Modelling crash risk on the New Zealand state highway network
March 2012

P Cenek, Opus Central Laboratories
R Henderson, Opus Central Laboratories
R Davies, Statistics Research Associates

NZ Transport Agency research report 477

ISBN 978-0-478-39423-8 (print)
ISBN 978-0-478-39422-1 (electronic)
ISSN 1173-3756 (print)
ISSN 1173-3764 (electronic)

NZ Transport Agency
Private Bag 6995, Wellington 6141, New Zealand
Telephone 64 4 894 5400; facsimile 64 4 894 6100
research@nzta.govt.nz
www.nzta.govt.nz

Cenek, P, R Henderson and R Davies (2012) Modelling crash risk on the New Zealand state highway network. *NZ Transport Agency research report 477.* 100pp.

Keywords: collective risk, crash prediction models, KiwiRap, out of context curves, personal risk

An important note for the reader

The NZ Transport Agency is a Crown entity established under the Land Transport Management Act 2003. The objective of the Agency is to undertake its functions in a way that contributes to an affordable, integrated, safe, responsive and sustainable land transport system. Each year, the NZ Transport Agency funds innovative and relevant research that contributes to this objective.

The views expressed in research reports are the outcomes of the independent research, and should not be regarded as being the opinion or responsibility of the NZ Transport Agency. The material contained in the reports should not be construed in any way as policy adopted by the NZ Transport Agency or indeed any agency of the NZ Government. The reports may, however, be used by NZ Government agencies as a reference in the development of policy.

While research reports are believed to be correct at the time of their preparation, the NZ Transport Agency and agents involved in their preparation and publication do not accept any liability for use of the research. People using the research, whether directly or indirectly, should apply and rely on their own skill and judgement. They should not rely on the contents of the research reports in isolation from other sources of advice and information. If necessary, they should seek appropriate legal or other expert advice.

Acknowledgements

The authors would like to acknowledge the assistance of the following Opus Central Laboratories staff: Iain McIver for his involvement in generating the databases, and Bill Frith and Ian Wells for their technical review and editing of the draft report.

Thanks also to the NZTA steering group members (Fergus Tate and Fabian Marsh) for their support and guidance, and the peer reviewers, Mike Jackett and Colin Brodie, for their many helpful comments.

Authors' note

This report assumes that the reader possesses a significant knowledge of statistical modelling techniques.

Abbreviations and acronyms

ADT	average daily traffic count
AS	advisory speed
CAS	crash analysis system
DOT	Department of Transport
ESC	equilibrium skid resistance
IHSDM	Interactive Highway Safety Design Model
IL	investigatory level
IRI	International Roughness Index
MOT	Ministry of Transport
MVMT_IDA	first letter of the MOT's crash movement classification
OOCC	out-of-context curve
RAMM	Road Asset Maintenance Management
SCRIM	the Sideways-force Coefficient Routine Investigation Machine
TL	threshold level

Contents

Executive summary

A considerable number of crash prediction models have been developed in both Australia and New Zealand. Most are created for a specific application and need. In 1997–2002, Cenek and Davies created a statistical model, based on Poisson regression crash prediction, specifically for the New Zealand state highway network. This model, which is notable for its sophistication, utilises traffic flow, road geometry and road condition as inputs, and predicts the number of fatal/injury crashes on rural road networks with any moisture level, and the number of fatal/injury crashes on wet rural road networks. Significantly, this model appears to be the sole Australasian crash prediction model to consider road condition.

This report updates and builds on that earlier model. It assumes that the crashes are statistically independent and the number in each 10m segment of road follows a Poisson distribution. (Of course, for most segments the number will be zero.) Fitting is by maximum likelihood.[1] As the exponential of a linear combination of the road characteristics is being used, the actual model is multiplicative. The actual rate of reported crashes in a 10m length of road is the average of generating rates over the 10m lengths in its immediate neighbourhood (on the same road) and summed over the two sides of the road. In the results reported here, the average is over 10m lengths within 100m of the length being considered. Typically, this gives an average over 210m on each side of the road. There is no weighting down of the more distant 10m lengths. This averaging allows for error in reporting the location and the possibility that a crash ends at a location some distance from the piece of road involved in generating the crash.

Several 'what if' studies were conducted to investigate the application of the refined model.

The most striking result is the close agreement with previous analyses of the 1997–2002 data. The inclusion of the term for interaction between roughness and curvature suggests that roughness is a factor for curves where traffic is going at close to full speed but there still is some curvature.

There is a suggestion that skid resistance is more important on curves than on straight roads. This makes sense, but as yet the dependence has not been precisely quantified. The agreement between the analyses when we look at all casualty crashes, and those when we consider only serious/fatal crashes, suggests that the low reporting rates associated with minor-injury crashes is not a serious problem. Similarly, there is little change when the year-x-region interaction is included; this also suggests that reporting rates are sufficiently consistent for the analyses to be valid.

There is still more variability in the data than the Poisson model would predict. In this case the model is unlikely to fit exactly, as there are numerous things not included and the fit might be the best that one can reasonably expect. However it is possible that the problem lies in the estimates of average daily traffic and this might be worth investigating further.

---

1 The *maximum likelihood method* is a method of estimating statistical parameters which selects the 'most likely' value of a parameter with regard to other possible values.

Abstract

This report presents an updated statistical analysis of data relating to crash rates on New Zealand roads. The research was carried out during 2007–2009 and it precedes the changes in 2010 to the New Zealand *T10* specification. The refinements presented are associated with accounting for differences between the local and the general (ie design) speed environment, crash severity and interactions between curvature and roughness. The addition of these refinements will extend the present model's usefulness for guiding safety initiatives and providing economic justifications.

The regression model used in the analysis assumes that crashes are statistically independent and the number of crashes in each 10m segment of road follows a Poisson distribution. Inputs to the model include the average daily traffic (per side) and is a linear combination of the road characteristics, being transformations of terms that include factors such as gradient, curvature, out-of-context-curve effect, skid site classification, skid resistance, region and an urban/rural classification.

There is still more variability in the data than the Poisson model would predict. However, the results indicate the availability of a robust crash prediction model that takes into account both road condition and road geometry, allowing proactive identification of existing engineering-related road safety deficiencies and more importantly, the ability to quantify the potential for improvement.

