

Road Data Aggregation and Sectioning Considerations for Crash Analysis

Revised: November 15, 2008

(Word Count: 6493 + 4 figures/tables)

Glen Koorey, MIPENZ

Senior Lecturer in Transportation Engineering

Dept of Civil & Natural Resources Engineering, University of Canterbury

Private Bag 4800, Christchurch, New Zealand

Tel: +64-3-364 2951, Fax: +64-3-364 2758

Email: Glen.Koorey@canterbury.ac.nz

ABSTRACT

Increasingly, roading authorities are collecting a variety of datasets related to their networks, including horizontal/vertical alignment, cross-section, traffic volumes, crashes, and the location of features such as intersections and passing lanes. This data may be a mixture of point locations, fixed-length records, and variable-length records.

A critical question, both in terms of computational ease and practical usefulness, is how best to aggregate or section the available data into appropriate road segments for operational and safety analyses. This issue is becoming more pertinent with the development of tools such as *Interactive Highway Safety Design Model* and the *Highway Safety Manual*, which require a logical partition of roads based on many different attributes. The guidance on how to do this however is rather scant. Analysis of traffic exposure versus crash risk is also affected by using fixed or variable length segments.

Research is nearing completion in New Zealand to combine road feature, geometry and crash data on the national rural State Highway network. The resulting database will enable better analysis of crash patterns against different types of road elements and be used to calibrate IHSDM for New Zealand use.

This paper outlines the investigation done to determine a rational method for aggregating the available data into logical road segments. The resulting method uses horizontal alignment, significant cross-section changes, and changes in speed limit. It also attempts to minimize the number of very short segments. The resulting dataset contains approximately 83,400 segments generated from 20,900 lane-km (13,000 lane-miles) of highway.

INTRODUCTION

Increasingly, roading authorities are collecting a variety of datasets related to their networks, including horizontal/vertical alignment, cross-section, traffic volumes, crashes, and the location of features such as intersections and passing lanes. This data may be a mixture of point locations, fixed-length records, and variable-length records.

A critical question, both in terms of computational ease and practical usefulness, is how best to aggregate or section the available data into appropriate road segments for operational and safety analyses. This issue is becoming more pertinent with the development of tools such as *Interactive Highway Safety Design Model* (IHSDM) (1) and the *Highway Safety Manual* (HSM) (2), which require a logical partition of roads based on many different attributes. The guidance on how to do this however is rather scant. The choice of using fixed or variable length segments also affects the analysis of traffic exposure versus crash risk.

Research is nearing completion in New Zealand (NZ) to investigate (predominantly two-lane) rural highway safety (3). The main objectives of the research are:

- (1) To identify road and environmental factors affecting (non-intersection) crashes on rural roads in NZ, particularly at horizontal curves.
- (2) To identify the tasks required to make IHSDM suitable for use in NZ, including the necessary calibration of parameters.
- (3) To validate IHSDM against NZ road and crash data and confirm its suitability.
- (4) To develop a suitably robust model for predicting the relative safety of a rural road alignment in NZ, based on the previous investigations.

A key part of the work is the development of a database combining road feature, traffic and crash data on the national rural State Highway (SH) network. This database will enable better analysis of crash patterns against different types of road elements. It is also being used to determine suitable calibration parameters for using IHSDM in NZ. This paper outlines the investigation done to determine a rational method for aggregating the available data into logical road segments.

HIGHWAY DATA COLLECTION

There are approximately 9600km (5900 mi) of rural State Highway throughout NZ. Although comprising less than 1/7 of the country's rural road length, State Highways carry the bulk of rural traffic (>65%), and subsequently have the bulk of rural crashes (>60%); therefore it is prudent to concentrate on them.

The following data sets were obtained for use in this study:

- RAMM (Road Assessment and Maintenance Management) inventory data for the entire SH network (as of 2002) from Transit NZ, the national state highway agency. This included information on traffic volumes, cross-sections, road geometry, signs, and pavement construction dates.
- Crash data from Land Transport NZ, the national road safety and funding agency. All reported rural crashes (injury or non-injury) between 1996-2006 were obtained from the

national Crash Analysis System (CAS). The dataset comprised over 90,000 records, including more than 37,000 injury and fatal crashes

The collected data were combined in a single Microsoft Access database for subsequent processing. It should be noted that most of the original data processing was done on entire lengths of state highway, including the urban sections. This was for computational convenience, enabling a single run down each highway; urban or other anomalous sections were subsequently identified and removed later.

RAMM Database

Transit NZ's RAMM system (4) provides an inventory of all SH data that describes the physical road environment and demands placed upon it. This includes:

- Inventory data held about the physical features along the road (e.g. surfacing type, drainage, signs, pavement structure, road geometry).
- Condition data held about the road or associated road features (e.g. roughness, texture, condition rating, pavement strength).
- Information held about the local environment and demands placed on the asset (e.g. traffic volumes, road hierarchy, land use, regional administration area).
- Historic information held about events that occur on the asset that are not part of physical inventory (e.g. crash records, construction/maintenance dates).

State Highways in NZ use a system of linear referencing to identify the locations of parts of the network (5). 'Route positions' (RP) define a location along the highway in terms of the displacement from 'reference stations' (RS), regularly spaced along the route. Sometimes, for convenience there are also intermediate reference posts in between the RSs, known as "established route positions" (ERP). These may be used to assist with accurate location in the field or to locate significant features, e.g. a large roundabout, a major bridge, or the start of a divided carriageway section.

