Australasian Transport Research Forum 2010 Proceedings
29 September – 1 October 2010, Canberra, Australia
Publication website: http://www.patrec.org/atrf.aspx

Hot Spot Identification using frequency of distinct crash types rather than total crashes

Alex Sims[1], Dr. Sekhar V. C. Somenahalli[2]

[1] School of Natural and Built Environment, University of South Australia, Adelaide

[2]School of Natural and Built Environment, University of South Australia, Adelaide

Email for correspondence: alex@softgrow.com

Abstract

Hot Spot (or "black spot") identification is an important part of traffic engineering as it offers high safety benefits on investment and responds to community and media interest in places with high crash frequencies. One of the difficulties is the low frequency of crashes and the subsequent long periods of data collection needed to make statistically valid decisions.

Identification of crash hot spots by ranking the frequency of the predominant crash type at each site instead of total crash count was investigated using data from metropolitan Adelaide. This method was tested on both historical and simulated data and found to have poor predictive value and efficacy. The use of simulated data is found to be useful in investigating evaluation techniques for hot spot identification where long periods (years) of data are necessary for evaluation.

The use of techniques from statistical process control on the interval between crashes instead of the count of crashes in a period is also examined. This appears a promising tool for identifying trends in crash frequency and may faster to show changes than simple counts of crashes in a period.

1. Introduction

A fundamental problem for a practising Traffic Engineer is to allocate resources to improve a road network in general, when the resources are limited and there is no clear direction as to exactly where or how to use the resources. A technique for Hot Spot identification along with the use of control charts is explored in this paper to meet this end.

1.1. Hot Spot Identification

Hot spot identification is an important part of Traffic Engineering practice for two reasons, the community is focussed, particularly by media interest, on "black spots"; the identification, if possible of black spots, and their rectification is economically efficient. Governments are always constrained by funds and hot spots are attractive for remediation as they offer a greater return. Guerts and Wets (2003) warn though of diminishing returns in the identification and rectification of black spots as once the worst sites have been treated, the returns will diminish and eventually become uneconomic. They suggest that ongoing evaluation is necessary if treatment beyond the point of economic return is to be avoided.

How to find hot spots has been subject to much research and debate as they are a result of infrequent random events which exhibit low frequency and great variation as a consequence of their Poisson nature.

Maher and Mountain (1986) explored the possibility of allowing for the flow through an intersection or road segment when ranking sites, in that high traffic sites should not dominate the rankings, but that those that are empirically dangerous on an exposure basis should. They found however that the gains of the proposed criterion, potential crash reduction to be small in theory and may disappear in practice.

Cheng and Washington (2009) reviewed existing statistical methods for hot spot identification, then proposed new methods and measured their predictive performance. This is a key issue as any method must be capable of identifying the high crash locations that will continue to be so, not merely those that have random fluctuations. They found that accident (crash) frequency performed well in site consistency but that overall (with caveats) they recommended an Empirical Bayes approach. The Empirical Bayes method compares the crash history of each site with similar reference sites. The Empirical Bayes method was also supported by Elvik (2008), although only for use in mid-blocks. Miranda-Moreno et al (2007) developed another two Bayesian methods and show they can be used not only to produce a ranked list of hot spots, but also for identifying sets of sites worthy of further engineering investigation.

Simulation of crash data as used by Cheng and Washington (2005) can be used to experiment with methods for evaluating hot spots. In this method “ideal” data is generated entirely by random number generation and can be produced in any desired number of sites and time period.

GIS forms an important part of hot spot identification allowing both visualisation and management of data (Erdogan, Yilmaz, Baybura & Gullu, 2007). Lambert, Peterson & Joshi (2006) argue that GIS and crash data can be combined not only for short-term approaches to mitigate crashes but also for longer term planning.

1.2. Control Charts

An alternative approach is the use of control charts to find changes in crash rate instead of ranking sites by overall crash rate. The crash rate, either improving or worsening, could be used as a basis for investigation of a particular site to find what the newly introduced underlying improvement or defect might be. This is on the basis that an improvement could be investigated further for application elsewhere and a defect that can be “fixed”.

Control charts are commonly used for manufacturing processes when an object is produced by a manufacturing process and this process is monitored by measuring the number of defective objects or some critical attribute of the objects produced. This measurement can then be plotted on a control chart and when the line moves beyond either the upper or lower control limit the control limit is said to be “broken”. With a broken control limit a change to the process has more than likely occurred and the manufacturing process requires attention as it should work consistently to the same level of quality.

The same techniques could be applied to crashes at an intersection where the manufacturing process is the intersection and the defective product is when a crash occurs. However the relative infrequency of crashes brings with it the same problems for control chart production as for high-quality or near zero-defect processes. The issues around monitoring high-quality processes are dealt with by Pham (2006) at some length and detail.

By considering the length of time between events instead of the number of events in a period, information in the continuous data is not lost. For example the statement “six crashes in a year” conveys much less information than either six crashes evenly spaced throughout the year or alternatively clustered around a period of bad weather.

In this paper control charts have been generated for intersections for all crashes along with the predominant crash type as an exploration of the data and possible benefits that might be gained from their production.

2. Comparison of methods of Hot Spot Identification

Two methods of hot spot identification are to be compared; the first is a straightforward approach of considering the crash frequency at an intersection over a period of time. This method is proposed to be varied and further refined by partitioning the total crashes by type and then counting each type. It is postulated that this might make random variation less problematic by reducing and isolating the variability each of the underlying crash causes and their corresponding crash types.

In the first case it is assumed that the crash frequency *AF* over a period *Time* is given by the formula:

$$AF = \sum_{Time} f(V, v, x_1, x_2 \ldots) \qquad (1)$$

where *f()* is some random Poisson distributed function that is affected by variables *V* volume, *v* velocity and x_1, x_2… other variables. The function *f* would be unique to each site reflecting its geometry, location with the network and so on.