1 Introduction

This report provides the findings of research that refine a statistical model first developed by Cenek et al (1997) for predicting injury crashes on New Zealand two-lane state highways. This model is believed to be one of the first to successfully relate crash rates to road geometry and road condition. The former Transit New Zealand and its consultants used the original model to analyse the safety performance of the state highway network, leading to safety improvement projects in the Bay of Plenty (Karangahake Gorge), Nelson/Marlborough (SH6) and Wellington (Centennial Highway). The need for model refinements has resulted from this use.

The refinements presented are associated with:

- accounting for differences between the local and the general (ie design) speed environment
- better separation of road gradient effects
- consideration of interaction effects such as roughness and curvature
- more statistically robust relationships through analysis of a much larger crash dataset covering the 10-year period 2000–2009.

These refinements extended the usefulness of the statistical model for guiding safety initiatives and providing economic justifications, leading to its application to several recent NZ Transport Agency (NZTA) safety-related initiatives, including KiwiRAP and out-of-context-curves.

This report details the model in its most refined form, as well as providing example applications of its use. These applications show that the availability of a robust crash prediction model that takes into account both road condition and road geometry will allow proactive identification of existing engineering-related road safety deficiencies and more importantly, quantify the potential for improvement. The use of the model in the safety management of road networks will therefore directly contribute to making the driving environment for New Zealand motorists safer.

A prime function of this report is also to document the crash prediction models currently being used in the KiwiRAP and curve-context initiatives introduced by the NZTA and its partners to assist in achieving the aims of Safer Journeys: New Zealand's Road Safety Strategy 2010–2020.

2 Literature review[2]

A number of crash prediction models have been developed in Australasia, particularly in New Zealand for application in New Zealand (eg Davies et al 2005), and to a lesser extent in Australia for application in Australia. An excellent summary of models developed in both countries has been given by Turner and Wood (2009a).

2.1 New Zealand crash prediction models

The Poisson regression crash prediction model specifically developed for the New Zealand state highway network outlined by Cenek and Davies (2006) is notable for its sophistication; it utilises traffic flow, road geometry and road condition as inputs and predicts the number of fatal/injury crashes on rural road networks with any moisture level, and the number of fatal/injury crashes on wet rural road networks. Significantly, this model, as far as the report authors are aware, is the sole Australasian crash prediction model to consider road condition. In their paper, as well as detailing the model, Cenek and Davies presented the results of a case study where the model was applied to the Karangahake Gorge in New Zealand.

Turner, Durdin and Jackett (2003) provided details on other crash prediction models developed from reported injury crash data and traffic counts in New Zealand for major crash types and total crashes. Possible applications of the models were discussed, including economic evaluation, developing performance measures, assessing safety management systems, and optimisation of network flow patterns to improve safety.

Turner, Dixon and Wood (2004) discussed a model that focused on predicting crash risk as a function of roadside hazard. Such models were claimed by the authors to be useful in targeting resources to remove roadside hazards.

Turner, Roozenburg and Francis (2006) discussed crash model predictions for cyclists and pedestrians, and reported a noticeable 'safety in numbers' effect.

Turner, Wood and Roozenburg (2006) noted that crash occurrence was typically low at rural priority-controlled intersections, compared with priority-controlled urban intersections, due to low traffic volumes. They went on to discuss the production of crash prediction models for rural priority-controlled intersections based on traffic volume, sight distance, approach speed and geometric design. The paper also outlined some of the more important statistical methods that were used to assess the quality of the models produced.

Turner, Persaud and Chou (2007) observed that over the previous 15 years (from 2007), a multitude of crash prediction models had been developed for rural roads, urban intersections and mid-block sections in New Zealand. They emphasised that there was a growing need for more-comprehensive models similar to those developed internationally (eg in the US and Europe).

Partially addressing this perceived need, Turner, Persaud, Chou, Lyon and Roozenburg (2007) presented New Zealand-based comparisons of selected crash prediction models from New Zealand, the US, Sweden and Australia. Their results suggested that it was possible to transfer models from one country to the next, but there were a number of differences between countries that needed to be accounted for.

---

2 This review concentrates on post-2004 Australasian crash prediction models.

In their technical note, Turner, Tate and Koorey (2007) identified the important attributes required in a road safety evaluation tool for New Zealand conditions. They reviewed overseas software, considered whether it could be applied in New Zealand, and what local software could be required. They placed particular emphasis on the issues involved in New Zealand adopting the Australian package SIDRA.

Notably, crash modelling work by Turner and Tate was incorporated in the NZTA's *Economic evaluation manual* (2010).

Turner et al (2008) focused on crash prediction models for intersections and noted that a significant proportion of urban crashes occurred at traffic signals, and that many of the 'black spots' in both Australian and New Zealand cities coincided with high-volume and/or high-speed traffic signals. The crash prediction models considered have been thought to have enabled a better understanding of the impact of various factors on safety to be quantified. Turner, Roozenburg and Smith (2009) noted that management of speed was considered an important safety issue at roundabouts. They went on to discuss crash predictions for pedestrian-versus-motor-vehicle and cyclist-versus-motor-vehicle crash types.

Turner and Wood (2009a) presented an overview of the statistical methods they used in their development of crash prediction models, and presented many key findings from New Zealand and Australian crash prediction models.

In another publication, Turner and Wood (2009b) summarised a crash prediction model for intersections. Before describing this model they stated that a large number of crash prediction models had been developed in New Zealand for different road elements and speed limits. Such models were deemed to have potential use, among other things, to predict the reduction in crashes that might result from an engineering improvement.

In the final publication included here by these two authors, Turner and Wood (2009c) appear to have covered the same information that was in a previous paper (2009a), presenting a literature review of New Zealand and Australian crash risk models, and going on to focus on the statistical methods used in their predictive models.

In a PowerPoint presentation, Turner (date unknown) overviewed New Zealand land transport organisations, New Zealand and US crash/accident trends, crash (accident) prediction models, and the application of crash prediction models in economic evaluation.

