A NETWORK SCAN OF HORIZONTAL ROAD GEOMETRY

D. O'Connor
National Roads Authority, Ireland
DOCONNOR@NRA.IE

ABSTRACT

This paper examines GPS data collected from road condition surveys carried out on approximately 2,500 kilometres of largely unimproved roads in Ireland. Such surveys are carried out periodically across the network. The analysis exploits some of the underused data collected during these surveys, principally speed data. By utilising this GPS data, it was possible to detect significant speed differential along routes at specified intervals. This allowed for inconsistencies in the horizontal road alignment to be identified. The volume of data that is currently accessible facilitated a network scan to be conducted on 2,570 km of national secondary roads in Ireland. Collision and traffic flow data were processed to calculate collision rates for locations characterised by large speed differentials associated with challenging horizontal alignment. In addition horizontal route consistency was assessed using a sinuosity value. These two results were used together to show how horizontal road curves can be classified in the absence of detailed geometry information.

The results show that rural roads, characterised by isolated bends, are the greatest to the road user. The variance of the observed speed differential is not as critical a contributory factor of collision risk as originally anticipated.

1. INTRODUCTION

Speed recorded along a route can be used to calculate zones of acceleration and deceleration. Speed differential related to design quality have been discussed by Lamm et al. [18] amongst others [14], [29]. Speed differentials of less then 10 km/h are considered good quality while 10 to 20 km/h are acceptable. Anything in excess of 20 km/h is an indication of poor design quality. In general the larger the difference, then the probability of safely traversing the curve reduces. The probability of a collision increases if the speed reduction is sudden. A sharp reduction in speed is likely to occur when a solitary curve is encountered after a long straight section of road [3].

By measuring the differential speed along routes, a reasonable indication of where any abrupt change in the horizontal road geometry takes place, can be observed. de Waard [9] demonstrates how rural roads can change the mental workload experienced by the driver as the road environment becomes more challenging. Figure 1 shows how demanding the environment can be to a driver regardless of road geometry inconsistencies.

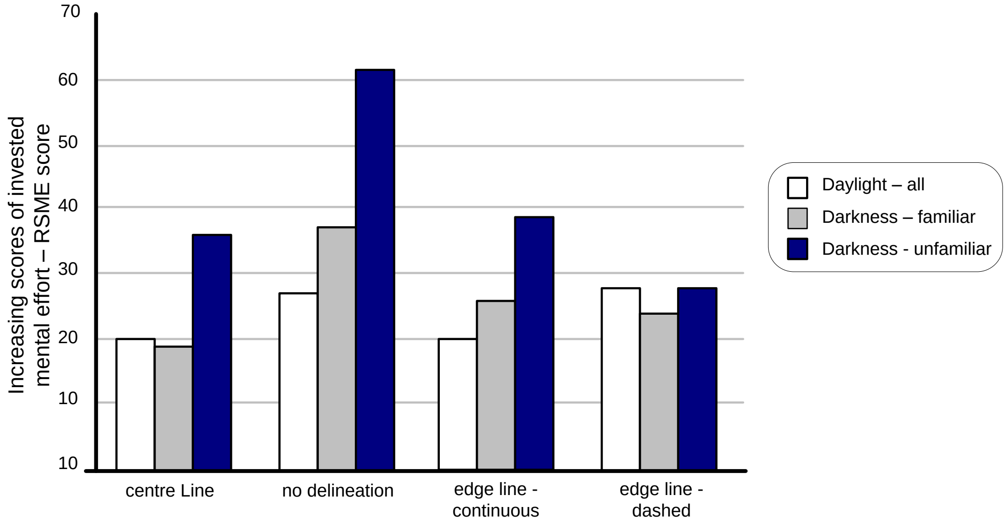

Figure 1 - Results from de Waard (2002) showing mental workload on differently delineated rural roads [9]

It has been put forward by Smith et al. [26] that drivers experience concentration difficulties on lower demand roads rather then high demand roads. In addition it was noted that the transition from high to low demand and vice versa are areas where collisions can occur. This can be attributed to drivers failing to cope with the changing driving demands.

Of the 220 fatal collisions that were recorded on Irish roads in 2008 some 66 per cent occurred outside urban areas. Of these collisions, 30 per cent were reported to have occurred where the main character or the road was described as a curve [25]. International, studies [7], [18], [21] have shown a similar situation with significant number of collision occurring at bends on rural roads. Although the number of road collision has reduced on the rural road network over recent years, the percentage of fatal collisions occurring has grown to 30 per cent. Examining earlier collision figures for Ireland [2] it has shown that the percentage of collisions occurring on curves has been lower, at 20 per cent.