In the second case the crash frequency is given by an expanded formula

$$AF = \sum_{Time} \sum_{i=1}^{n} f_i(V, v, x_1, x_2 \ldots) \quad (2)$$

where $f_i()$ is some random Poisson distributed function for crashes of type *i* that is affected by variables *V* volume, *v* velocity and x_1, x_2… other variables. Each of the $f_i()$ functions for different crash types are summed to give the total crashes at that site. This is to build upon the approach for crash prediction as proposed Kim (2006).

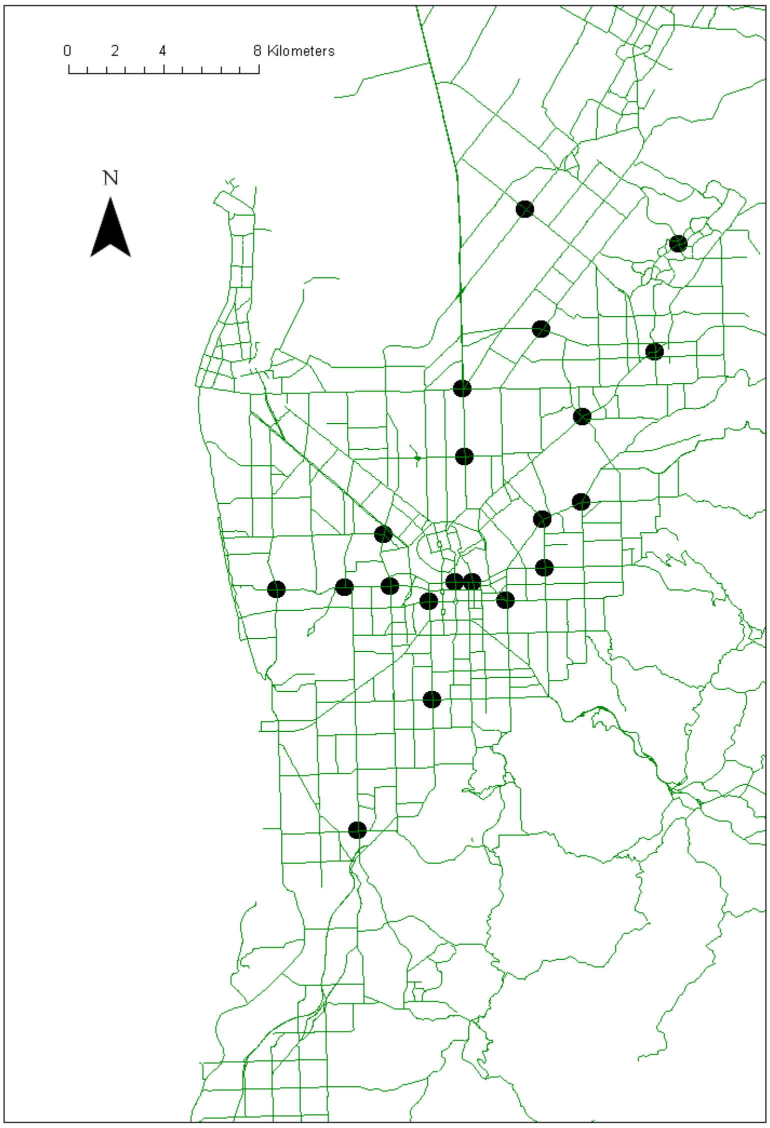
This should have an advantage where once a hot spot has been identified a treatment will be proposed that will correct the site to avoid a particular type or classes of crashes. By removing the “noise” of other crash types it should be easier to identify relevant sites. This offers an approach to use in relatively small networks that lack homogeneity to enable classification and comparison with reference sites.

2.1 Data selection

Crash data for Metropolitan Adelaide was obtained for the years 2001 to 2008. This data was partitioned into two data sets 2001-2004 and 2005-2008 with the intention of using the 2001-2004 data to identify the prospective hot spot sites and then test the predictive value against the 2005-2008 data set. Due to legislative changes in 2003 affecting the level at which property damage only crashes are included, these crashes were excluded from all data sets.

The top twenty intersections by total count of crashes over the first four year period (2001-2004), were then identified and are illustrated in Figure 1