Roozenburg and Turner (2005) stated that Beca, Carter, Hollings and Ferner Ltd had developed a number of crash prediction models for crashes at signalised intersections. Non-flow variables such as intersection geometry and signal phasing were concluded to be important predictor variables. They went on to present and discuss in detail those models developed for signalised intersections in New Zealand. A predecessor of such models was included in appendix 6 of Transfund NZ's now-superseded *Project evaluation manual* (Transfund NZ 1997).

Harper and Dunn (2005) detailed findings of research to develop more-advanced urban roundabout crash prediction models. Crash prediction model forms and findings from studies in the UK, Australia and New Zealand were briefly outlined. The authors commented that a considerable database had been collated, including traffic movement volumes, geometric site characteristics, and crash data for 95 urban roundabouts throughout New Zealand. Employing this database, advanced conflicting-flow crash prediction models had been developed to predict major vehicle crash types on urban roundabouts in New Zealand in relation to traffic volumes and geometric variables.

Koorey (2006 and 2010) noted that crash prediction models were an increasing feature of rural highway design practice internationally. Koorey went on to consider the issues involved with New

Zealand using the *Interactive highway safety design model* (IHSDM) (FHWA 2006). For an excellent case study of the application of the IHSDM, refer to Bansen and Passetti (2005).

2.2 Australian crash prediction models

The Australian-developed crash prediction models, or models applied to parts of Australia, appear to be associated mainly with Blair Turner, although there are some notable exceptions to this generalisation.

Affum and Goudens (2008) discussed the application of Australia's NetRisk Road Network Safety Assessment Tool as they envisaged it applying to the North Coast Hinterland District in Australia.

Bobevski et al (2007) discussed the application of generalised linear models of road trauma outcomes, to assess the safety benefits of countermeasures in Victorian roads in Australia from 1998 to 2003. They concluded that generalised linear models of crash outcomes as a function of potential explanatory factors needed realistic assumptions to be made about viable functional forms connecting each factor and the outcomes.

Cossens and Cairney (2008) summarised literature covering crash risk and road characteristics. Skid resistance was concluded to be the best-established of the road surface condition characteristics to affect crash risk. This conclusion was in agreement with the crash prediction model of Cenek and Davies (2006).

McInerney et al (2008) focused on risk maps and the AusRAP Star Ratings initiative for state highways in Australia that was launched by the Australian Automobile Association in October 2006, whereby roads are rated with between one and five stars according to their crash risk, based on road design elements that are known to affect crash risk. The analysis presented provided a strong indication of the improvement in crash costs that could be expected as a road network improved from a two-star, to three-star, to four-star, and ultimately a five-star road.

Prinsloo and Chee (2005) provided detail of a project undertaken for the Roads and Traffic Authority, New South Wales, Australia. They defined the approach and method for the derivation and computation of rural road crash rates.

Prinsloo and Goudanas (2003) presented the tools and results achieved from the prediction of the safety performance of rural highways in New South Wales, Australia. The crash rate prediction model setup consisted of a series of base models relying on a plethora of roadway parameters. Notably, road condition variables did not appear to be among them.

We found that Blair Turner had published very widely on the Australian Road Research Board's (ARRB's) NetRisk Manager safety initiative (eg Turner 2007, 2008a and 2008b). Among these publications was the paper of Turner and Jurewicz (2008) who, in their PowerPoint presentation, gave an overview of ARRB's activities, including road safety improvements. They mentioned AusRAP and NetRisk as being among these initiatives.

2.3 Selected international crash prediction models[3]

Chen et al (2006) discussed a non-country-specific hypothetical conceptual framework for a system whereby an in-vehicle system of sensors and computation algorithms would warn the driver of places where the crash risk was relatively high. This approach was judged by the paper's authors to give the driver sufficient time to react promptly, and would potentially promote safe driving and decrease curve-related injuries and fatalities.

Easa and You (2009) described crash prediction models developed using Washington State road data collected from 2002 through to 2005. In total, the authors developed five statistical models for different combinations of three-dimensional alignment (eg curve on crest).

Elvik (2008) compared five techniques for identifying locations that had a high expected number of crashes. These techniques were tested and evaluated by using data for Norwegian roads. It was concluded that hazardous road locations were most reliably identified by the empirical Bayes technique.

Hildebrand et al (2008) presented comparisons of crash prediction estimates for three models with actual crash data for sections of rural two-lane arterial highways in the province of New Brunswick, Canada. All three of the models used were found to overestimate actual crashes. Notably, all of the models did not appear to include any road surface condition variables.

In a US National Highway Traffic Safety Administration report, Liu and Subramanian (2009) concluded that significant factors related to the high risk of fatal single-vehicle run-off-road crashes were driver sleep deprivation, alcohol use, roadway alignment with curves, speeding and adverse weather.

Montella (2010) compared and evaluated various hot-spot identification methods for allocating resources to improve the safety management of roads. He used data from Italian roads and concluded that the Empirical Bayes estimate of total crash frequency performed the most consistently of the methods evaluated.

Montella et al (2008) described the development of a crash prediction model for the rural motorway between Naples and Canona in Italy (the A16). Notably, the model did not appear to include any road surface condition variables.

In their *National Cooperative Highway Research Program* (NCHRP), Pigman and Agent (2007) sought to formalise US Department of Transport (DOT) reconstruction activities and concluded that very few state DOTs conducted crash reconstructions on a routine basis.

3 While some of the papers do not address crash prediction models explicitly, they are included in this review as they cover some of the important risk factors for crashes.

3 Crash risk relationships: 2000–2009 data

3.1 Introduction

This chapter is concerned with presenting the findings of fitting a Poisson regression model to allow prediction of crash rates from road condition and road geometry parameters. Data stored in the NZTA's Road Assessment and Maintenance Management (RAMM) database for the 10-year period 2000 to 2009 was utilised. Most of the statistical modelling is concerned with vehicle crashes in which at least one person has been killed or suffered serious or minor injuries. These are referred to as *casualty crashes* or simply *crashes*.