RSs and ERPs are generally used as boundaries of RAMM data records. However, subsections between these points are often used to further divide data. For example, roadway data records may be terminated at side-road intersections or at changes in roadway width, surfacing, kerbing or speed limit. Because of the significant highway maintenance investigation work done using the data (e.g. pavement deterioration modelling), sectioning is also often done on the basis of similar pavement condition measures. It is notable however that RAMM data is not ever sectioned in terms of horizontal or vertical geometric elements.

RAMM data is stored as tables in a relational database. For this research, some of the most relevant RAMM tables are:

- **hsd_geom:** Contains road geometry data at 10 m intervals in each travel direction, as described below (approximately 20,000 lane-km or 2 million records).
- **pave_layer:** Contains information on each new pavement layer (e.g. basecourse, sub-base) installed when a section of road was last reconstructed. There are over 22,000 pavement layers on the SH network, although typically a single location has a multiple number of construction layers and subsequent overlays recorded.

- **carr_way:** Summary data for each road section, including traffic volumes, land-use information, number of lanes and road width. Over 8000 SH carriageway sections are recorded in RAMM, with boundaries typically at significant changes in cross-section or traffic.
- **sign:** An inventory of every traffic sign located next to the road, including the position, size, type, and legend of each sign. Nearly 150,000 signs are recorded on the SH network.

By linking these RAMM tables and performing particular queries and matching operations on them, a list of suitable road sections (partitioned by various measures) could be obtained. The process is described in the following sections.

State Highway Road Geometry Data

Since 1992, NZ's SH network has been regularly surveyed to collect data on horizontal curvature, gradient and cross-fall at 10-meter (33 ft) intervals. Originally this was undertaken using the Australian Road Research Board's Road Geometry Data Acquisition System (RGDAS) instrumented vehicle (6), a dedicated vehicle travelling at highway speed. More recently, Transit NZ have repeated the exercise by means of a "SCRIM+" vehicle, which surveys the highway network annually at speeds of up to 60 km/h (37 mph) and also collects a variety of pavement condition measures such as skid resistance and road roughness.

The geometry items recorded by high-speed data collection are:

- relative position along highway (km)
- (longitudinal) gradient (%)
- crossfall or superelevation (%)
- horizontal radius (meters)
- centerline GPS co-ordinates (since 2000)

Currently, geometry data is smoothed with a 30 m (100 ft) moving average and reported at every 10 m. One limitation of this smoothing is that it may be difficult to identify accurately short tight curves. This provides relatively conservative values for analysis, so any identified effects or deficiencies will always be understated. The other practical problem is accuracy of data location, especially when comparing data at the same location from opposing directions. The addition of GPS data has helped to mitigate this problem.

The introduction of road geometry data on NZ's SH network opened up a whole range of studies that previously would have been too difficult to complete. In particular, relatively large sample sets can be analyzed for patterns by integrating geometry data with other RAMM road inventory data, such as traffic volumes and seal widths, and CAS crash data. Koorey (7) describes in more detail a number of other novel applications in highway research and operations that this valuable data source has lent itself to.

It should be noted that, prior to a study of this nature, the raw geometry data should be checked for any built-in survey biases. A bias in horizontal curvature data, for example, might produce more right-hand curves or more deflection to the right (when both directions are surveyed). Although not detailed in this paper, analysis of the NZ geometry data identified

corrections to be made to the raw values for horizontal curvature, gradient and crossfall before further processing.

AGGREGATION OF HIGHWAY DATA

The data were stored in records with a varying range of fixed or variable lengths. To be able to combine the information in a practical manner for analysis, some thought was needed on techniques for aggregating data. Two techniques used in the past are worthy of discussion here.

Cenek *et al* (8) examined the relationship between crashes and road geometry, using over 8000 lane-km (5000 lane-mi) of rural NZ SH data, divided into fixed-length 200 m (660 ft) sections. This fixed-length dataset was also used for research into crashes on curves (9), and the effects of pavement condition data (such as skid resistance and texture) on crashes (10).

Meanwhile, Austroads (11) examined nine lengths of highway in Australia (2900 lane-km) and related geometric attributes such as pavement width and gradient to recorded crash rates. Compared with the previous studies mentioned, the roads were divided to produce sections that were reasonably uniform in terms of geometry, resulting in sections of varying length. Similar studies have been made elsewhere using inventories of roadway “elements” (e.g. 12, 13), although in some cases, the geometric data available has been limited to attributes such as cross-section rather than horizontal/vertical alignment, or only categorized in broad groupings or longer “homogenous” sections.

The latter (“road element”) approach seems intuitively more useful than sections of constant length. In many cases, a fixed-length section may have variable geometry within it, by virtue of where the section boundaries lie in relation to road features such as curves and gradients. Although the very large number of sections analyzed by Cenek *et al* (8) and other “fixed-length” studies helped to minimize the effects of this problem, a more logical division by similar road attributes (such as geometry) should improve the relationships developed.

Which approach is used, however, often depends upon the type of data available. If road geometry data has already been collected systematically at constant intervals along a road then, in practical terms, it is easier to divide road data using a fixed interval length, and one suspects that this was a key factor in the method used by the “fixed-length” studies in NZ. For the Austroads (11) study (involving a smaller dataset), it appears that the division of road sections was undertaken manually via inspection of the geometry data and maps. Other “road element” studies have either manually collected data for each element on site, or used route data from GIS databases or aerial photos to identify road elements.