In response to the current situation an analysis, scanning the complete secondary road network, was undertaken by the National Roads Authority (NRA). This paper describes the work undertaken, to establish the extent of significant speed differentials, on the national secondary routes. These routes are largely comprised of unimproved rural roads. On comparing the findings with the published literature, a robust estimate of the quality of route, in terms of horizontal consistency, was made. The calculation of collisions rates for these roads was not in itself the goal. There are well documented issues with selecting road sections for treatment based on historical traffic and collision data. These mainly deal with the regression to mean effect [4],[28],[30] and concerns about underreporting of injury collisions [23]. The focus was on developing a methodology, using the data already in use within the NRA, to facilitate a network scan to be carried out that would assist in identifying routes comprised of poor horizontal alignment. Research has shown that "*62% of fatalities and 49% of other incidents occurring in curves, the first manoeuvre that led to the accident was made at the beginning or the end of the curve*" (Council cited in [3]). Therefore any additional information concerning the typical conditions experienced by drivers negotiating curves is beneficial in addressing these site specific incidents

Some work has been recently published [17], [29] specifically on Irish roads relating to horizontal geometry. This paper intends to expand on previous work in this area making use of the existing data that has been collected along the Irish road network, specifically GPS data.

The structure of this paper is set out as follows; Section 2 looks briefly at the available literature on the subject of identifying hazardous road sections etc. Data collection is dealt within Section 3 whilst Section 4 looks at how this data was used and the process it was put through to gain new information from these existing sources of data. This new information is presented in Section 5. The final section, Section 6 discusses this information and provides conclusions based on the work conducted.

2. LITERATURE REVIEW

To ascertain current practice in analysing and quantifying inconsistencies in horizontal alignment within a road network, the contemporary literature on the subject was reviewed. This was carried out with the understanding that gathering new data was not possible and whatever methodology was followed it necessitated the use of the existing road related datasets currently accessible within the NRA. The literature review examined papers and reports, specific to Irish roads, on the subject of collisions and road geometry as well as looking to the international research.

Work in relation to identifying roads of similar types, and establishing an associated safety risk for comparative purposes, has been detailed and widespread [10], [12], [27]. However, due to the often dispersed nature of published road safety research, some published without sufficient appraisal, it is difficult to establish what approach to take to ensure best practice is followed. Consequently "*costly research findings may go unused, valuable experience may be overlooked, and full consideration may not be given to available practices for solving or alleviating the problem*" [29].This section looks at some of the available research on the subject of assessing risk to the road user through collision analysis, as well as looking at horizontal curvature and its contribution to road safety. The processing of GPS data and the concepts around data mining are also addressed within the literature review.

2.1. Collision analysis

Typically there are three basic types of collision analysis that can be carried out to assist in indentifying hazardous locations within a road network. These include the calculation of collision frequencies at sites along the road network. This approach is more usually referred to as 'black spot' analysis. When a number of collisions occur in excess of a predetermined number over a period of time the site is considered hazardous. These hazardous sites are then reviewed in more detail to see what appropriate countermeasure can be put in place. The second analysis type involves factoring in an exposure to risk indicator, such as traffic volumes to allow for the calculation of collision rates. A site or section of road may be considered hazardous if the rate is in excess of the average rate for a particular type of road. The final approach is to identify hazardous sites based on a topology indicator. Typically this analysis attempts to isolate groups of similar types of collisions based on similar environmental or infrastructure contributory factors [16].

A key goal in road safety analysis is to identify sites capable of improving their safety performance rather than just the sites with the highest number of collisions. This requires the analyst to define reference populations. These populations are used to subdivide the road network into homogenous road sections [4]. An example of a reference population would be all the motorways within a network. It is reasonable to assume that motorways have been designed to the same standard and driver behaviour would be consistent throughout the motorway network. That is, we expect the road geometry element of any motorway to be consistent from beginning to end, with, for example, an absence of learner drivers and a prohibition on stopping in the hard shoulder etc.

Desktop analysis of collision data can increase the pool of knowledge available to be drawn upon when devising a solution to a specific road safety problem. This solution should be the most effective intervention aimed at eliminating, or at least reducing, the severity of any future collisions at a specific location, identified through analysis of the relevant data. One of the key pieces of relevant data is location details and the surrounding road geography. As expressed by Elvik [11], geography can assist in both the identification and diagnosis phase of collision analysis. Some road safety issues can be identified by "*a distinct geographic dimension … typically problems related to the quality of infrastructure*" [11].

In classifying road types into homogenous sections, the analyst makes the assumption for practical reasons that "*all the sites of the same reference population would have the same long term average accident frequency"* [4]. Therefore any fluctuation in the numbers of collisions is either attributable to the random nature of collisions or evidence of an actual decrease or increase in road safety due to localised conditions. Consideration of how to deal with the 'regression to mean effect' should be considered when selecting a site for remedial safety treatment. As reported by SWOV [27] there is no absolute approach to employing a correction method for the regression to mean effect. Controlling the effect by examining untreated sites at similar locations is one solution. However, in practice this is unacceptable. The identified control group may have had previous safety works carried both in terms of engineering and enforcement programmes. Also it would raise serious moral and legal issues if known hazardous locations were left untreated for the sake of addressing the regression to mean effect.

2.2. Assessment of horizontal curves

"*The relative safety of a curve depends not only on its design speed but also on how it fits into its local speed environment. Signage and other curve treatments should reflect the relative safety of the curve not just the maximum desirable speed*." [7]. Therefore it is as important to describe a curve in the context of the road immediately adjacent to it, as well as describing the curves individual horizontal geometry. The availability of detailed road alignment data for unimproved rural roads in many cases isn't readily available [29]. In practise surveying and collecting voluminous amounts of road geometry data is both costly and time consuming [13]. Some work [24] has been conducted, in the absence of detailed road geometry data, on establishing the curvature of a road using a measure of sinuosity. The ratio of the length of a section of road and the shortest path between the end-points of the same road section is defined as sinuosity index (SI). For a level, straight section an SI value of one would be expected. The SI value increases as the road becomes more interspersed with curves.

