control*, Vol. 31, Issue 10, 2007, pp. 3255–3280.
32. Cobb L., P. Koppstein and N. Chen. Estimation and moment recursion Relations for multimodal distributions of the exponential family. *Journal of the American Statistical Association*, Vol. 78, Issue 381, 1983, pp. 124–130.
33. Guastello S.J. Moderator regression and the cusp catastrophe-Application of 2-stage personnel selection, training, therapy and policy evaluation. *Behavioral Science*, Vol. 27, Issue 3, 1982, pp. 259–272.
34. Wagenmakers E.J., P. Molenaar, R. Grasman, P. Hartelman and H. van der Maas. Transformation invariant stochastic catastrophe theory. *Physica D*, Vol. 211, 2005, pp. 263–276.
35. Barunik J. and M. Vosvrda. Can a stochastic cusp catastrophe model explain stock market crashes? *Journal of Economic Dynamics and Control*, Vol. 33, Issue 10, 2009, pp. 1824–1836.
36. Gilmore R. *Catastrophe Theory for Scientists and Engineers.* Dover, New York, 1983.
37. Cobb L. and R.K. Ragade. Applications of Catastrophe Theory in the Behavioural and Life Sciences. *Behavioral Science*, Vol. 23, 1978, pp. 291–419.
38. Cobb L. Estimation theory for the cusp catastrophe model. *Proceedings of the Section on Survey Research Methods*, 1980, pp. 772–776.
39. Cobb L. and B. Watson. Statistical catastrophe theory: an overview. *Mathematical Modelling*, Vol. 1, Issue 4, 1980, pp. 311–317.
40. Cobb L. and S. Zacks. Applications of the catastrophe theory for statistical modeling in the biosciences. *Journal of the American Statistical Association*, Vol. 80, Issue 392, 1985, pp. 793–802.

41. Cobb L. *An introduction to the cusp surface analysis.* Technical Report, Aetheling Consultants, Louisville, CO, USA. http://www.aetheling.com/models/cusp/Intro.htm. Accessed Mar. 5 2014.
42. Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, Vol. 19, 1974, pp. 716–723.
43. Schwarz G. Estimating the dimension of a model. *The annals of Statistics*, Vol. 6, 1978, pp. 461–464.
44. Hartelman, P. Stochastic Catastrophe Theory. *Ph.D. thesis*, University of Amsterdam, Amsterdam, the Netherlands, 1997.
45. Van der Maas H., R. Kolstein and J. van der Pligt. Sudden transitions in attitudes. *Sociological Methods and Research*, Vol. 23, Issue 2, 2003, pp. 125–152.
46. Tobin J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, Vol. 26, 1958, pp. 24–36.
47. Anastasopoulos P., A. Tarko and F. Mannering. Tobit analysis of vehicle accident rates on interstate highways. *Accident Analysis and Prevention*, Vol. 40, 2008, pp. 768–775.
48. Henningsen A. Estimating censored regression models in R using the censReg Package. *University of Copenhagen*, 2010.
49. Cobb L. Parameter estimation for the cusp catastrophe model. *Behavioral Science*, Vol. 26, 1981, pp. 75–78.
50. Maddala G.S. *Limited-dependent and Qualitative Variables in Econometrics*. Cambridge University Press, New York, 1983.