Clearly, it would be desirable to be able to automate the process of defining road elements, where constant-interval geometry data is available, as in NZ. For example, an automated routine in the database could identify where horizontal curves began and ended. This could be by means of monitoring the recorded curve radius and triggering the start of a new curve when the radius fell below a particular value (e.g. 2000 m or 6600 ft). This technique was previously used successfully in research on rural simulation models (14) to generate straight and curved road elements for Paramics models that required input data in this format. Koorey (7) describes how similar methods were also used to identify curves on highways for reviewing curve advisory speeds and for assessing curve widening requirements for large vehicles. Figure 1 illustrates the conceptual approach applied; here the road data elements recorded at fixed intervals are grouped into logical horizontal elements.

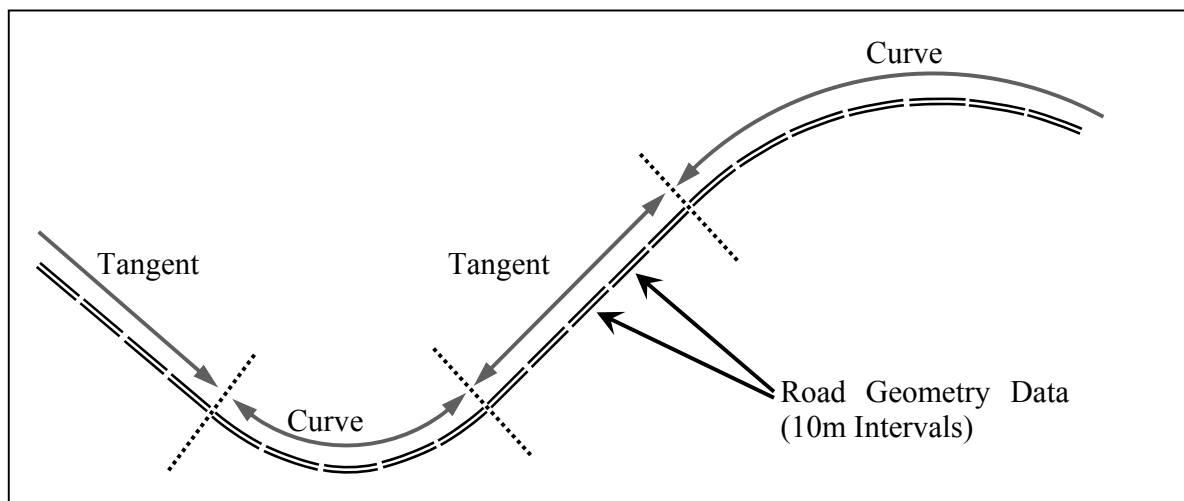


FIGURE 1 Sectioning of Road Geometry Data into Horizontal Elements

Choice of Road Segment Sectioning Approach

A critical question, both in terms of computational ease and academic usefulness, is how best to section the existing road geometry data into appropriate road segments. While the above examples used horizontal curvature as the sole sectioning criterion, this may not necessarily be the most pertinent measure to use (or at least not the only criterion to use). Either way, another question is what “trigger value” (or values) should be used to determine when to create a new road segment.

In NZ, this question has already been widely investigated for the sectioning of road data for asset management purposes. Although the application is not very relevant to this study, the sectioning techniques used may provide some useful insight. HTC (15) investigated various ways to consistently identify “homogenous” road sections for pavement deterioration modeling of state highways. The final methodology chosen used both traditional road inventory data (number of lanes, traffic volumes, road surfacing details, etc) and changes in pavement condition measures (roughness, skid resistance, rutting, etc) to determine the most suitable section lengths for treatment.

Of note was the “CUMSUM” cumulative deviation process to monitor variations in pavement condition measures along the road compared with their target values; this is similar to the way that a golfer monitors how well they are doing compared with “par”. Significant and sustained changes in direction of the CUMSUM trend were used to identify section breaks. The subsequent sections were then further reviewed to minimize the presence of very short sections or adjacent sections with similar characteristics. Another measure used to assess the homogeneity of the resulting sections was to evaluate the coefficient of variation ($CV = \text{standard deviation} \div \text{mean}$) of various road condition parameters, with smaller CV values desired. Although much of the process was automated, a graphing tool was also used to visually confirm the validity of the resulting sections, and a field inspection was used to confirm the final sections.

Cafiso *et al* (13) used Curvature Change Rate (CCR, generally defined as the rate of angular deflection per length of curve), average carriageway width, traffic volume (AADT) and roadside hazard rating to divide 92 km (57 mi) of Italian two-lane local rural roads into sections for crash risk modeling. Each measure was assessed individually first, using statistical methods similar to HTC (15) to divide the roads into homogenous sections. For example, significant changes in the cumulative deflection angle deviation defined section boundaries for CCR. These were then combined to further sub-divide the sections to be homogenous across all four measures. These segments ranged from 150 m (490 ft) in length to over 4 km (2.5 mi) long, generally with a mixture of curves and tangents. Interestingly, some adjacent sections were then aggregated back together to ensure that at least one recorded crash was present in each section.

Fitzpatrick *et al* (16) investigated the applied use of the forthcoming *Highway Safety Manual* guidelines; in particular testing the use of the crash prediction model for rural two-lane highways on two sections of Texas road. They noted that the HSM advises on how to divide a roadway into homogeneous segments, when one of the following variables changes:

- Average Daily Traffic
- lane or shoulder width
- shoulder type
- driveway density (number of accesses per mile)
- roadside hazard rating; usually based on the seven-point subjective scale developed by Zegeer *et al* (17)
- presence of an intersection
- beginning or end of a horizontal curve
- point of intersection of a vertical curve
- beginning or end of a passing lane or short four-lane section
- beginning or end of a two-way left-turn lane (TWLTL)

The authors commented that, in practice, this is not always easy to achieve. Firstly, any such division will depend on the data available; most roading jurisdictions are not likely to have information on all of these variables easily to hand (the Texas DoT only had about half). Secondly, such divisions may not always make sense, for example if an intersection occurs midway through a horizontal curve.