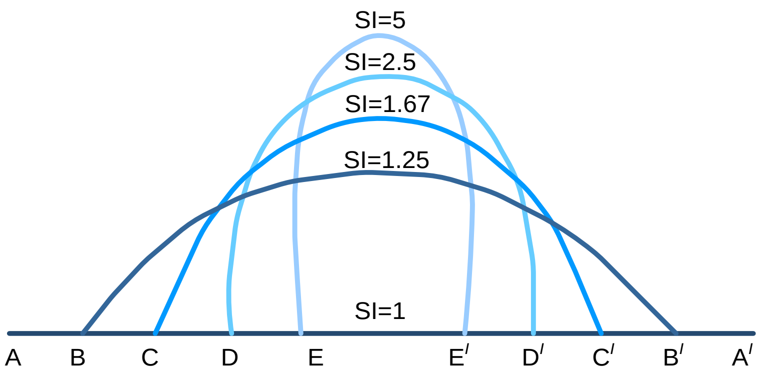

Figure 2 – Schematic diagram depicting increase of sinuosity index (SI) with increasing sinuosity (curvature) of a given road stretch. SI values for road stretches of equal length increase with reducing aerial distance between their ends (A-A/, B-B/, C-C/) [25]

2.3. GPS data and data mining

"Successful applications of data mining are not common, despite the vast literature now accumulating on the subject. The reason is that, although it is relatively straightforward to find a pattern or structure in data, establishing its relevance and explaining its cause are both very difficult problems." [5]. To explain this further it is worth considering the work carried out by Chen et al. [8] on data mining of text fields within a large database (65,56 records) from a car insurance company. Their approach, while novel, lacked the awareness of the inherent bias within the database. The rules they applied focused in on the driver age, the vehicle age and the presence of key words, such as lost control, narrow, wet, sharp, gravel. A frequency count of key words was then associated with a particular type of crash. The results of this data mining showed that the highest risk factors that contribute to collisions at bends can be classified as those drivers with 25 to 42 years driving experience and that drive a relatively new vehicle [8]. Although a pattern was discovered through this data mining analysis, these findings say as much about those individuals who claim on car insurance from a particular insurer, rather than the contributory road collision factors at bends on rural roads.

Despite the assertion by Buttenfield et al. [8] that data mining is not a straightforward process it is something most of use do on a daily basis without any need to resort to some black box technology [31].

The increasing use of real time GPS systems and the other spatial data collections systems such as remote sensing have allowed substantial amounts of spatial data to be collected. New approaches to collision analysis are required to realise the sometimes difficult to discerned information stored within sets of data routinely collected by police and road authorities [8]. Classification is one of the ways in which information is discovered within large volumes of data. Classification of data can be carried out based on rules. The rules themselves must be devised by the individuals with sufficient knowledge to interact with the data, often referred to as supervised machine learning. It is 'supervised' as it necessitates the individual or analyst to identify a field within the data containing attributes to be mined [32].

Other research supports the use of GPS data in establishing places of significance [1] or systems to warn of excessive speed as part of car navigation systems [19]. There is now a growing view that GPS is not only fast and reliable in gathering road geometry [6] but is often the most cost effective solution for capturing important road related geometry. In practice referring to a set of detailed 'as-built' road designs is not an option as many of rural roads were never designed to a standard. GPS data is an excellent primary data source. GPS receivers are becoming more accurate and cheaper. The GPS signals from the constellations of GPS satellites are free and accessible everywhere in the world. In many ways GPS, for this type of analysis, is superior to secondary sources of route data such as cartographic maps.

3. DATA COLLECTION

Road condition data is recorded during annual surveys carried out on behalf of the NRA by means of a vehicle equipped with a Road Surface Profileometer (RSP). All the data gathered during the road survey is GPS referenced. The RSP/GPS operates at 10 Hz and 5 metre intervals. So for every 5 metres, data is stores as a string of information within the Global Positioning System Fix Data (GPGGA) file. This is one of the international standards for storing GPS data and includes recorded records such as time, latitude, longitude, number of satellites in view etc, along with the data from the RSP. Of particular interest is the time (t) stamp recorded every 5 metres (d). From this basic data the velocity (v) of the vehicle can be calculated as …

$$V = d/t \qquad \text{eqn. (1)}$$

All the GPS data is post processes to correct GPS errors, known as differential GPS (DGPS). DGPS is use to account for and cancel out the small errors attributed to the receiver itself, atmospheric conditions and multipath errors (reflected signals from local obstructions).

The extent of this analysis was limited to the national secondary network. The national secondary network consists of over 2,690 km. Due to some processing errors the N51 and the N78 have been removed for the dataset. This brings the total extent of the network under analysis to 2,570 km. It was assumed that these omissions would not affect the overall result.

