Accident Analysis and Prevention 112 (2018) 30–38

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Application of Fractal theory for crash rate prediction: Insights from random parameters and latent class tobit models

Sai Chand[a,*], Vinayak V. Dixit[b,c]

[a] *Research Centre for Integrated Transport Innovation (rCITI), School of Civil & Environmental Engineering, University of New South Wales, Sydney, NSW 2052 Australia*
[b] *Research Centre for Integrated Transport Innovation (rCITI), School of Civil & Environmental Engineering, University of New South Wales, Sydney, NSW 2052 Australia*
[c] *IAG Research Centre, IAG Research Labs, Sydney, NSW 2000 Australia*

ARTICLE INFO

Keywords:
Congestion
Crash rate
Hurst exponent
Latent class model
Random parameters
Tobit regression

ABSTRACT

The repercussions from congestion and accidents on major highways can have significant negative impacts on the economy and environment. It is a primary objective of transport authorities to minimize the likelihood of these phenomena taking place, to improve safety and overall network performance. In this study, we use the Hurst Exponent metric from Fractal Theory, as a congestion indicator for crash-rate modeling. We analyze one month of traffic speed data at several monitor sites along the M4 motorway in Sydney, Australia and assess congestion patterns with the Hurst Exponent of speed (H_{speed}). Random Parameters and Latent Class Tobit models were estimated, to examine the effect of congestion on historical crash rates, while accounting for unobserved heterogeneity. Using a latent class modeling approach, the motorway sections were probabilistically classified into two segments, based on the presence of entry and exit ramps. This will allow transportation agencies to implement appropriate safety/traffic countermeasures when addressing accident hotspots or inadequately managed sections of motorway.

1. Introduction

Reducing congestion and accidents are the primary goals of any transport agency. These two issues are interdependent on one another as significant improvements to one could result in substantial impacts on the other (Chang and Xiang, 2003). Furthermore, relieving congestion could also reduce travel delays, such that air quality and economic productivity can be improved. The connotation between congestion and road safety is always a debated issue, by transport planners and safety experts. Some studies argue that the increased level of traffic congestion reduces traffic speeds and therefore leads to less severe crashes. While that may be true, there will also be an increase in the likelihood of exposure and the number of potential conflicts in the congested conditions. This may lead to more crashes, although of a less severe nature (Quddus et al., 2009).

Existing studies explore the applications of different statistical models in estimating crash frequencies, crash rates, and injury severities at specific locations. For example, count data models such as Poisson and Negative Binomial models were extensively used to estimate crash frequencies (Shankar et al., 1995; Famoye et al., 2004; Bhat et al., 2014); Tobit model was used to study the accident rates (Anastasopoulos et al., 2008); the Mixed Logit (Moore et al., 2011), the Ordered Logit and the Ordered Probit models (O'Donnell and Connor (1996), Kockelman and Kweon (2002) were used to predict injury severity. These studies considered road geometric variables (ramps, shoulder width, and gradient), pavement characteristics (roughness, rutting, and friction), environmental factors (rainfall, cross wind speed, and snow) and traffic variables (AADT, posted speed limit and the heavy vehicle proportionality). Traffic variables were considered a proxy for traffic congestion in most of the earlier studies. However, modeling the aggregated crashes at a road segment level with such proxies for congestion may obscure the actual relationships (Quddus et al., 2009). In this regard, some researchers have used disaggregated crash records and a measure of traffic congestion during the accident period (Zheng et al., 2010; Ahmed et al., 2012; Yeo et al., 2013). Nevertheless, studies of this category require extensive traffic and crash data, which can be computationally cumbersome.

Within this study, we apply the Hurst Exponent, a metric from Fractal Theory, as a congestion indicator, as reported by Chand et al. (2017) for crash-rate modeling. We analyze one month of traffic speed data at several monitor sites along the M4 motorway in Sydney, Australia and assess the congestion patterns in terms of the Hurst Exponent of speed (H_{speed}). Finally, we use Random Parameters and Latent Class Tobit models to estimate the effect of congestion (in addition to other

* Corresponding author.
E-mail addresses: saichand.chakka@unsw.edu.au (S. Chand), v.dixit@unsw.edu.au (V.V. Dixit).

https://doi.org/10.1016/j.aap.2017.12.023
Received 26 June 2017; Received in revised form 29 October 2017; Accepted 30 December 2017
0001-4575/

fundamental variables) on long-term crash rates. The detailed procedure is discussed in the following sections.

2. Methodology

2.1. Fractal theory and the Hurst exponent

Fractal Theory, introduced by Benoit Mandelbrot (1967) is useful for studying irregularities in a time series (Mandelbrot, 1967). Objects that can be viewed with a similar appearance at different magnification scales are called fractals. While classical geometry deals with objects of integer dimensions (1-d lines and curves, 2-d squares and circles, etc.), fractal theory deals with objects of non-integer dimension, called Fractal Dimension. This dimension depends on the complexity of the shape, i.e., a shape with a higher fractal dimension is more complicated or rough, than one with a lower dimension and fills more space (Breslin and Belward, 1999). There are various methods to estimate the fractal dimension of a time series, such as the box-counting method, Hurst exponent and Higuchi method. However, a widely used practice for researchers is to calculate the Hurst Exponent using rescaled range (or *R/S*) analysis (Hurst, 1951; Bo and Rashed, 2004; Chand et al., 2016). Despite having several applications in many scientific fields, applications of fractal analysis of time series data is rarely explored in the transportation domain. Presently, research is limited to observing the fractal nature of traffic flow variables, rather than looking at potential real-world applications (Chand et al., 2017).