The approach used ultimately by Fitzpatrick *et al* was to subdivide the roadway into horizontal curve and tangent segments and predict the number of mid-block crashes for each segment. Separately, the number of intersection-related crashes at each intersection was then predicted.

Fitzpatrick *et al* noted that, even without subdividing for the variables not available, the sub-division process sometimes produced some extremely short segments (as small as 5 m or 16 ft long). They concluded that the HSM needed to provide better advice on minimum section lengths (and how to avoid or manage very short segments) as well as advice on which are the more important variables to consider when subdividing the roadway.

Another issue to consider when determining sectioning lengths is its effect on the distribution of crash numbers over the sections. For example, Lord *et al* (18) noted that the incidence in many studies of a higher-than-expected proportion of sites with zero crash counts can be attributed to (among other reasons) data with relatively small spatial or time intervals. Under-reporting of actual crash numbers was another reason cited and this issue is particularly prevalent in rural areas. Smaller size sites will also tend to produce relatively low mean crash numbers, which can be difficult to model or validate.

Cafiso *et al* (13) attempted to minimize this problem by aggregating road segments so that each segment had at least one recorded crash. Other studies have attempted to identify a minimum length for road segments e.g. Resende & Benekohal (19) concluded that, to get reliable prediction models, rural crash rates should be computed from road sections 0.5 miles (0.8 km) or longer. Thomas (20) noted that the statistical form of the crash count distribution varies as the length of the road segment increases, with a Poisson distribution most appropriate only for very short segments and much more variability evident in longer segments (where a Negative Binomial distribution may be more appropriate).

Another likely problem with shorter segment lengths is the chance that a feature of the road in one segment triggered a crash officially located in another segment. This can be particularly an issue in remote rural locations, where precise location by the attending traffic officer may be less likely. The attending officer may also be basing their location on where a vehicle came to rest, whereas the initial cause of the crash may have occurred further back along the road; at typical highway speeds, vehicle momentum can lead to a location discrepancy of hundreds of meters. Thus, it is prudent to try to produce relatively long segment lengths where the crash and potential contributory features are more likely to be located in the same segment.

AGGREGATION METHOD USED IN THIS STUDY

For the NZ situation, as with many of the studies mentioned above, only some of the possible highway data is automatically recorded by location, such as traffic volume and (average) roadway widths. By default, there is no record of horizontal and vertical curves, although the 10 m road geometry data allows inference of these geometric elements to be determined. Similarly, intersection locations (and their other attributes), roadside hazards or driveways are not recorded in a specific database, although their locations and features can be determined manually or indirectly through other sources (e.g. traffic signs database, video log of the road).

For a study like this, it was therefore rather difficult to relate the existing RAMM data to physical geometric road elements present on the highway, without extensive data collection and/or processing. It is instructive to contrast this with the Texas DoT approach where a special road inventory table, called “Geo-Hini”, has been created to link road positions with horizontal curvature and another table, “P-Hini”, records point features such as intersections (16).

Given the importance of horizontal curvature in this study, segmentation by horizontal element seemed an appropriate approach. Many existing curves in NZ have “evolved”, rather than having ever been formally designed, so identification of spiral transition curves was considered a redundant exercise. Instead, the study data was simply divided into relatively “straight” segments (i.e. tangents) and relatively “bendy” segments (i.e. curves).

If horizontal radius is used to determine the division between segments, some thought needs to be given to a suitable cut-off value. Sections of state highway in NZ that are essentially straight are recorded as having large radius values (e.g. typically >3000 m or 9,800 ft), the value and sign (i.e. direction of curvature) somewhat dependent on the path variations of the survey vehicle. A low cut-off value (e.g. 1000 m or 3300 ft) may be appropriate given that it is typically curves with radii below this figure that produce safety problems; in fact, studies would suggest that only curve radii below ~400 m (1300 ft) are of real significance (e.g. 21). However, if curves are to be recorded in their entirety (i.e. the part that deflects from the tangent lines) then a low cut-off value may omit a significant portion of many curves (it may also limit the number of curve-related crashes that are identified within each curve element). Conversely, too high a value may create too many “false curve” records that contribute little to the analysis of crashes on curves. A curve radius of greater than 2000 m (6600 ft), for example, produces a horizontal deflection of less than 3° per 100 m (330 ft). False curvature readings, due to path variations made by the survey vehicle, are also more likely to be seen at higher radius values.

To assess the suitability of the cut-off radius used, a trial was undertaken to divide the road geometry data available into elements using two different cut-off values for curves; 1000 m (3300 ft) and 2000 m (6600 ft). Visual Basic code was created to “walk down” each highway, identify the start of new horizontal elements by virtue of the recorded radius, and to collate statistics on the attributes of each element. A “look-ahead” routine was included in the algorithm code so that three consecutive geometry records had to be above or below the cut-off value before terminating an element; this was to avoid one-off rogue values. Some trial and error was also required to identify aspects of the algorithm logic creating other spurious elements and subsequently minimize their occurrence. Table 1 summarizes the key details of the resulting element distributions.