4. DATA PROCESSING

Simple classification of data is a common data mining technique. The methodology followed for this classification is rule based. For the purposes of this analysis a rule was devised to establish the difference in velocity or speed (Sd) of the survey vehicle as recorded by the GPS data captured along the route. A number of assumptions have been made. It has been assumed that the survey vehicle has travelled unimpeded at a speed typical of the traffic on the route. Another assumption was that the weather conditions haven’t adversely affected the speed of the vehicle and that the speeds haven’t differed greatly over previous year’s data.

The differential of the vehicle's speed is given by

$$S_d = S_a - S_{a+v} \qquad \text{eqn. (2)}$$

Where S_a is the speed at a point and S_{a+v} the speed at a point along the route in the direction travelled offset by distance v. Where v is a constant value and represents the distance over which the differential occurs, based on research on brake reaction and vehicle breaking distances [3],[23]

Equation (2) was run several times in an iterative process to establish the beginning of the speed differential section, the end of the section (i.e. chainage descriptors). A standard spreadsheet application was used to execute this function in the form of a conditional IF statement. It was run until a pattern of speeds associated with breaking and accelerating emerged. The classification process filtered through approximately 520,000 records of speed events measured at 5 metre intervals over approximately 2,600 km of national secondary roads. The results of this data mining exercise showed that there were some 2,109 sections of road within the national secondary network with speed differentials in excess of 10 km/h.

The NRA maintain and update the route data periodically within a geographic information system (GIS) and the events related to these routes, such as road condition survey data, are stored as separate files. This data’s location within the network is described as a series of paired ‘chainage start’ and ‘chainage end’ records. This route event data can be manipulated both internally and externally of a GIS. Performing operations on this data in a spreadsheet, for example, does not adversely affect the reuse of this data within GIS if managed carefully. The GIS tools in use with the NRA allow for tabular data, holding route and chainage information, to be referenced onto a common network of roads. This is often referred to as linear referencing [20]. This ability to reference events, either point or linear, to a defined route is a key requirement for this type of analysis.