3.2 The data and its processing

The following sets of data were available to be used for the statistical modelling:

- data collected by the SCRIM[4]+ machine at 10m intervals on each side of the road comprising:
  - mean texture depth in terms of mean profile depth (MPD) for each wheelpath
  - SCRIM coefficient for each wheelpath
  - T10 (TNZ 2002) skid site classification
  - geometry: gradient; curvature; crossfall; GPS coordinates (2010 only)
- data collected by the SCRIM+ machine at 20m intervals on each side of the road, comprising:
  - International Roughness Index (IRI) roughness for left and right wheelpaths
  - 3m, 10m, and 30m wavelength profile variances (2006–2009 only)
  - mean rut depth and standard deviation of rut depth and related measurements
- carriageway data: urban/rural; number of lanes; lane width; estimated average daily traffic (ADT) count
- road names: state highway number; road region
- crash data:
  - crash location (two versions); crash details; movement code etc
  - crash vehicle details
  - crash causes
- annual high speed pavement condition survey details, including survey number, survey year and survey vehicle type and ID.

The data utilised in the statistical modelling detailed in this report pertains to sealed road sections of New Zealand's state highway. These state highways are divided into road sections that can be identified through a unique name and a unique identification number. Road section names are important because they give us the order of the road sections on a state-highway. The road sections are divided into 10m segments for which we have the *SCRIM+ data*. The road sections are also divided

4 SCRIM: Sideways Force Coefficient Routine Investigation Machine

into longer segments for which we have the *carriageway data*. Further details of the data and its processing are given in appendix A.

3.3 The Poisson regression model

The model assumes that the crashes are statistically independent and the number in each 10m segment follows a Poisson distribution. (Of course, for most segments the number will be zero). Fitting is by maximum likelihood.

Suppose each side of each 10m length of road can generate crashes at the rate (per year):

$$\alpha \exp(L)$$ Equation 3.1

where α is the ADT (per side) and L is a linear combination of the road characteristics, being transformations of terms including:

- a constant
- gradient
- curvature
- out-of-context-curve (OOCC) effect
- T10 skid site classification
- skid resistance
- $\log_{10}$(ADT)
- year
- region
- urban/rural classification.

The coefficients in the linear combination are the unknown parameters to be estimated.

Equation 3.1 represents the collective risk or crash density in terms of expected number of crashes per year per 10m lane section of state highway.

As the exponential of L is being used, the actual model is multiplicative. Note that the ADT appears in the model in two places; α in equation 3.1, and as a component of L. These could have been combined into a single term in L. However, by using the formulation in equation 3.1, the component in L is present only if the personal crash risk (or crash rate) , in units of expected number of crashes per 100 million vehicle kilometres, depends on ADT. When there is dependence, this dependence is modelled by the size of the coefficient of log (ADT) in L. The personal crash risk is given by:

$$\frac{10^{10}}{365} \exp(L)$$ Equation 3.2

The *actual* rate of crashes that are reported in a 10m length of road is the average of *generating* rates over the 10m lengths in its immediate neighbourhood (on the same road) and summed over the two sides of the road. In the results reported here the average is over 10m lengths within 100m either side of the length being considered. Typically, this gives an average over 210m on each side of the road. There is no weighting down of the more distant 10m lengths. This averaging allows for error in reporting the location, and the possibility that a crash ends at a location some distance from the piece of road involved in generating the crash.

Because the method involves combining the sides of the road, it is not necessary to know the directions of vehicles involved in the crash.

3.4 Error estimates

The error estimates and significance tests produced by the model are based on the assumption that the values of the *dependent variable*, the number of crashes in each 10m segment each year, are distributed as independent Poisson variables. The results of this research and data provided in appendix B demonstrate that this assumption is not correct.

The method of choice in this situation is to use the *residual deviance* calculated as part of the maximum likelihood estimation to provide a correction to the error estimates. This does not work in the present analysis because the average number of crashes per segment is very small – much less than 1.The discussion in appendix section B.1 suggests error estimates should be increased by around a factor of around 2.3. This corresponds to increasing the critical points for the tests in the analysis of variance tables by a factor of around 5.4.

The increased error is probably due at least partially to an unknown factor that affects a length of road in the same way. It is possible that the ADT estimate is subject to sufficient error to cause a problem.

If this is the case, the 'variables' that vary gradually along a road (such as those derived from SCRIM or IRI) are likely to be subject to the increased error, but variables that change rapidly (such as curvature), and their interactions with other variables, will be less affected. A more advanced analysis could attempt to allow for this extra variability.

3.5 The regression analysis

The regression analysis was carried out using the four categories of crash data described in table 3.1. A more complete description of each of the four crash categories is provided in appendix section A.3.4.

Table 3.1 Subsets of crash data used in the statistical modelling

Group	Criteria
All	All casualty crashes
Wet	All casualty crashes with the road wet field being W or the cause code was 801, 823 or 901
Selected	All casualty crashes with MVMT_IDA being one of A, B, C, D, F[a]
Wet & selected	Satisfying both the wet and selected criteria

a) Refer to MOT (2009), figure 14.

The analysis also includes an interaction term between curvature and the adjusted IRI value. Table 3.2 summarises the predictor variables and any lower and upper bounds that were applied.

The resulting tables of variance for each of the four crash categories investigated are given in tables 3.3–3.6. The analysis of variance tables shows two versions of the chi-squared values. The type III value is such that each variable is tested in the presence of all other variables. This can be misleading if two variables are highly correlated, since both can appear non-significant when tested in the presence of the other. This version does not make sense when you test a main effect when that effect is also part of an interaction term (curvature and IRI in our analyses). The type I version is when each variable is tested only in the presence of the variables above it in the table. The order of variables is arranged so that the most important variables come first, with interactions coming after main effects, but even then, apparent significance can be misleading when variables are highly correlated (as is the case with OOCC and curvature in the analyses presented in this report). In an earlier analysis (Davies et al 2005) the standard errors given by the Poisson model seemed to be underestimated. Appendix B suggests an adjustment to 5.4.