TABLE 1 Distribution of Geometric Elements with Varying Cut-Off Values

Cut-Off Radius Used	2000 m (6600 ft)	1000 m (3300 ft)
Number of Curve Elements	67,737	53,441
Average Curve Length	142 m (465 ft)	134 m (439 ft)
Number of 10 m Curve Elements	2269 (3.3%)	451 (0.1%)
Number of Curve Elements < 100 m	28,322 (41.8%)	21,837 (40.9%)
Number of Curve Elements > 1000 m	107	14
Maximum Curve Length	9.84 km (6.11 mi)	4.91 km (3.05 mi)
Number of Tangent Elements	39,456	37,778
Average Tangent Length	282 m (924 ft)	366 m (1200 ft)
Number of 10 m Tangent Elements	14	14
Number of Tangent Elements < 100 m	16,625 (42.1%)	13,308 (35.2%)
Number of Tangent Elements > 1000 m	2113 (5.4%)	2876 (7.6%)
Maximum Tangent Length	15.01 km (9.32 mi)	18.28 km (11.35 mi)

As might be expected, the smaller cut-off value produced fewer elements overall and reduced the proportion of short (< 100 m) elements as well. Either way, there were a still large number of short elements created, which may not be useful for analysis purposes. It may be

prudent to add a post-processing step to combine some of the shortest elements with their adjacent counterparts.

Interestingly, the average curve element length was hardly affected by the cut-off value, whereas the average tangent length increases by more than 30% with the lower cut-off value. A smaller cut-off value also tends to result in more of the longer elements being tangents rather than curves. This reflects the fact that previous sequences of (say) “tangent → 1500 m radius curve → tangent” have now been replaced by a single, longer tangent. These global statistics however mask the underlying distribution somewhat. Figure 2 and Figure 3 show the distribution of curve and tangent element lengths (up to 1000 m long) using the cut-off values of 2000 m and 1000 m respectively.

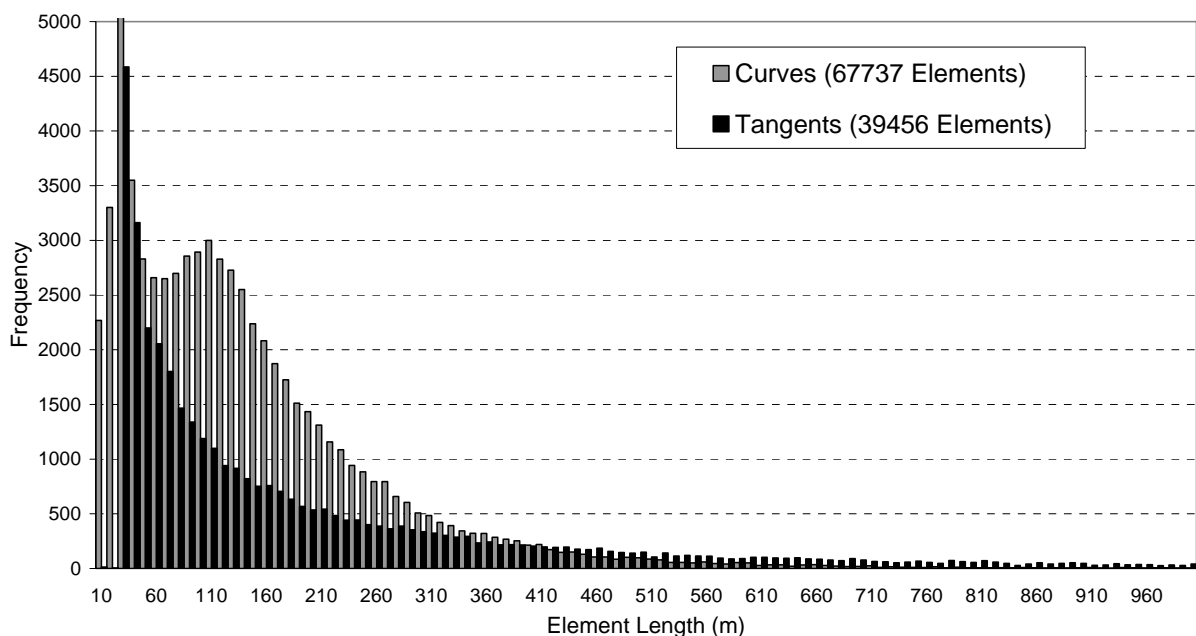


FIGURE 2 Distribution of Road Geometry Elements (2000 m Cut-Off Value)