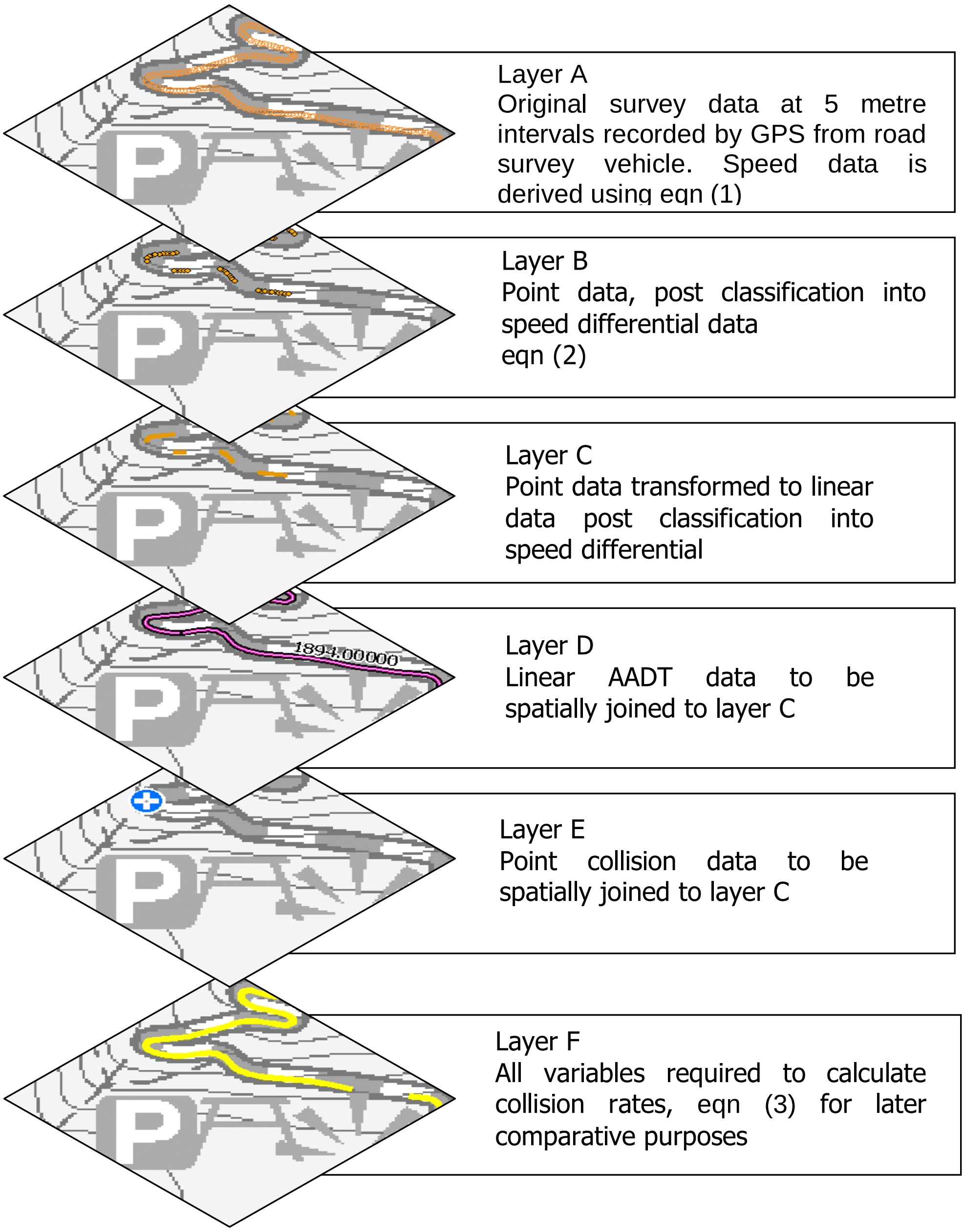

Figure 3 - Description of the GIS data layers produced as part of the data mining exercise

5. RESULTS

The results are presented in two sections. The first deals with the results from calculating collision rates based on speed differentials. The second section deals with the results from the sinuosity analysis of the national secondary routes.

5.1. Speed differential and collision rate results

The relatively short length of roads identified as having speed differentials in excess of 10 kilometres per hour was of concern when calculating collision rates. These sections (L_i) of road accounted for only 416 km or just 16 per cent of the national secondary roads. Therefore the rates are based on a relatively small dataset of collisions (1339). These 1339 collisions(f_j) occurred over a seven year period (P) from 2001 to 2007. Mid year traffic volume data (AADT 2004) was used when calculating collision rates. The calculation assumes the traffic volume, over the time period under review, is best represented by the mid year traffic volume (Q_i) data [4].

$$R_j = \frac{f_j \times 10^8}{365.25 \times PL_jQ_j} \qquad \text{eqn. (3)}$$

As a benchmark, collision rates were calculated (eqn. (3)) for all sections of the national secondary network, referred to as population A. This population of roads consisted of rural and urban sections. It is evident looking at Table 1 that collision rates are considerably higher (1.6 times higher) on the urban sections than the rural sections.

Reference population B is comprised of all sites having a speed differential in excess of 10 km/h for all national secondary roads. Again this population of roads is composed of urban and rural sections and again the collision rates are much higher (2.25 times higher) on urban sections than rural road sections. However, within the urban environments, the difference in speeds observed in the data cannot be attributed to challenging horizontal alignment. It is more likely to do with the start-stop nature of traffic in towns and villages.

Reference population C is comprised of all sites having a speed differential in excess of 20 km/h for all national secondary roads. As above the collision rate and speed differential for the urban sections is unlikely to be explained by the horizontal alignment and can be attributed to a more complex traffic movements and the greater presence of vulnerable road users. However there is little difference between the collision rates when comparing the rural population of B and C. Despite the greater differential speed in C there is a slight decrease in the collision rates. This result was not expected.

Table 1 – Summary of collision rates calculate for the national secondary routes

Reference Population	Sum of site lengths (km)	Sum of injury collisions	Exposure data vehicle kilometres	Collision Rate / 10^8 veh km
A	2547	4327	12257773	13.81
Urban A	1777	2380	8041659	18.06
Rural A	770	1947	4216114	11.58
B	416	1339	2023816	25.88
Urban B	124	850	882393	37.68
Rural B	293	489	1141423	16.76
C	223	928	1234034	29.41
Urban C	92	695	682625	38.82
Rural C	131	233	551381	16.53
D	253	666	1209517	21.54
Urban D	50	304	387733	30.67
Rural D	203	362	821784	17.23
E	163.7	673	814299	32.33
Urban E	74.2	546	494660	44.81
Rural E	89.5	127	319639	14.03

In order to investigate the difference observed between rural B and C the number of speed differential zones within both populations were compared. Typically an isolated curve within a straight will be identified by a single sudden drop in speed. Reference population D is comprised of all road sections having one or two speed differential zones within a single site. The collision rates are then compared to reference population E. This population of road sections is comprised of sites with four or more speed differential. It is hypothesised that isolated curves present greater risk to the road users than a series of bends. The results show that this is indeed the case and collision rates for population rural D is greater than population rural E. Within the rural D population there is just over 200 km of road available for analysis. This length of road is characterised by localised speed differentials (> 10 km/h). On average each speed differential has bee observed over 80 m. To gauge the severity of the horizontal alignment adjacent to these isolated speed differentials a sinuosity index (SI) value is calculated.

5.2. Sinuosity analysis results

As an alternative assessment of the horizontal alignment of the national secondary roads a sinuosity analysis was carried out. As stated in section 2.2 the availability of detailed road alignment data for unimproved rural roads aren't readily available. Collecting voluminous amounts of road geometry data is both costly and time consuming. In the absence of detailed road geometry data typically required to establish the radius of curvature, a measure of the roads sinuosity was used to assess the consistency of horizontal alignment at a 1km resolution, for the entire national network of roads (approximately 5,500 km). The ratio of the actual length (L_{al}) of a section of road (X_1,Y_1 to X_2, Y_2) and the shortest path (L_{sp}) of the same road section is defined as sinuosity index (SI).

$$SI = \frac{L_{al}}{L_{sp}} \qquad \text{eqn. (4)}$$

For a level, straight section an SI value of 1 would be expected. The SI value increases as the road becomes more interspersed with curves. Figure 4 shows the spatial layers used within GIS to gauge the SI values for each 1km section of the national road network.