Figure 1: Intersections Selected in Metropolitan Adelaide

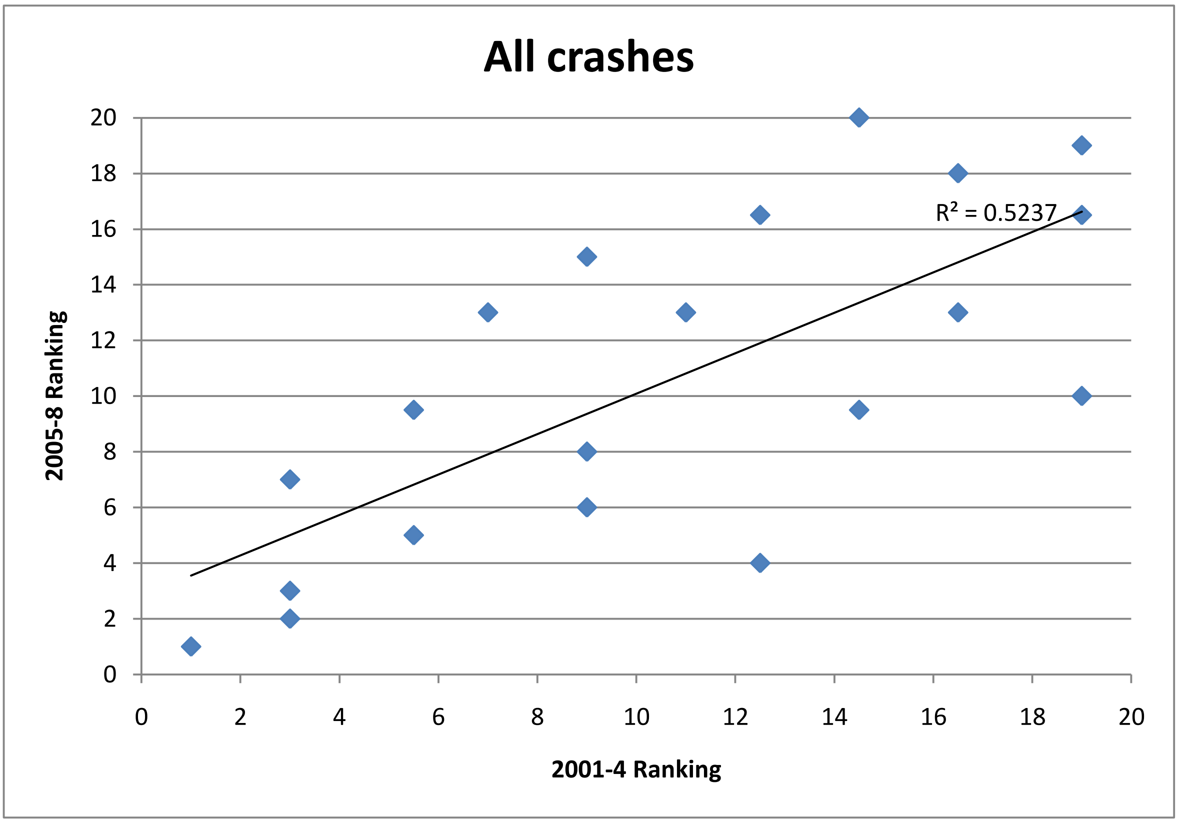


2.2 Ranking results

For each of the twenty sites the count of crashes in the 2001-2004 and 2005-2008 periods were calculated. A rank score was then calculated for each site and each period giving 1 for the site with the highest number of crashes and 20 for the site with the least number of crashes. If two or more sites had the same number of crashes then each of these sites was given the average ranking of the set, e.g. four sites ranked equally for ranking 2, 3, 4 & 5 would all be given the arithmetic average of 3.5.

The before and after rankings for each site were plotted against each and shown in Figure 2. If the count is a predictor of crash propensity and suitability as a hot spot, then the ranking should remain the same and be strongly correlated between periods.

Figure 2 Comparison of rankings for all crashes



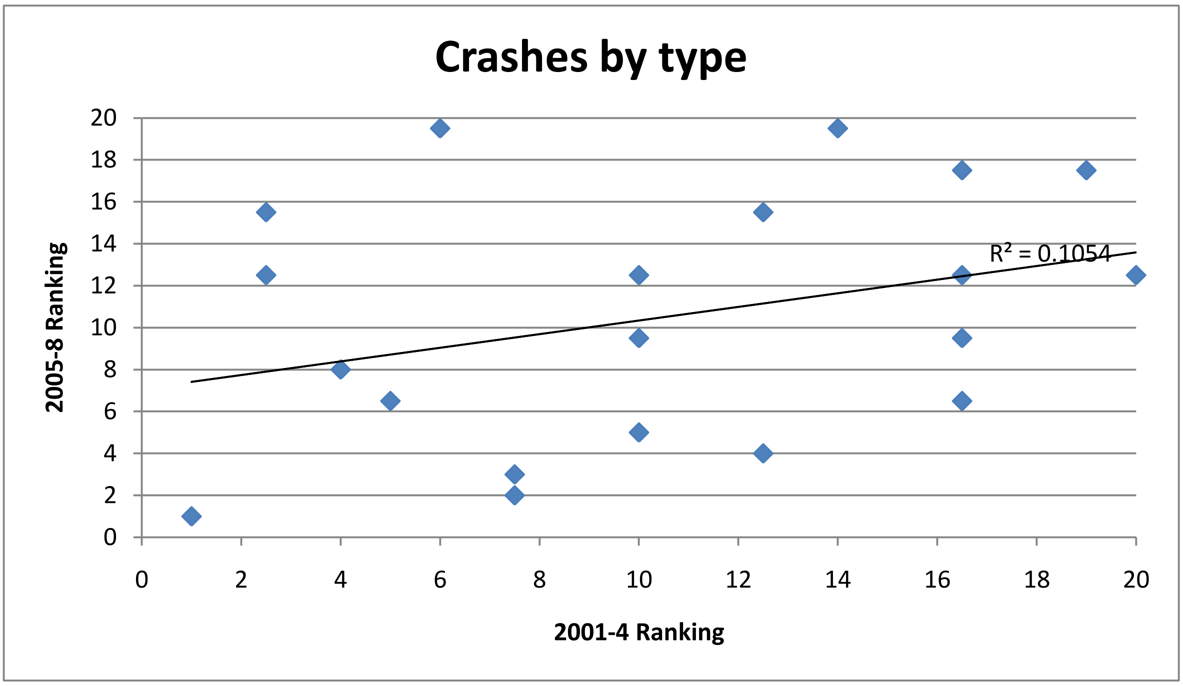
Then for each of the sites identified the count of the predominant crash type at each site in 2001-2004 was determined. The predominant crash type was determined to be the crash type with the most occurrences in the entire 2001-2004 period. The frequency of the type selected is shown in Table 1.

Table 1 Crashes by Site, Type and Year

Crash Type	Number of sites
Right Turn	9
Right Angle	1
Rear End	10

For each of the twenty sites the count of predominant crashes in the 2001-2004 period and the 2005-2008 period was calculated. A rank score was then calculated for each site as before and the before and after rankings for each site were plotted against each and shown in Figure 3.

Figure 3 Comparison of rankings for crashes by type



This shows that by considering crashes by predominant type had much less predictive power than the total crashes.

2.3 Simulation results

Given the unexpected results with the crash type data, a second set of crash data was generated by simulation. It was assumed that the number of crashes both in total and type would follow a Poisson distribution. The lambda(λ) of the distribution was assumed to be the average number of crashes in each four year period. A single set of crash counts was for each year 2001 through 2008 and then totalled and ranked as before. Thus a second set of tables and rankings was generated. The before and after rankings are plotted as Figures 4 and 5.