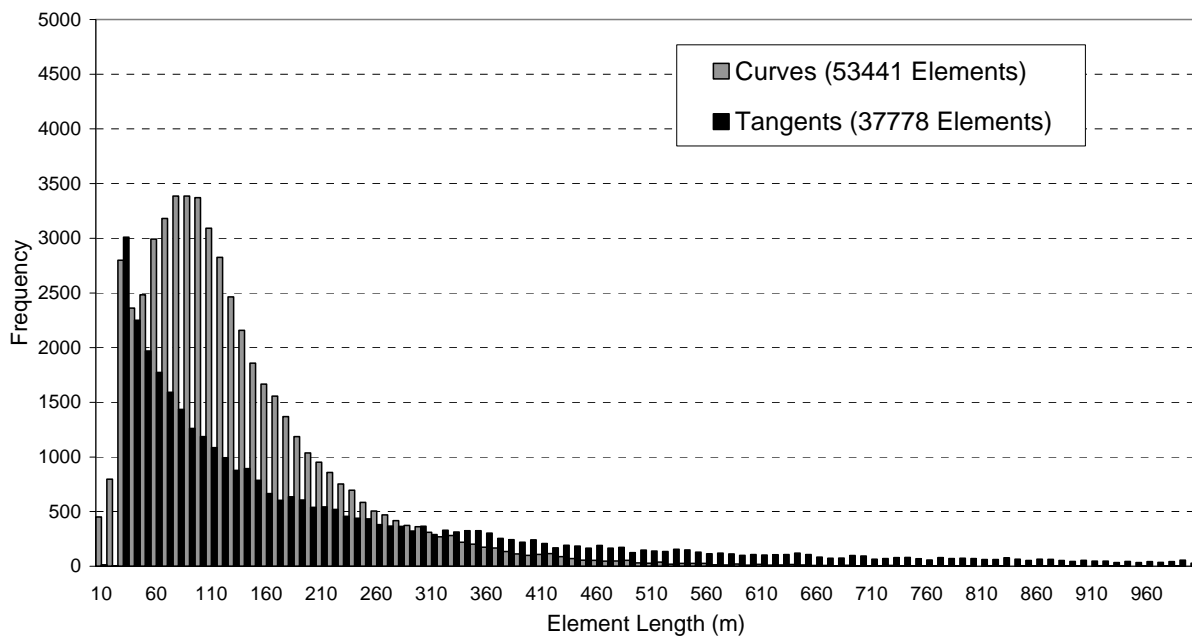


FIGURE 3 Distribution of Road Geometry Elements (1000 m Cut-Off Value)

It can be seen that tangent elements peak at very short lengths, irrespective of cut-off value, with the frequency systematically decreasing as the length increases. Curve elements also have a peak at 30 m (100 ft) in length; however, the lower cut-off value reduces its impact and accentuates the second peak at around 100 m (330 ft) in length.

The results would suggest that a 1000 m (3300 ft) cut-off is somewhat more efficient and minimizes the worst excesses of very short elements. One could contemplate an even shorter cut-off value; however, this would be at the cost of capturing most of the physical curve length in the element record. Even with the 1000 m cut-off value there are a considerable number of “tangents” exhibiting significant deflection mainly due to the large radius curves included in them.

Segments could also be further sub-divided by other features suggested above:

- **Intersections** introduce their own safety problems but are otherwise not necessarily a major factor in crashes for travelers continuing along the main road. For this study, the focus was on non-intersection crashes so, by ignoring crashes clearly related to intersections, the presence of an intersection should not be a reason to sub-divide a horizontal curve for example. The more pragmatic problem is that RAMM does not record intersections; so their identification would have to be by manual means or inference from data sources like the “sign” table.
- **Average Daily Traffic** is not likely to change greatly along rural highway sections, except at major intersections. Although this could mean that a road segment straddling an intersection has somewhat differing traffic volumes either side of it, the need for entering or departing traffic to speed up or slow down near the intersection should mean that this traffic does not greatly contribute here to non-intersection crashes, which are typically at

full highway speed. Therefore, the existing segment divisions should be sufficient for traffic volume purposes.

- **Sealed lane or shoulder widths** tend to be relatively consistent over long distances, reflecting the practice of resealing relatively long (>1 km) road lengths at a time. The exception may be on some tight horizontal curves where additional widening is provided to allow large vehicles to negotiate them without encroaching on the opposing traffic lane. Given that this widening is usually related to a curve segment, using the same segment boundaries for width and curvature seems prudent. For significant (>1 m) seal width changes midway along a tangent, it is also sensible to further sub-divide the road there.
- **Vertical alignment** can also be determined from the available geometry data, such as the location of grades and vertical curves. Inspection of the state highway data suggests that, due to the vagaries of the existing alignments, such division would be at least as hard to do as that for horizontal curvature and would produce many relatively short segments, often splitting horizontal elements into even shorter lengths. Of more practical interest would be the interaction of vertical geometry with horizontal elements, particularly the presence of sharp vertical curves or steep grades. This can be done instead by analyzing the changes in vertical grade over each horizontal element and categorizing the nature of the vertical alignment present.
- **Passing lanes or short four-lane sections** are identified by the number of lanes recorded in RAMM's "carr_way" table, although passing lane direction is not identified and such sections may be divided into multiple carriageway records. Again, it would be immensely useful for similar studies if a separate record of passing lane locations and attributes was maintained, avoiding the need for further database interrogation.
- **Speed limits** usually reflect changes in both land-use and cross-section, so are a logical division point. Given that this study was focusing on rural highways, such a demarcation was also necessary to eliminate urban sections of highway. As it is, most speed limit boundaries are already reflected in RAMM's "carr_way" records; the "sign" table also records the location of speed limit signs. Neither of these is consistent however; again, a separate table of speed limit zones would be of immense value.
- **Other potential dividing features** include changes in accessway density, centerline treatments, roadside hazards, and adjacent land-use. However generally there was no readily available data source for this information for the entire State Highway network. Some of their attributes are also captured sufficiently using other features; for example, a speed limit change often reflects changes in accessway density and land-use.

On balance, it seems that division by horizontal element should remain the primary means of data aggregation. Major changes in cross-section (seal width or number of lanes) and speed limits should also be identified and used to further demarcate highway data, so long as they occur not too close to a horizontal curvature boundary (e.g. ≥ 50 m or 160 ft). Summary information about vertical alignment can be collected over each resulting segment. This approach was adopted for this study.