The step-wise procedure of the standard *R/S* analysis is shown below:

For a time-series, $X = X_1, X_2, ...X_n$

(1) Calculate average,

$$m = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1}$$

(2) Calculate mean adjusted series Y,

$$Y_t = X_t - m \quad t = 1,\ 2,\ ...,n \tag{2}$$

(3) Calculate cumulative deviate series Z

$$Z_t = \sum\nolimits_{i=1}^{t} Y_i \quad t = 1,\ 2,\ ...,n \tag{3}$$

(4) Calculate range series R

$$R_t = \max(Z_t) - \min(Z_t) \quad t = 1,\ 2,\ ...,n \tag{4}$$

(5) Calculate the standard deviation series S

$$S_t = \sqrt{\frac{1}{t}\sum_{i=1}^{t}(X_i - m)^2} \quad t = 1,\ 2,\ ...,n \tag{5}$$

(6) Calculate rescaled range series (R/S)

$$\left({}^{R}\!/_{S}\right)_t = {}^{R_t}\!/_{S_t} \quad t = 1,\ 2,\ ...,n \tag{6}$$

$(R/S)_t$ is averaged over the regions $[X_1, X_t]$, $[X_{t+1}, X_{2t}]$ until $[X_{(m-1)t+1}, X_{mt}]$ where m = floor (n/t). The value of t is normally chosen so that it is divisible by n.

(7) Finally, Hurst exponent (H) is the slope of the regression line that estimates the relationship between (R/S) versus t in log-log axes.

The following formula relates the Hurst exponent, H and fractal dimension, D.

$$D = 2\text{-}H \tag{7}$$

For a time-series,

i A value of H in the range 0.5–1 is indicative of a time series having a long-term positive autocorrelation. A high (or low) value will likely be followed by another high (or low) value in the series, i.e., the future trend is more likely to follow an established trend. For example, a very high H value (say $H = 0.9$) means a greater level of determinism (as shown in Fig. 1a) and easily predictable. Large H_{speed} values can be seen after the occurrence of an incident or removal of a bottleneck on a roadway. Incident does not necessarily imply an accident; vehicle breakdown, fire hazard, police pull over, sudden braking and lane changing in dense traffic can result in high H values.

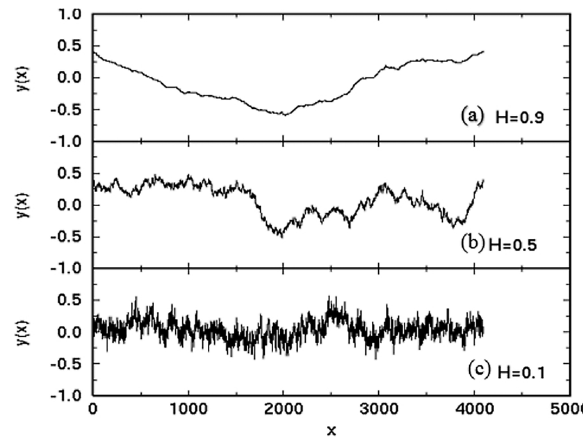

Fig. 1. Demonstration of the Hurst Exponent.
(Source: Cooperative Phenomena Group, n.d.).

ii H values close to 0.5 (Fig. 1b) indicates a completely uncorrelated series. The values in the time series are random and potentially indicating Brownian motion. It becomes extremely challenging to predict the future values for such time series.

iii H value of 0–0.5 suggests the long-term fluctuation between high and low values in the time series. A low H value (say $H = 0.1$) indicates a strong determinism and good predictability despite being volatile (Fig. 1c). A single high value will likely be succeeded by a small value or vice-versa. For example, small H values can be observed for traffic counts on downstream links at signalized intersections, when the measurement interval is shorter than the cycle time of the signal.

Fundamentally, large H values indicate weak dynamics, whereas small H values indicate frequent changes, i.e. high dynamics. From these phenomena, researchers have applied the Fractal Dimension and the Hurst Exponent for incident detection, short-term predictions, and accident warning models. In a recent study, researchers demonstrated that high H_{speed} could be used as a congestion indicator (Chand et al., 2017). They observed high H_{speed} values at several traffic monitor sites during weekdays and evening peak hours. Furthermore, they also observed a strong correlation between high H_{speed} and historical incidents. Reinforcing this premise, it was observed that strong dependence (high H_{speed}) would manifest after the occurrence of an incident. There can be multiple instances in a time-window where a location has high H_{speed}. These occurrences may not be limited to police-reported incidents such as; accidents, breakdowns and police stops, but also unreported incidents such as; abrupt braking and lane changing in dense traffic conditions, which were considered as surrogate safety measures in earlier studies (Gettman and Head, 2003; Lee et al., 2006; Oh and Kim, 2010; Bagdadi, 2013). Therefore, the frequent occurrence of high H_{speed} at a location, can indicate a potentially higher number of crashes.

2.2. Tobit model

There has been an enormous emphasis in past research on factors that determine the frequency of accidents (Anastasopoulos et al., 2008). Using exposure-based accident rates (continuous variable) instead of traditional accident frequencies (count variable) as the dependent variable has significant appeal because accident rates are typically used