Figure 4 Correlation of Rankings All Crashes

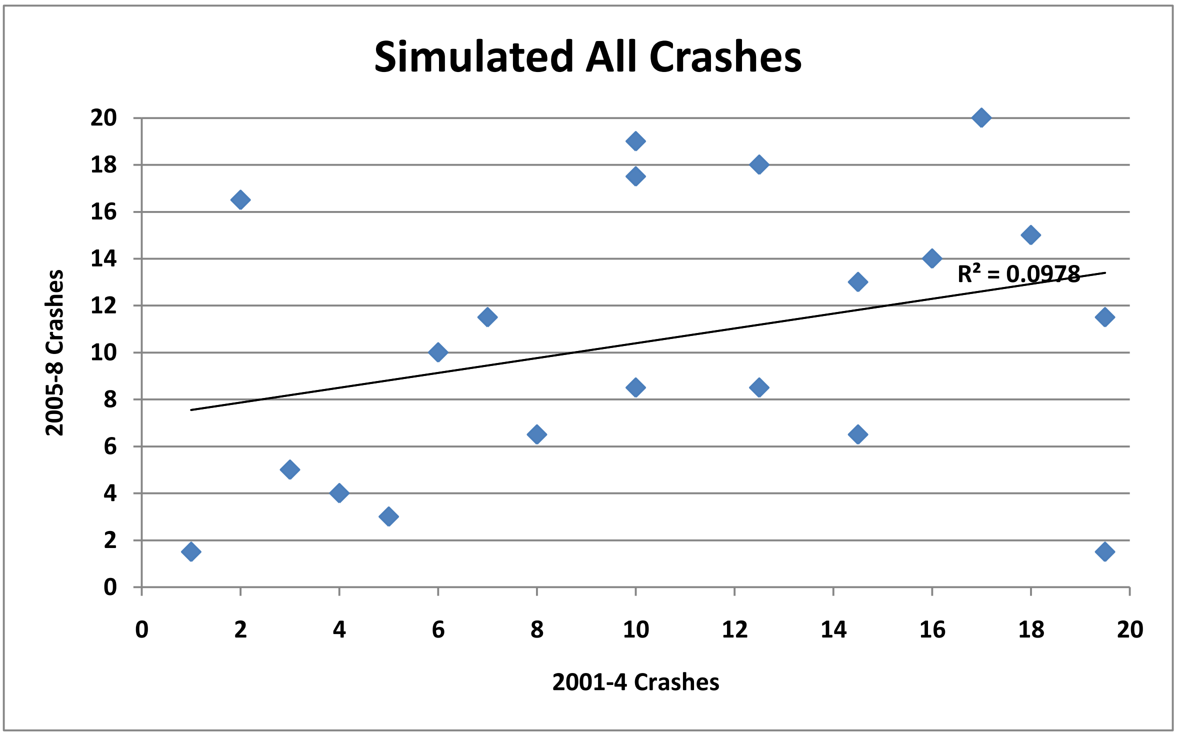
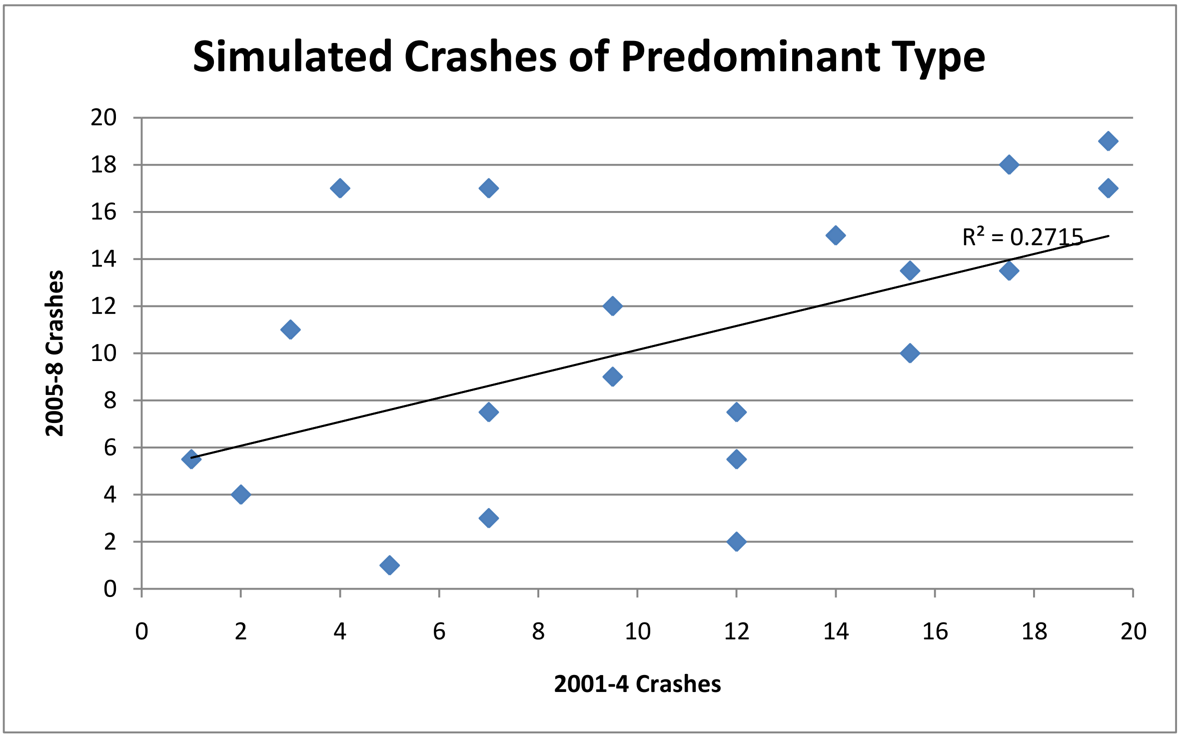


Figure 5 Correlation of Rankings Crashes of Predominant Type



The simulated crashes show that the predictive value from period to period is poor, and suggests the possibility that the underlying distribution of crashes is not Poisson. To test this the period between crashes in 2001 to 2004 for two sites were tested to see if this fitted an exponential distribution but neither could be rejected as not fitting the exponential distribution and hence cannot be rejected as not being Poisson distributed.