Table 3.2 Predictor variables

Predictor variable	Bounds	Notes
year		discrete variable, 10 levels
region		discrete variable, 14 levels
urban_rural		discrete variable, 2 levels
adj_skid_site		discrete variable, 3 levels
poly3_bound_OOCC	0, 35	3rd degree polynomial of bounded version of OOCC
poly2_bound_log10_abs_curvature	2,4	2nd degree polynomial of bounded version of log of absolute curvature
poly2_log10_ADT		2nd degree polynomial of ADT
poly2_scrim-0.5000		2nd degree polynomial of (scrim – 0.5)
poly3_bound_abs_gradient	4,10	3rd degree polynomial of bounded version of absolute curvature
poly3_bound_adj_log10_iri	-0.3, 1.2	3rd degree polynomial of bounded version of adjusted log IRI
poly2_bound_log10_abs_curvature × poly2_bound_adj_log10_iri	as above	interaction between 2nd degree polynomial of bounded version of absolute curvature and 2nd degree polynomial of bounded version of adjusted log IRI

Table 3.3 Analysis of variance – 'all' casualty crashes

Predictor variable	df	1% pt[5]	Chi-squared	
			Type III	Type I
year	9	21.70	519.00	526.35
region	13	27.70	298.88	665.07
urban_rural	1	6.63	27.81	485.58
adj_skid_site	2	9.21	4693.30	6288.50
poly3_bound_OOCC	3	11.30	454.48	5399.50
poly2_bound_log10_abs_curvature	2	9.21	110.88	459.57
poly2_log10_ADT	2	9.21	576.57	518.17
poly2_scrim-0.5000	2	9.21	221.65	265.24
poly3_bound_abs_gradient	3	11.30	46.33	65.98
poly3_bound_adj_log10_iri	3	11.30	83.13	108.94
poly2_bound_log10_abs_curvature × poly2_bound_adj_log10_iri	4	13.30	107.14	107.14

5 Refers to the confidence limit.

With the exception of *urban_rural* (in the type III column), table 3.3 shows all variables are statistically significant, although *gradient* is fairly marginal.

Table 3.4 Analysis of variance – 'wet' casualty crashes

Predictor variable	df	1% pt	Chi-squared	
			Type III	Type I
year	9	21.70	194.10	134.70
region	13	27.70	221.21	485.06
urban_rural	1	6.63	38.93	8.13
adj_skid_site	2	9.21	684.88	1005.10
poly3_bound_OOCC	3	11.30	246.90	4211.50
poly2_bound_log10_abs_curvature	2	9.21	77.20	428.00
poly2_log10_ADT	2	9.21	141.10	104.84
poly2_scrim-0.5000	2	9.21	389.38	436.88
poly3_bound_abs_gradient	3	11.30	68.25	83.44
poly3_bound_adj_log10_iri	3	11.30	32.17	39.79
poly2_bound_log10_abs_curvature × poly2_bound_adj_log10_iri	4	13.30	42.74	42.74

With reference to table 3.4, the chi-squared value of the *scrim* variable has increased, showing the importance of skid resistance for wet roads. Otherwise, the chi-squared values have mostly decreased, as would be expected with the smaller number of crashes. The *iri* and *curvature* × *iri* terms are not significant when we use the 1% point multiplied by the 5.4 factor discussed in section 3.4.

Table 3.5 Analysis of variance – 'selected' casualty crashes

Predictor variable	df	1% pt.	Chi-squared	
			Type III	Type I
year	9	21.70	430.40	417.60
region	13	27.70	207.98	607.61
urban_rural	1	6.63	98.65	265.56
adj_skid_site	2	9.21	465.85	971.36
poly3_bound_OOCC	3	11.30	382.23	7057.30
poly2_bound_log10_abs_curvature	2	9.21	146.39	711.96
poly2_log10_ADT	2	9.21	710.14	600.84
poly2_scrim-0.5000	2	9.21	285.08	304.36
poly3_bound_abs_gradient	3	11.30	41.35	53.91
poly3_bound_adj_log10_iri	3	11.30	34.04	57.84
poly2_bound_log10_abs_curvature × poly2_bound_adj_log10_iri	4	13.30	59.93	59.93

With reference to table 3.5, the chi-squared values are similar to those for 'all' casualty crashes. The *iri* and *curvature* × *iri* terms are marginal when we use the increased value of the 1% point. The *urban_rural* variable is now significant in the type III version of the chi-squared value. The results indicate that when the types of crashes that tend to be associated with urban areas are removed, the increased safety due to the lower speed limit becomes apparent.

Table 3.6 Analysis of variance – 'wet selected' casualty crashes

Predictor variable	df	1% pt	Chi-squared	
			Type III	Type I
year	9	21.70	166.94	114.38
region	13	27.70	170.96	464.49
urban_rural	1	6.63	75.63	251.10
adj_skid_site	2	9.21	80.89	196.44
poly3_bound_OOCC	3	11.30	194.90	4606.20
poly2_bound_log10_abs_curvature	2	9.21	87.78	570.04
poly2_log10_ADT	2	9.21	181.86	138.04
poly2_scrim-0.5000	2	9.21	429.95	462.80
poly3_bound_abs_gradient	3	11.30	58.00	69.77
poly3_bound_adj_log10_iri	3	11.30	17.23	23.96
poly2_bound_log10_abs_curvature × poly2_bound_adj_log10_iri	4	13.30	32.33	32.33

Table 3.6 shows that the chi-squared values are similar to those for the 'wet' casualty crashes. However, the *scrim* value has increased slightly; the others have generally decreased slightly; the *adjusted skid site* affect seems a lot smaller.

3.5.1 Predicted crash rate graphs

In order to see how each variable in the model affects crash rate, the figures in this section show the crash rate predicted by the model as each variable, in turn, is varied. For the terms not being varied, the values tabulated in table 3.7 have been used.

Table 3.7 Baseline parameter values used in generating crash rate trend plots

Parameter	Baseline value
year	2008
region	R03
urban_rural	R
adj_skid_site	4
OOCC	0
curvature	5000
ADT	1000
gradient	0
scrim	0.5
adj_log10_iri	0.3

Crash rates are in crashes per 100 million vehicle kilometres travelled. The error bounds show 2 standard deviations (roughly 95% confidence) and are based on the Poisson model – so lengths should be roughly doubled, in line with the discussion in appendix B, to adjust for the variance underestimation resulting from the use of the Poisson model. However, these figures are for the overall crash rate and there is some error that is common to all the points on a graph. Observation of the differences shows that the error may be less than is suggested by the graph (after the length has been doubled).