Short Road Sections

The question remains about how to deal with the remaining very short segments, say those <50m (160 ft). While the winding geometry in many parts of NZ will indeed result in some relatively short curves, it may be difficult to confidently ascribe the features of such a curve to a particular crash located there (especially given the vagaries of accurate crash location, discussed above). It may either be better to group short segments with adjacent segments or at least not terminate them until they have reached a specified minimum length. This may be particularly so if the segment in question does not involve significant angular deflection or speed reduction.

A consistent automated means of identifying when to extend/join short segments is difficult to achieve. A manual inspection of the short segments created revealed a few common features:

- Some of the short segments were tight curves at an intersection; for example, at the very start/end of a highway, where the highway traversed a roundabout, or where the highway changed to another road. In many cases, this occurred in an urban area and so would ultimately be excluded from analysis anyway. It seemed also appropriate to identify and remove other short right-angle bends from analysis, where intersection crashes are likely to be most prevalent anyway.
- For short “tangents”, many included curve radii in the 1000-2000 m range, leading into the subsequent curve. Invariably, all tangents <50 m long had no more than 2° of total angular deflection, which would not greatly influence the description of the subsequent curve element if combined together.
- Approximately 1000 short “curve” elements had a total angular deflection of less than 2° (with a maximum length of 90 m or 300 ft). Motorists would hardly notice such curves and, in some cases, they appear to be actually parts of essentially straight road sections. It was therefore pragmatic to combine these sections with adjacent tangents.
- While many short segments could be extended into or joined with the next adjacent segment, there remain some for which this does not seem appropriate, usually near tight reverse curves.

A pragmatic approach for the latter point might be to create an additional segment type called (say) “reverse curves” that can encompass a series of short mixed elements. However, while this may be useful for crash analysis, it poses extra difficulties for use with IHSDM, which requires the specification of tangent or curve elements when defining highway alignments.

Therefore, this study retained the use of just “tangent” and “curve” segments. However, any tangents <50 m long or curves with less than 2° of total deflection were not terminated, but instead combined with the subsequent data. It should be noted that these cut-off values are somewhat arbitrarily chosen, but should at least eliminate the most unnecessarily short segments.

FINAL PROCESSING AND DISCUSSION

The resulting method used considers horizontal alignment, significant cross-section changes, and changes in speed limits. It also attempts to minimize the number of very short segments. The following procedural steps were used:

- (1) State Highway 2002 RAMM data for all of NZ were imported into Microsoft Access tables. This included a “Geometry” table for the high-speed road geometry data, sorted by running distance.
- (2) A separate table, “SegmentDivides”, was created of locations where segments should be divided, including changes in the number of traffic lanes (such as passing lanes and one-lane bridges) and significant changes in carriageway width (such as bridges or urban boundaries). Over 1500 such locations were identified, mainly by queries of various RAMM tables.
- (3) A “GeomElements” table was created to store details of the road elements identified in the Geometry table. A Visual Basic routine scanned the Geometry table and determined where each geometric element started and finished. As discussed above, curve radius (1000 m cut-off) was the main criterion used to determine when an element (curve or tangent) had ended, with minimum length and deflection checks also in place to minimize very short segments. The SegmentDivides table was monitored in parallel to identify additional segment boundaries, so long as they weren’t too near (i.e. within 50 m of) the most recent boundary. When a segment was terminated, running statistics on attributes such as length, radius, gradient, deflection angle were collated and a new record was created containing the details of that segment.

The resulting dataset contains approximately 83,400 segments generated from 20,900 lane-km (13,000 lane-miles) of highway. This represents an average segment length of 250 m (820 ft). Notably however, only 3600 segments (4.3%) are less than 50 m and 26,000 segments (31%) are less than 100 m; this compares favorably with the earlier trial analyses presented in Table 1, despite the additional sectioning criteria used in addition to horizontal curvature. Many of these shorter segments were subsequently excluded when other filters such as urban areas or sharp intersections were applied to the dataset. However, for safety analyses, it may be useful to check segments against crashes in adjacent segments as well.

Having created a suitable partition of road data, this dataset can now be combined with traffic volume and crash data within each segment to produce a large catalog of variable-length road segments, each with relatively homogenous road and environmental characteristics. The potential now exists to undertake a variety of investigations using this database to relate these characteristics to their crash potential.

Because the length and traffic volume of each segment is known, crash risk can be related to exposure (i.e. crashes per veh-km). The isolation of each significant curve element also allows the analysis to consider whether it is more appropriate to treat curves as “point hazards” or continuous hazards.

CONCLUSIONS

In considering the challenge of dividing road network data into segments, the following suggestions are given:

- Variable-length road segments seem intuitively more useful than fixed-length segments, because of the mixed attributes contained in the latter. However, that advantage is less

when shorter lengths are used, and fixed-length segments are computationally easier to create from constant-interval raw data.

- Subdivision by horizontal geometry is normally an important factor. Where curves haven't been manually located, curvature monitoring is needed to identify when to trigger the start of a new segment. A lower radius cut-off value helps to identify clearly the curves of most concern and minimize the number of spurious segments. However, a larger cut-off captures more of the curve extents and improves the chances of matching curve crashes to the appropriate sections.
- Subdivision by major cross-section changes (number of lanes or significant width changes) and speed limit changes should capture the other main relevant division points. Indirectly these measures are also likely to reflect changes in land use, accessway density, and traffic volumes.
- Short segments can be minimized by not creating new segments when a short distance (e.g. < 50 m) has elapsed since the last new segment. Similarly, short curves with minimal deflection (e.g. < 2°) can be incorporated into adjacent tangents.
- Rather than segmenting data at changes to every possible attribute, a more pragmatic approach is to just summarize parameters for some attributes. For example, vertical alignment can be described over each segment in terms of the general shape (grade, sag, summit, mixed), direction (uphill, downhill, level) and statistics such as average/minimum/maximum grade.
- To simplify the creation of useful data segments, it is strongly recommended that roading agencies create and maintain databases of the location of key features including curves, intersections, passing lanes, and speed limits.