5. Use of control charts

For convenience, control charts were generated for only the eight Northern most intersections for the all crashes and the predominant crash type as selected by the above method of Hot Spot identification.

For each chart the control limits were calculated assuming that the number of crashes was generated by a Poisson process of which the period 2001-2004 was a representative sample. The mean occurrence rate (λ) in crashes per day is calculated according to equation 3 below where C_y is the number of crashes at the intersection in the year *y*

$$\lambda = \frac{\sum_{y=2001}^{2004} C_y}{365 \times 4} \quad (3)$$

Then the interval $I_{j,j+1}$ in days and fractions of days between each crash was calculated and the probability of an interval of at most that many days was calculated according to the distribution function in equation 4

$$F\left(I_{j,j+1}\right) = 1 - e^{-\lambda I_{j,j+1}} \quad (4)$$

This probability was then normalised (i.e. converted to a normal distribution) by finding the value of *z* for which the area under the standard normal distribution is equivalent as in equation 5.

$$F\left(I_{j,j+1}\right) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \quad (5)$$

This can be readily calculated using a spreadsheet and a formula like “NORMSINV(1-EXP(-B5*C5))”. This then produces a series of values where if they are between plus or minus three they can be considered to be “in control” as the possibility of them being outside the three sigma limits is 0.027%, a standard value used for control charts.

For each intersection for all crashes and for the predominant crash type a control chart was generated and examined for the signs of the process moving and breaking its control limits. This would indicate some improvement or worsening of conditions at the intersection.

5.1. Salisbury Highway and Kings Road Intersection

Figure 6 below shows all crashes at this intersection and Figure 7 for the predominant crash type. The road authority was consulted and there were no phasing or geometry changes in the period 2005-2008. The first chart shows a series of points centred around the zero line, not reaching plus or minus three indicating a relatively stable process.

Figure 7 is more interesting appearing to show a process shift where all of the points (bar one) are on one side of the zero. This would indicate a strong likelihood that the mean has changed and is moving in a positive (fewer crashes) direction.

Figure 6 Salisbury Highway and Kings Road Intersection, All Crashes

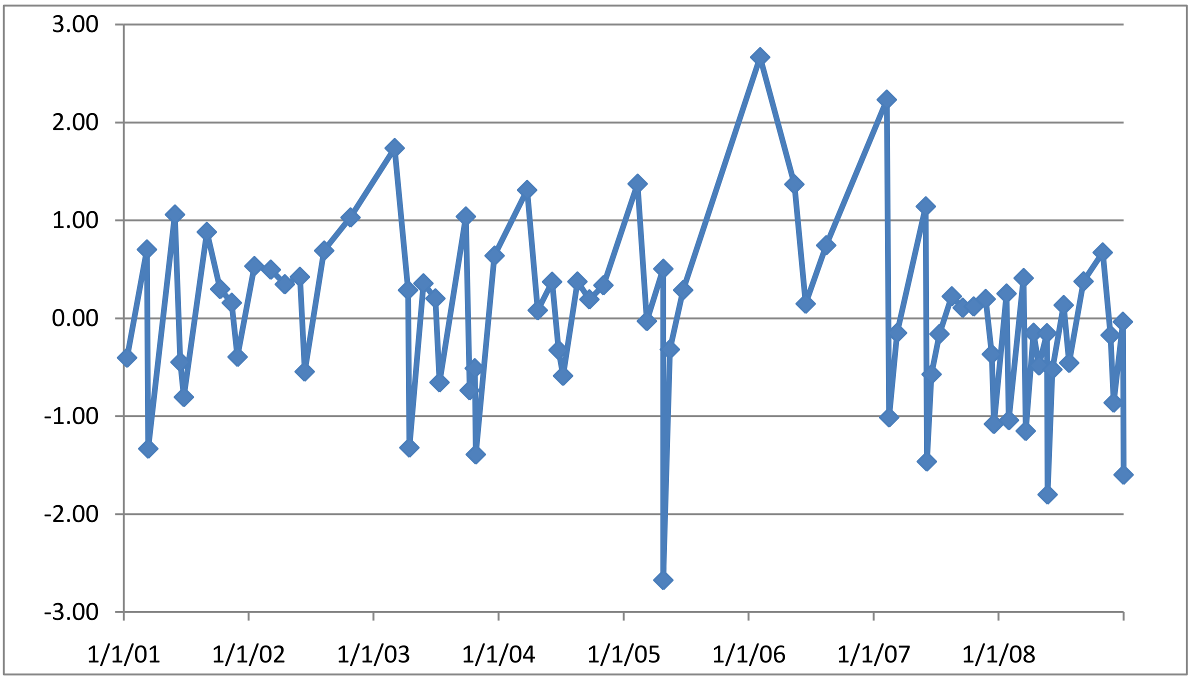
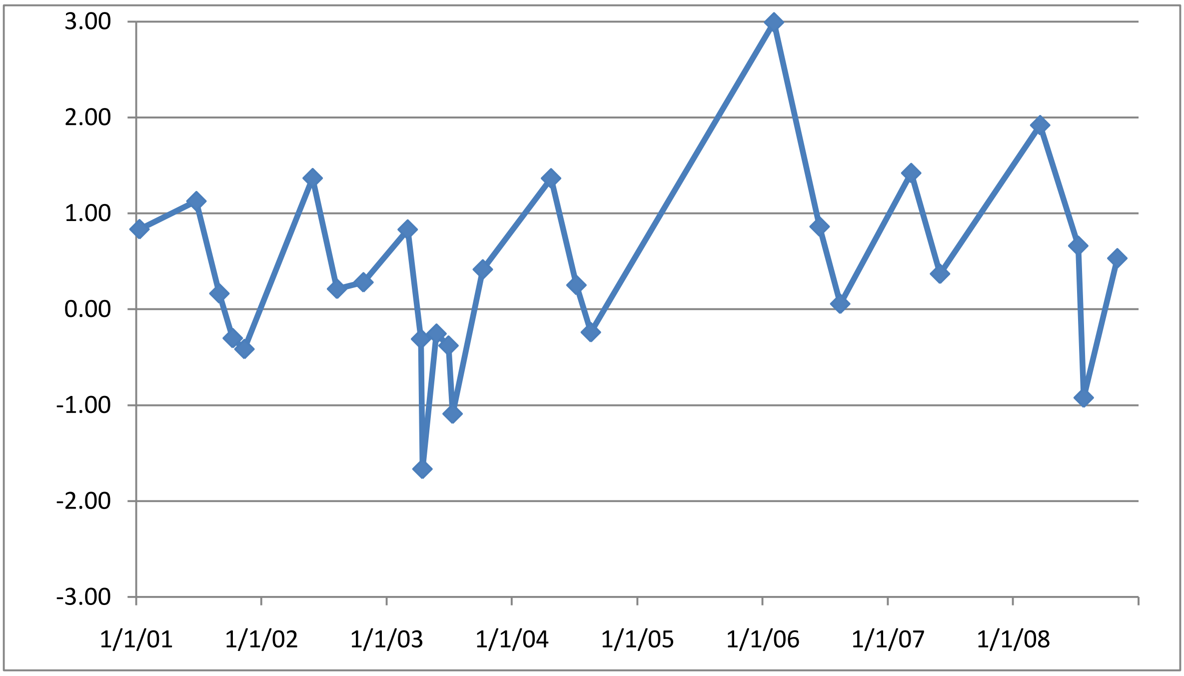