The following figures show all the graphs for the '*all' casualty crashes*, but only a selection for the others. Note that '*all' casualty* means all reported crashes that involve at least one fatal, serious or minor injury.

'W*et' casualty crashes* and '*wet selected' casualty crashes* can occur only when the road is wet. However, they are normalised by the total traffic, not the traffic when the road is wet. Hence the calculated crash rates are much less than for '*all' casualty crashes* and '*selected' casualty crashes*.

In all the crash rate graphs the vertical scale is logarithmic and there is a ratio of 40 between the crash rate at the lower end and the upper end.

Figure 3.1 shows crash rate to have an overall increasing trend over the 8-year period from 2000 to 2007. The downturn in 2008 was possibly due to the global recession.

Figure 3.2 shows some regional variation. It is unknown whether the variation is a real effect or if it is due to different reporting rates. To assist in the reading of figure 3.2, table 3.8 lists the 14 NZTA regions used for regionalising data contained in the RAMM database.

Table 3.8 NZTA regions and their statistical modelling identifier

NZTA region	Identifier for statistical modelling
Northland	R01
Auckland	R02
Waikato	R03
Bay of Plenty	R04
Gisborne	R05
Hawkes Bay	R06
Taranaki	R07
Manawatu-Whanganui	R08
Wellington	R09
Nelson-Marlborough	R10
Canterbury	R11
West Coast	R12
Otago	R13
Southland	R14

Figure 3.1 'All' casualty crash rate versus year

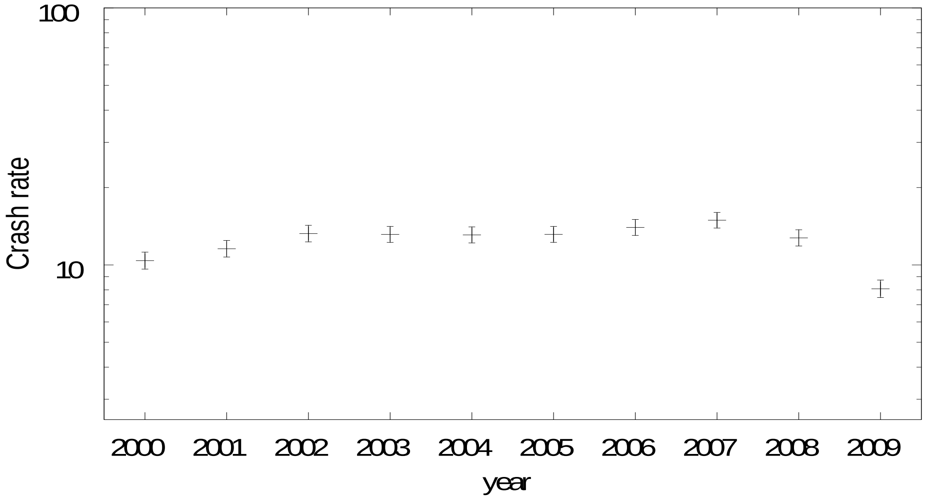


Figure 3.2 'All' casualty crash rate versus NZTA region

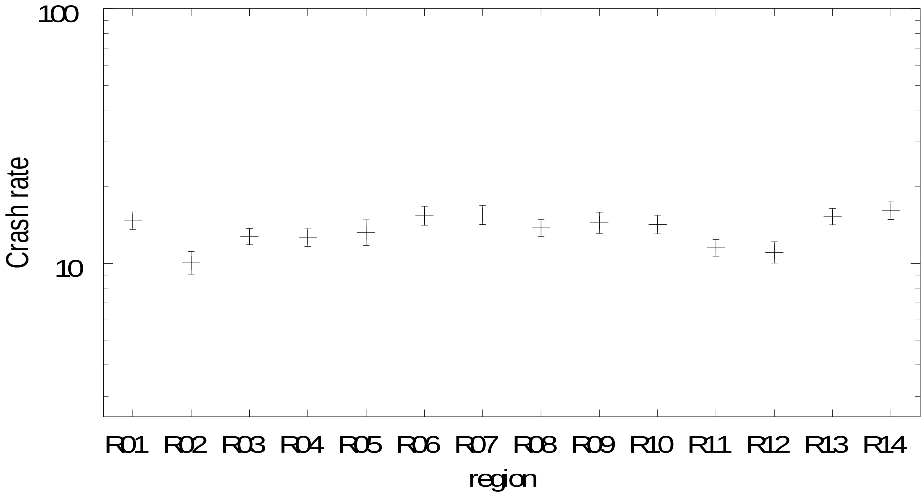
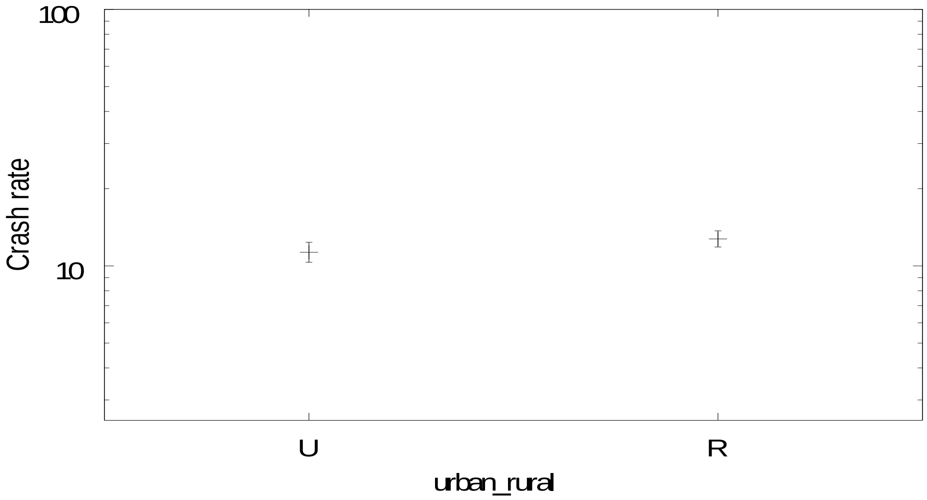


Figure 3.3 shows very little difference in crash rates between urban and rural state highways. Possibly the decreased risk due to a lower speed limit in urban areas is balanced by additional causes of crashes.