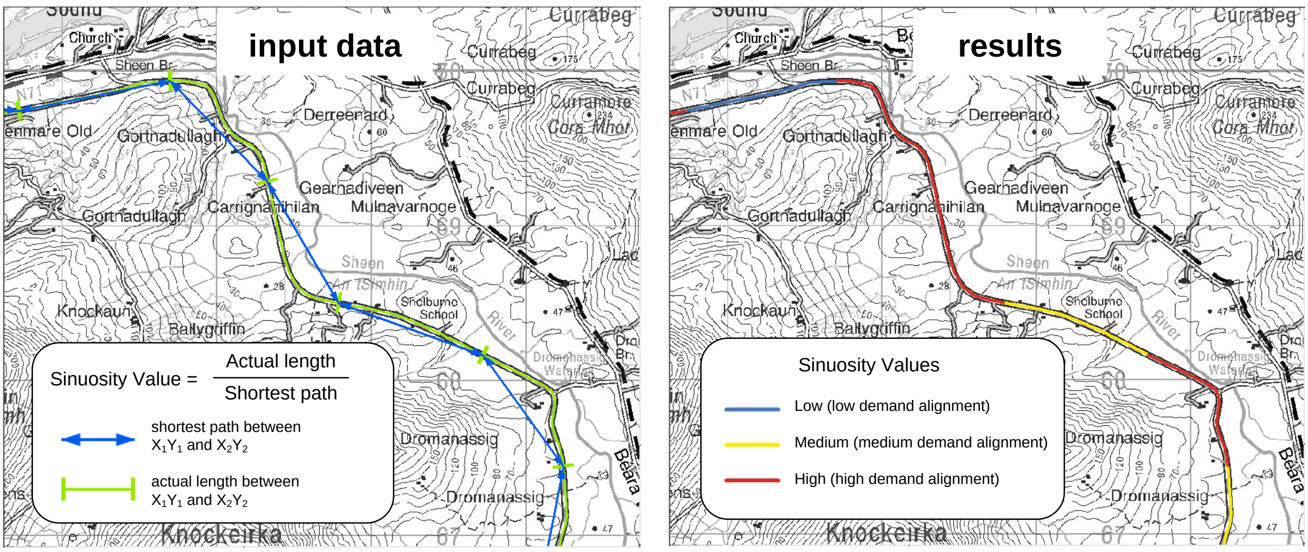

Figure 4 – Linear referenced, 1 km sections, along existing route data (actual length) and straight lines distances (shortest path) generated from paired X,Y data within GIS

Figure 4 show the results of a quantile classification of the sinuosity values generated from dividing the actual driven length of a section of road to the shortest path between the start and end of each section. Quantile classification is a standard method within GIS to group data values equally amongst a number of categories. This classification included both the national secondary and primary networks, in total accounting for approximately 5,500 km of road, including motorways. In this case three categories were used to help define the alignment (i.e. low, moderate and high sinuosity). The results for the national secondary network, considered in section 5.1, are presented in Table 2.

Table 2 – Results of sinuosity analysis by route and summarised by length

Route No.	Sinuosity			
	Low	Moderate	High	% High
52	43.0	68.0	91.6	**45.2%**
53	9.0	6.2	3.0	**16.5%**
54	10.0	12.0	13.5	**38.1%**
55	27.0	17.2	35.0	**44.2%**
56	19.0	43.0	94.4	**60.4%**
58	1.0	3.0	7.3	**64.5%**
59	58.0	85.0	154.9	**52.0%**
60	31.0	31.4	30.0	**32.5%**
61	27.7	21.0	26.0	**34.8%**
62	39.0	33.0	23.6	**24.7%**
63	46.0	20.0	28.8	**30.4%**
65	10.0	21.8	21.0	**39.8%**
66	9.0	10.5	7.0	**26.4%**
67	27.0	42.0	60.3	**46.7%**
68	19.0	17.0	4.9	**11.9%**
69	26.0	37.2	38.0	**37.6%**
70	23.0	31.0	88.7	**62.2%**
71	15.0	38.0	137.2	**72.1%**
72	45.0	53.5	67.0	**40.5%**
73	11.0	16.0	7.5	**21.7%**
74	4.0	9.0	7.1	**35.4%**
75	4.0	1.0	2.8	**35.7%**
76	3.0	26.0	14.7	**33.7%**
77	4.0	7.0	16.2	**59.5%**
80	41.0	50.0	45.9	**33.5%**
81	9.0	34.0	40.0	**48.2%**
82	0.0	0.0	2.6	**100.0%**
83	22.2	12.0	11.0	**24.3%**
84	28.0	26.0	20.0	**27.0%**
85	5.0	5.0	22.2	**69.0%**
86	10.0	13.0	26.6	**53.6%**
87	2.0	11.1	15.0	**53.5%**
Total length km	**628.0**	**801.0**	**1163.8**	

Almost 45 per cent (1,164 km) of the routes listed in table 5 are comprised of roads where the SI would be considered high, using quantile classification. Only 24 per cent (628 km) was classified as having a low SI value. Of the 32 routes considered, nine are comprised of mostly high SI values. That is, somewhere in excess of 50 per cent of their length would be considered a high demand road in terms of its horizontal alignment.