Figure 7 Salisbury Highway and Kings Road Intersection, Right-Turn Crashes



Examining the year by year crash data for the dominant crash type in Table two below appears to show a reduction in right turn crashes after 2005 but the ease of discernment is lower than that of the control chart.

Table 2 Crashes by Type and year, Salisbury Highway and Kings Road Intersection

Crash Type	2001	2002	2003	2004	2005	2006	2007	2008
Right Turn	5	3	7	3	0	3	2	4
All	10	7	11	8	6	4	13	17

5.2. Gepps Cross Intersection

Figure 8 below shows all crashes at this intersection and Figure 9 for the predominant crash type. The road authority was consulted and there were no geometry changes in the period 2005-2008. Phasing was altered on 6th December 2004 to ban right turns and remove an arrow in favour of an opposing through movement and again on 25th November 2008 to allow a left turn movement during an opposing right turn phase. The first chart shows a series of points centred on the zero line, not reaching plus or minus three indicating a relatively stable process. Figure 9 is similar, also showing a similar stable process. Of note is that latter part of the graphs has a different shape indicating that the rear end crashes and non-rear end crashes could have different underlying causation.

Figure 8 Gepps Cross Intersection, All Crashes

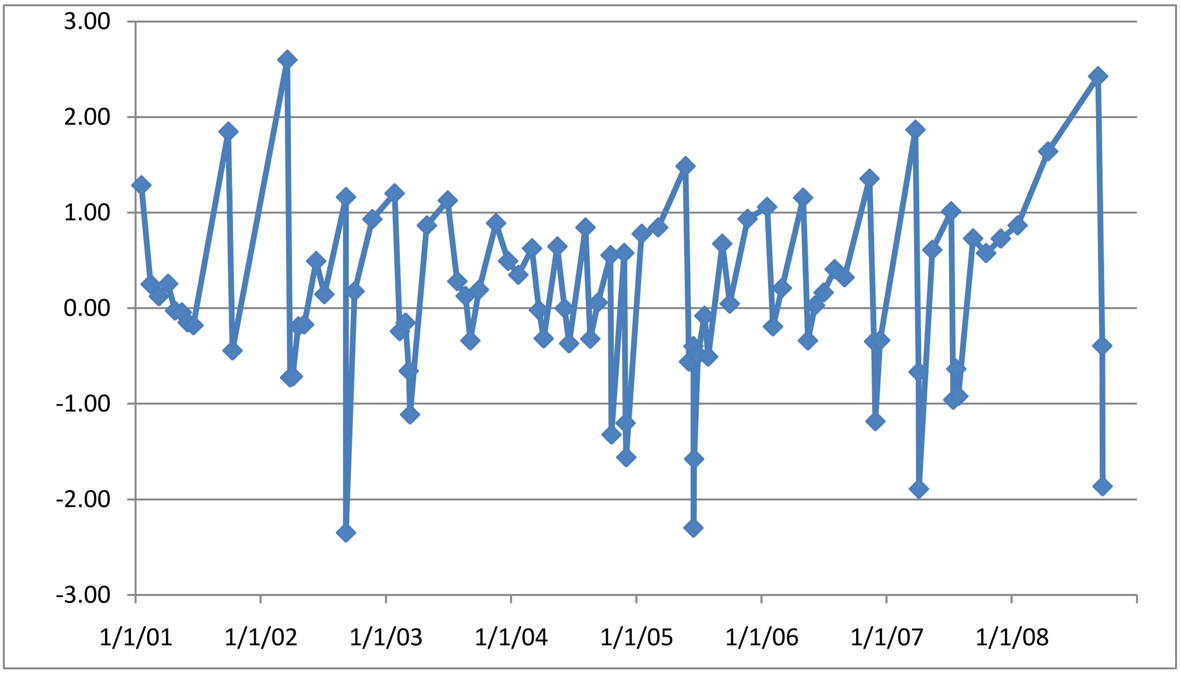
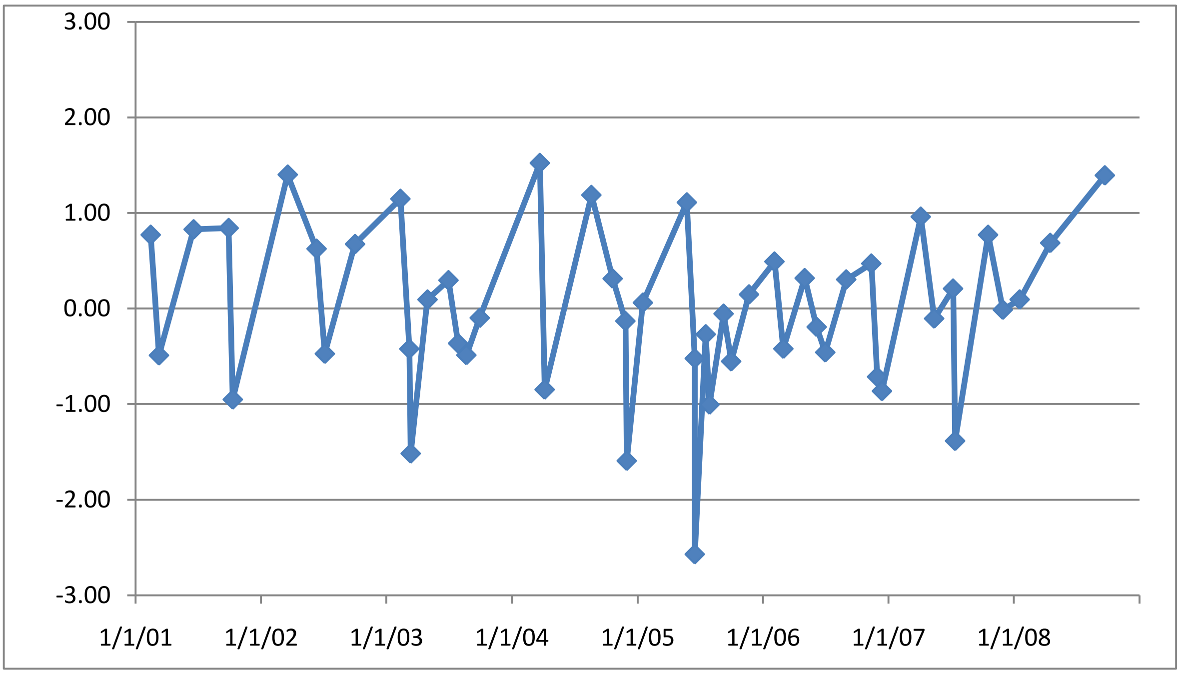