Figure 3.3 'All' casualty crash rate versus urban/rural environment, where urban means a speed limit less than or equal to 70km/h and rural means a speed limit greater than 70km/h



With reference to figure 3.4, the 'all' casualty crash rate for adjusted T10 skid site category 4 is substantially lower than for adjusted skid site categories 1 or 3. There needs to be some care in interpreting the actual rates for skid site categories1 and 3, since these are essentially point events and figure 3.4 depicts crash rates per kilometre.

Figure 3.4 'All' casualty crash rate versus adjusted T10 skid site category

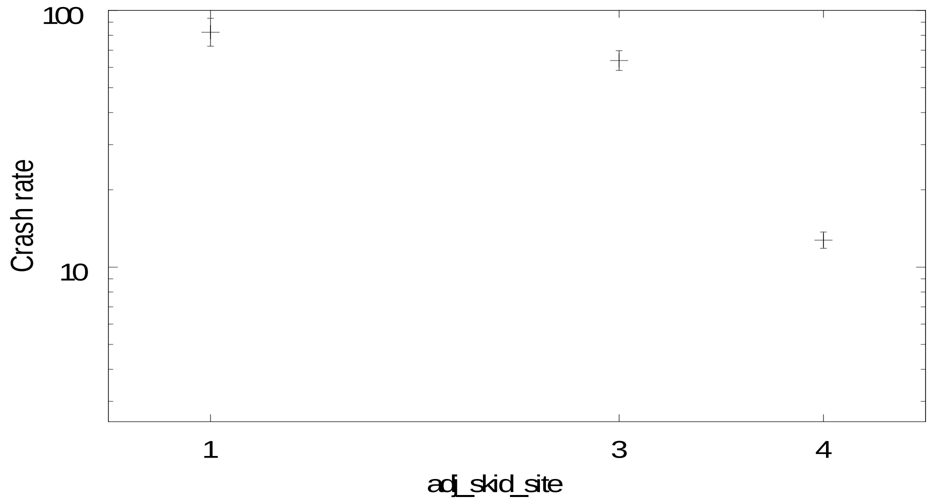
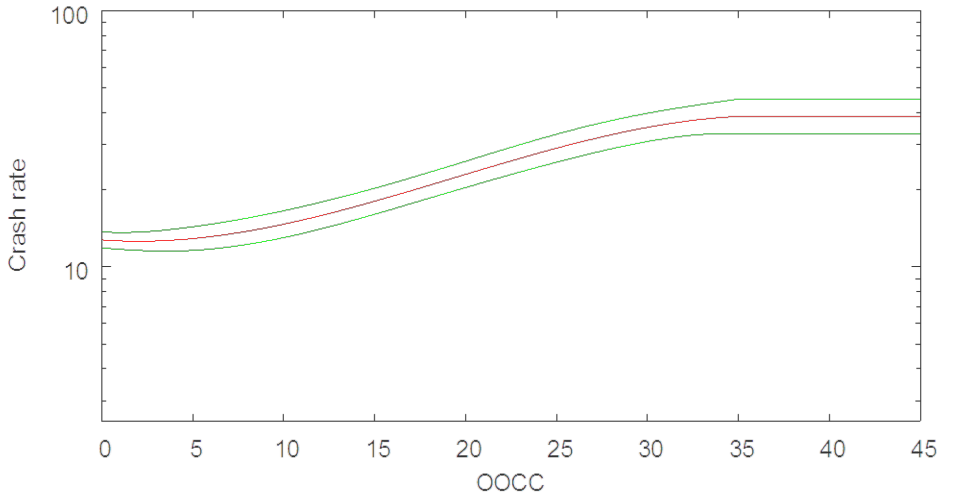


Figure 3.5 shows how the 'all' casualty crash rate varies with OOCC, where OOCC represents the difference between the local speed and the approach speed. The local speed is the average 85 percentile speed over the 10m section of state highway of interest and the preceeding two 10m sections. The approach speed is the average 85 percentile speed over the fifty 10m sections of state highway preceeding the three 10m sections used to calculate the local speed. The procedures used for calculating the 85 percentile speed from road geometry data in RAMM and also OOCC are given in appendix C.

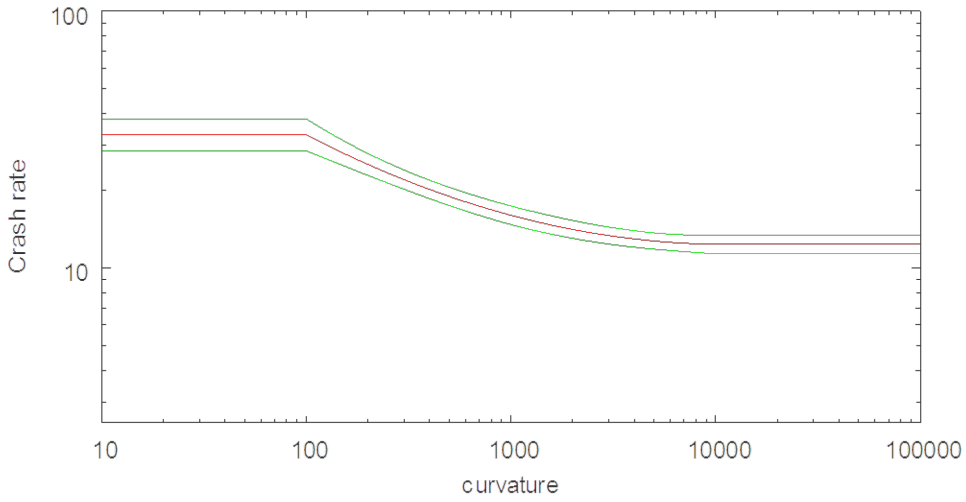
Figure 3.5 'All' casualty crash rate versus OOCC (values where OOCC is greater than or equal to 35km/h are grouped)



In figure 3.5, crash rate increases as the OOCC effect increases. Since we have grouped values where OOCC is greater than or equal to 35, we have a horizontal line for these values.