A similar analysis was carried out with reference population rural D. The hypothesis was that the speed differentials observed in the GPS data are a result from high demand roads with significant sinuosity values. Table 3 shows that 412 km of national secondary roads are characterised by isolated curves. These curves are likely to be demanding considering the high SI.

Table 3 – Results of sinuosity analysis by route and length on reference 'Rural D'

	Sinuosity			
Route No.	**Low**	**Moderate**	**High**	**% High**
52	9.0	19.0	36.5	**56.6%**
53	2.0	2.0	1.0	**20.0%**
54	1.0	2.7	5.0	**57.8%**
55	7.0	8.6	20.0	**56.1%**
56	4.0	14.0	42.0	**70.0%**
58	0.0	0.0	5.0	**100.0%**
59	11.0	18.0	53.0	**64.6%**
60	2.0	5.0	5.0	**41.7%**
61	5.7	2.0	5.0	**39.2%**
62	3.0	10.0	4.0	**23.5%**
63	6.0	3.0	8.0	**47.0%**
65	2.0	6.8	9.0	**50.5%**
66	1.0	0.0	1.0	**50.0%**
67	10.0	18.0	23.0	**45.1%**
68	3.0	0.0	1.0	**25.0%**
69	3.0	6.0	4.0	**30.8%**
70	5.0	9.0	51.0	**78.5%**
71	5.0	12.0	55.6	**76.6%**
72	9.0	12.5	23.0	**51.6%**
73	3.0	8.0	4.0	**26.7%**
74	1.0	2.0	2.1	**41.5%**
75	1.0	1.0	1.8	**47.0%**
76	0.0	4.0	5.0	**55.6%**
77	0.0	2.0	2.0	**50.0%**
80	1.0	3.0	4.0	**50.0%**
81	1.0	3.0	10.0	**71.4%**
82	NA	NA	NA	
83	3.0	4.0	4.0	**36.4%**
84	8.0	2.0	2.0	**16.7%**
85	0.0	0.0	1.0	**100.0%**
86	1.0	8.0	14.0	**60.9%**
87	2.0	5.1	10.0	**58.6%**
Total Length km	**109.8**	**190.7**	**412.0**	

Of interest in these results are the N56, N59, N70 and N71. All of these routes have considerable lengths of this type of challenging horizontal alignment. The results presented in Table 1 suggest that collision rates tend to be higher for this type of rural than any other. However these routes all have in excess of 80 km of roads classified by high SI values and thus can be considered high demand roads, requiring high levels of driver attention. As mentioned earlier and noted by Smith et al. [26] drivers experience concentration difficulties on lower demand roads rather then high demand roads. In addition, the critical areas occur at the transition from high to low demand and vice versa. Therefore attention should be payed to the transition between low and high SI scores along each route rather than just concentrate on the high demand, high SI sections.

6. DISCUSSION AND CONCLUSIONS

The results from the analysis have shown that by using existing spatial data sources it is possible to move beyond the traditional approaches of identifying hazardous sites based solely on historic collision data. Speed differential data is readily obtainable from GPS files that accompany road condition data, periodically gathered by road authorities. The bendiness of a road can be approximated by measuring its sinuosity. In combination with the speed differential data, this is a robust and repeatable process to gauge the horizontal consistency of a network of roads, in a quick and cost effective manner. This is particularly true for the older, legacy roads, which are unlikely to have been designed to any modern standard.

This study has shown that the data mining of GPS files can provide locations of vehicle breaking zones on approach to challenging horizontal curves. In the absence of detailed road geometry this information can be used to assess the quality of horizontal road curvature. Isolated curves pose more risk to the road user and this has been verified by the road collision data. This analysis has supported the decision to fund road safety remedial works at 25 locations in 2011 on the Irish national road network. The type of works funded range from minor improvements such as signing and lining, sight line improvements on bends to major improvements requiring substantial re-alignment of significant lengths (2 to 3 km) of road.

Over the coming years, it is likely that more data, very often spatial data, will be gathered by road authorities and similar organisations interested in road safety. Discovering new information by looking for patterns within these volumes of data is a real opportunity to gain valuable knowledge about the physical road environment.

In light of the new European directive on road infrastructure safety management (2008/96/EC), it is paramount that road authorities are able to assess the quality of road infrastructure on the TERN network in a consistent and robust manner. The directive requires that a road authority take a preventative approach to identify defects within their TERN network. Reuse of existing data, as demonstrated in this paper, is an opportunity to stretch their existing resources to achieve this objective efficiently.