ACKNOWLEDGEMENTS

The author would like to acknowledge Transit NZ and Land Transport NZ (now both part of NZ Transport Agency) for the road and crash data used in this study. He also acknowledges the support of the University of Canterbury for their support of the underlying doctoral research, and the contributions of his supervisors, Assoc. Prof Alan Nicholson and Dr Mofreh Saleh.

REFERENCES

1. FHWA, 2008. *Interactive Highway Safety Design Model (IHSDM)*, Federal Highway Administration, US Dept of Transportation. www.tfhrc.gov/safety/ihsdm/ihsdm.htm, accessed 12 July 2008.
2. TRB, 2008. *Highway Safety Manual*, Task Force on the Development of the Highway Safety Manual, Transportation Research Board of the National Academies, Washington, D.C., USA. www.highwaysafetymanual.org, accessed 12 July 2008
3. Koorey, G.F. 2008. *Incorporating Safety into Rural Highway Design*. PhD Thesis in Transportation Engineering, Department Of Civil Engineering, University of Canterbury, Christchurch, NZ. (forthcoming)
4. Transit New Zealand, 2006. State Highway Database Operations Manual. *Manual No. SM050*, Mar 2006, Wellington, NZ.
5. Transit New Zealand, 2004. State Highway Location Referencing Management System (LRMS) Manual. *Manual No. SM051*, Jul 2004, Wellington, NZ.
6. Rawlinson W.R., 1983. The ARRB Road Geometry Instrumented Vehicle - General Description. *ARRB Internal Report AIR 276-2*, ARRB, Vermont South, Victoria, Australia.
7. Koorey G.F., 2005. The Use of Road Geometry Data for Highway Applications, *3rd International Symposium on Highway Geometric Design*, Chicago, USA, Jun 2005.
8. Cenek P.D., Davies R.B., McLarin M.W., Griffith-Jones G., Locke N.J., 1997. Road Environment and Traffic Crashes. *Transfund New Zealand Research Report No. 79*, Wellington, NZ. 132 pp.
9. Koorey, G.F., Tate, F.N., 1997. Review of Accident Analysis Procedures for Project Evaluation Manual. *Transfund New Zealand Research Report No.85*, Wellington, NZ. 54 pp.
10. Cenek P.D., Davies R.B., 2004. Crash risk relationships for improved safety management of roads. *Towards Sustainable Land Transport Symposium*, Wellington, NZ, Nov 2004.
11. Austroads, 2000. Relationship Between Crash Risk and Geometric Characteristics of Rural Highways. *Austroads Publication AP-R162/00*, Sydney, NSW, Australia, 73pp.
12. O'Conneide D., Murphy J., Ryan T., 2005. The Effect of Geometric Elements on Interurban Accident Rates, *3rd International Symposium on Highway Geometric Design*, Chicago, USA, Jun 2005.
13. Cafiso S., Di Graziano A., Di Silvestro G., La Cava G., 2008. Safety Performance Indicators for Local Rural Roads: Comprehensive Procedure from Low-Cost Data Survey to Accident Prediction Model. *87th Transportation Research Board (TRB) Annual Meeting*, Washington DC, USA, Jan 2008
14. Koorey G.F., 2003. Assessment of Rural Road Simulation Modelling Tools. *Transfund NZ Research Report No.245*. Wellington, NZ, 124 pp.

15. HTC, 2002. Sectioning Improvement. *Transit New Zealand Project 801/004995 Final Report*. HTC Infrastructure Management Ltd, Auckland, NZ.
16. Fitzpatrick K., Schneider IV W., Carvell J., 2006. Using the Rural Two-Lane Highway Draft Prototype Chapter. *85th Transportation Research Board (TRB) Annual Meeting*, Washington DC, USA, Jan 2006
17. Zegeer C.V, Hummer J., Reinfurt D., Herf L., Hunter W., 1987. Safety Effects of Cross-Section Design for Two-Lane Roads. *Report No. FHWA-RD-87/008*. (US) Federal Highway Administration (FHWA), TRB National Research Council, Washington D.C., USA, Oct. 1987.
18. Lord D., Washington S.P., Ivan J.N. 2005. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis and Prevention*, Vol.37 No.1, pp.35–46, Elsevier.
19. Resende P., Benekohal R., 1997. Effects of Roadway Section Length on Accident Modeling. *Traffic Congestion and Traffic Safety in the 21st Century Conference*. ASCE, Chicago, USA, Jun 1997.
20. Thomas I. 1996. Spatial Data Aggregation: Exploratory Analysis of Road Accidents. *Accident Analysis and Prevention*, Vol. 28, No.2, pp.251–264.
21. Lamm R., Beck A., Zumkeller K., 2000. Analysis and Evaluation of Interrelationships between Traffic Safety and Highway Geometric Design on Two-Lane Rural Roads. *Proceedings of the 2nd International Symposium on Highway Geometric Design*, June 2000, Mainz, Germany, pp.557-570.