Figure 9 Gepps Cross Intersection, Rear End Crashes



These results are typical for most of the other intersections

5.3. North East and Sudholz Road Intersection

This third example has Figure 10 below shows all crashes at this intersection and Figure 11 for the predominant crash type. The road authority was consulted and there were no geometry changes in the period 2005-2008. On 17th September 2007 filtered right turns outside peak hours were removed in favour of controlled turns only. The first chart shows a series of points centred on the zero line, in some cases reaching or exceeding plus or minus three indicating an unstable process whose mean is not moving. Towards the end there appears to be a movement of the mean where all points are above zero indicating a possible decrease in crash frequency. Figure 11 has fewer deviations beyond plus or minus three indicating a more stable process. It has similar features to the all crashes chart with a pronounced period below the zero line and then a similar rise at the end.

This process shift would be consistent with the phasing change made in September 2007, confirming its positive effect.

Figure 10 North East and Sudholz Road, All Crashes

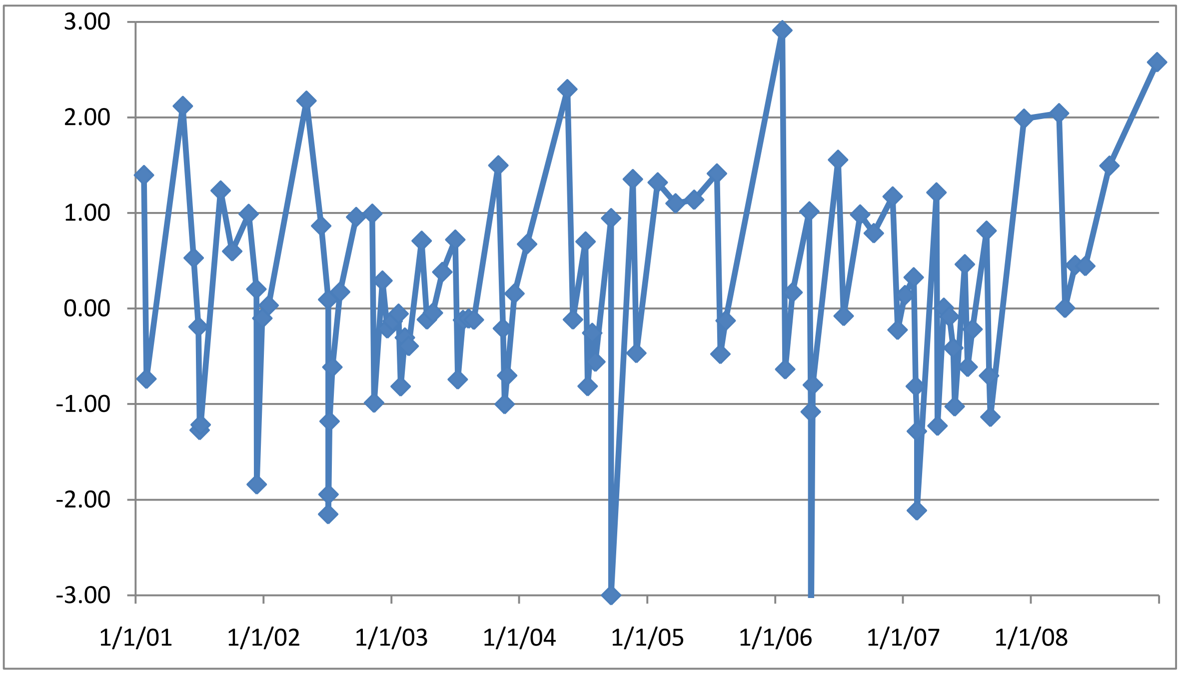
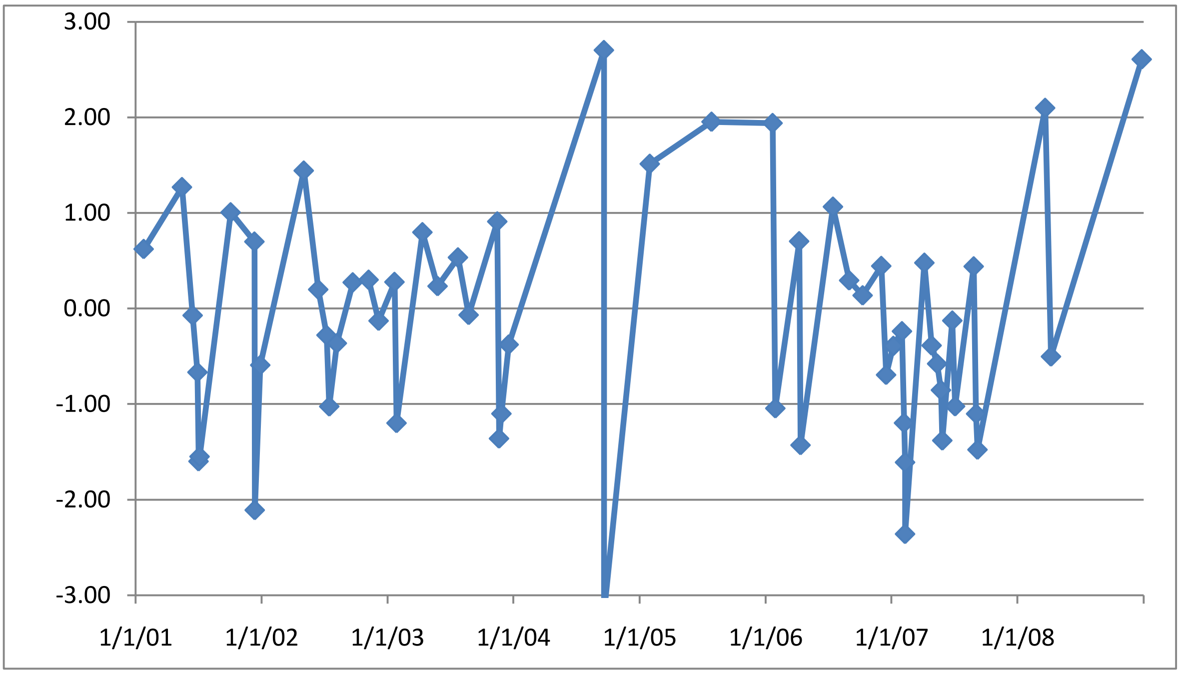