REFERENCES

1. Agamennoni, G. Nieto, J. Nebot, E (2008). Mining GPS data for extracting significant places. Proceedings from the 2008 Austrailian conference on robotics and automation, Canberra
2. An Foras Forbartha (1968). Road accident facts 1968 Ireland. An Foras Forbartha – The national Institute for physical planning and construction research, Dublin
3. Belanger, C. Barber, P. (2003). Horizontal alignment technical sheet, in Belanger, C. (ed.). Road safety manual - Recommendations from the World Road Association (PIARC). Route 2 Market, Kent
4. Belanger, C. Barber, P. (2003). Chapter 5 - Identification, in Belanger, C. (ed.). Road Safety Manual: Recommendations from the World Road Association (PIARC). Route 2 Market, Kent
5. Buttenfield, B. Gahegan, M. N. Miller, H. and Yuan, M. (2001). Geospatial data mining and knowledge discovery. Retrieved November 23, 2009 from http://www.ucgis.org/priorities/research/research_white/2000whitepapersindex.htm
6. Castro, M. Iglesias, L. Rodriguez-Solano, R. Sanchez, J. A. (2006). Geometric modelling of highways using global positioning systems (GPS) data and spline approximation. Transport research, part C, pp 233-243
7. Charlton, S. G. de Pont, J. J. (2007). Curve speed management. Research report 232. Land transport New Zealand
8. Chen, S. Rakotonirainy, A. Loke, S. W. (2009). Applying data mining to access crash risk on curves. Proceeding from the 2009 Australasian road safety research, policing and education conference, Sydney
9. de Waard, D. (2002). Mental workload in, Fuller, R. Santos, A. S. (eds.). Human factors for highway engineers. Elsevier science ltd, Oxford
10. Elvik, R. (2008). A survey of operational definitions of hazardous road locations in some European countries. Accident analysis and prevention, Vol 40, pp 1830-1835
11. Elvik, R. (2008). Dimensions of road safety problems and their measurement. Accident analysis & prevention. Vol 40, pp 1200-1210.
12. EuroRAP (2008). EuroRAP 2008: Ireland results, Risk and star rating Ireland’s major roads. AA Ireland & EuroRAP AISBL, Hampshire
13. Girish, N. Malgorzata, R. (2007). 3D calculation of stopping-sight distance from GPS data. Journal of transportation engineering. Vol 133, Issue 11, pp 645-546
14. Hassan, Y. (2004). Highway design consistency – Refining the state of knowledge and practice, proceedings from TBRannual meeting. Transport research board, Washington, DC
15. Hauer, E. (2007). A case for evidence-based road-safety delivery. AAA Foundation for road traffic safety, Washington DC
16. Labrousse, M. (2003). Chapter 3 – Road safety factors, in Belanger, C (ed), Road safety manual – Recommendations from the World Road Association (PIARC), Route 2 Market, Kent.
17. Laird, J. Harris, R. Shen, S. 2010, Accidents on rural roads – for better or worse. Proceedings from the European transport conference 2010
18. Lamm, R. Pasrianos, B. Mailaender, T. (1999). Chapter 9 – Alignment of non built up roads. Highway design and traffic safety engineering handbook. McGraw – Hill, New York
19. Li, K. Tan, H. S. Misener, J. Hendrick, K. J. (2007). Digital map as a virtual sensor – dynamic road curve reconstruction for a curve speed assistant. Vehicle System Dybanics,. Vol 46, No. 12, pp 1141-1158
20. Longley, P. A. Goodchild, M. F. Maguire, D. J. Rhind, D. W. (2001). Geographic information systems and science, Wiley, Chichester
21. Lynam, D. (2007). Rural road safety – policy options. Transport research laboratory, Berkshire
22. Noyce, D. A, Bahia, H. U. Yambo, J. M. Guisk, K. (2005). Incorporating road safety into pavement management, Maximizing asphalt pavement surface friction for road safety improvements. Midwest regional university uransportation uenter. Traffic operations and safety (TOPS) laboratory
23. O’Cinneide, D. Murphy, J.C. Ryan, T. (2004). Interurban accident rates by road type and geometric elements. Proceedings from the European transport conference 2004, Paris
24. Rautela, P. Pant, S. S. (2007). New methodology for demarcating high road accident risk-prone stretches in mountain roads. Current Science, Vol 92, No. 8 pp 1157-1161
25. Road Safety Authority. (2009), Irish Road Collision Facts 2008. Retrieved on November 14, 2009 from http://www.rsa.ie/en/RSA/Road-Safety/Our-Research/Collision-Statistics/
26. Smith, M. Oppenhuis, M. Koorey, G. (2006). Fatigue crashes: the extent to which terrain change has an influence on the fatigues (drowsy) driver. 22nd ARRB conference - Research into practice, Canberra
27. SWOV. (2007). SWOV fact sheet – The blackspot approach. Institute for road safety research. Leidschendam
28. Transport Research Board – National Research Council. (2001). Statistical methods in highway safety analysis, A systhesis of highway practice, National academy press, Washington, DC

29. Watters, P. O'Mahony, M. (2007). The relationship between geometric design consistency and safety on rural single carriageways in Ireland. Proceedings from the European transport conference 2010
30. Wegman, F. (2002). Review of Ireland's road safety strategy. SWOV Institute for road safety research, Leidschendam
31. West, D. (2009). What is data mining? - Interview with Dean West president of association laboratory Inc. Viewed on YouTube 18th Nov 2009
32. Yeung, A. Hall, G. B. (2007). Spatial database systems, design implementation and project management. The geojournal library. Vol 87, Springer