Figure 11 North East and Sudholz Road, Right Turn Crashes



6. Discussion and conclusions

It has been shown that identifying hot spot locations by crash-type is not necessarily a useful exercise, with the total number of crashes in a site appearing to have much greater predictive power. It is proposed that this variation is due to crashes being caused not entirely by the road environment and its physical factors but also by behavioural issues. There could be a motivation for drivers to move dangerously through an intersection to meet some other need, e.g. to be on time for an appointment. As the intersection is modified to prevent crashes of one type, the underlying behaviour remains the same so the crash type will simply alter. The only way to avoid this effect is to provide grade separation and other physical barriers that completely remove conflicts points between vehicles.

The use of simulation to test models has also been shown to be useful in investigation of methods where there is much underlying random variation. Simulation allows a statistical technique to be investigated without having to wait years for data to accrue. The simulation results suggest that it is difficult to rank sites with similar crash rates and a more useful approach may be to group sites by ranking and then check for consistency of grouping over time.

The use of techniques from statistical process control by following the interval between crashes and identifying possible process changes using control charts appears promising, providing a tool that might indicate trends sooner than other techniques. This could be useful in both identifying sites that have become more dangerous for whatever reason, along with confirming the efficacy of engineering changes. The use of control charts of crash data should be investigated further by doing this over a much larger set of sites.

References

Cheng, W, Washington, S 2005, ‘Experimental evaluation of hotspot identification methods’, Accident Analysis and Prevention, vol. 37, pp 870-881.

Cheng, W, Washington, S 2008, ‘New Criteria for Evaluating Methods of Identifying Hot Spots’, Transportation Research Record: Journal of the Transportation Research Board, vol. 2083, pp. 76-85.

Kim, D 2006, ‘Modeling Crash Types: New Insights into the Effects of Covariates on Crashes at Rural Intersections’, Journal Transportation Engineering, vol. 132, No 4, pp 282-292.

Elvik, R 2008, ‘Comparative Analysis of Techniques for Identifying Locations of Hazardous Roads’, Transportation Research Record: Journal of the Transportation Research Board, vol. 2083, 2008, p. 4.

Erdogan, S, Yilmaz, I, Baybura, T, Gullu, M 2008, ‘Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar’, Accident Analysis and Prevention, vol. 40, pp 174-181.

Guerts, K, Wets, G 2003, Black Spot Analysis Methods: Literature Review, RA 2003-07, Flemish Research Center for Traffic Safety, Diepenbeek, Belgium, available online at <http://www.roadsafetyresearch.be/nl/modules/publications/store/13.pdf>.

Lambert, J, Peterson, K, Joshi, N 2006, ‘Synthesis of quantitative and qualitative evidence for accident analysis in risk-based highway planning’, Accident Analysis and Prevention, vol. 38, pp 925-935.

Maher, M, Mountain, L 1988, ‘The Identification of Accident Blackspots: A comparison of Current Methods’, Accident Analysis and Prevention, vol. 20, No 2, pp 143-151.

Miranda-Moreno, L, Labbe, A, Fu, L 2007, ‘Bayesian multiple testing procedures for hotspot identification’, Accident Analysis and Prevention, vol. 39, pp 1192–1201.

Pham, H 2006 ‘Springer Handbook of Engineering Statistics, pp 281-290, Springer, London.