Figure 3.6 shows the decreasing crash rate as the absolute radius of curvature increases (ie the road becomes straighter). Since we have grouped values where absolute curvature is less than or equal to 100, or greater than or equal to 10,000, we have horizontal lines for these values.

Figure 3.6 'All' casualty crash rate versus horizontal curvature (values where absolute curvature is less than or equal to 100, or greater than or equal to 10,000, are grouped)



Because OOCC and curvature are closely related, figure 3.7 shows what happens when both are varied. Values have been omitted where a combination doesn't make sense – you can't have high radius of curvature and high OOCC. The contours show $\log_{10}$ crash rate, so 1.2 corresponds to a crash rate of 15.8 casualty crashes per 100 million kilometres travelled and 1.8 corresponds to a crash rate of 63 casualty crashes per 100 million kilometres travelled.

Figure 3.7 'All' casualty crashes versus OOCC and curvature

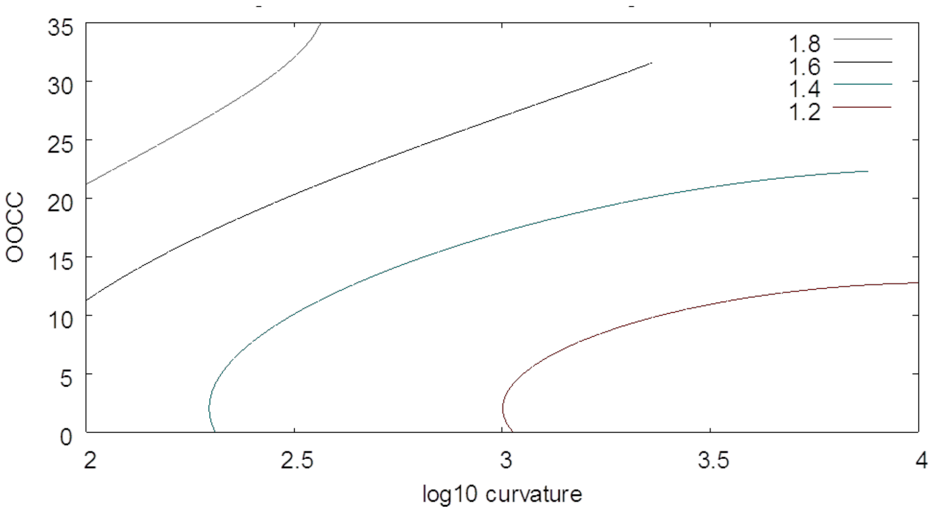


Figure 3.8 shows the crash rate dropping as ADT increases. Even after allowing for sharper bends, lower SCRIM measurements and the increased roughness that you might expect on less-used roads, these roads had a higher incidence of crashes than the high-ADT roads.

One could hypothesise that the effect is due to more risky driving, less policing, less signage and numerous other factors that are not included in the model and that are associated with lower-ADT roads.

Figure 3.8 'All' casualty crashe rate versus ADT

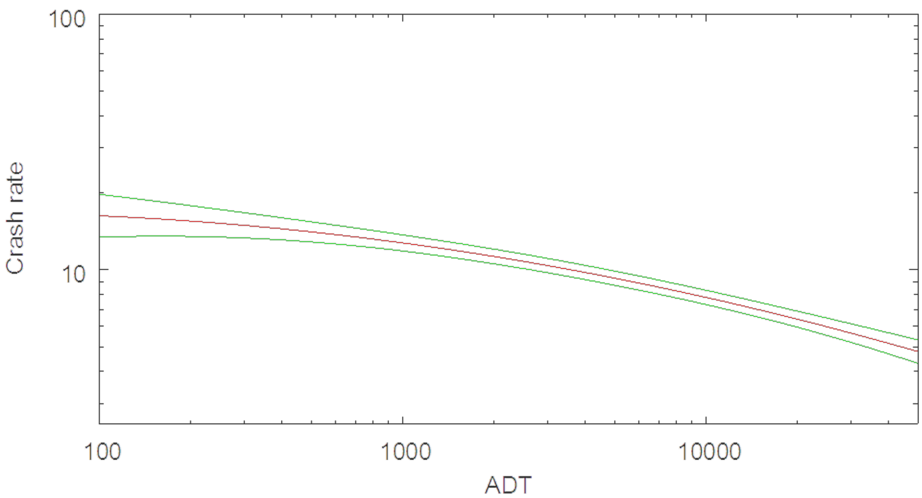


Figure 3.9 shows the crash rate decreasing as the SCRIM skid resistance values increase. The graph is a quadratic curve, although not much curvature is apparent. The error at the extreme ends is quite high, due to a lack of data.

Figure 3.9 'All' casualty crash rate versus SCRIM skid resistance

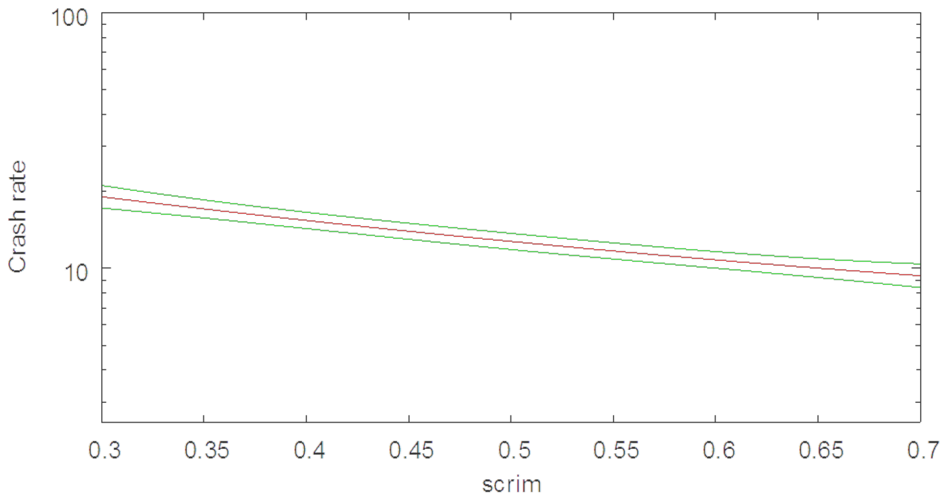
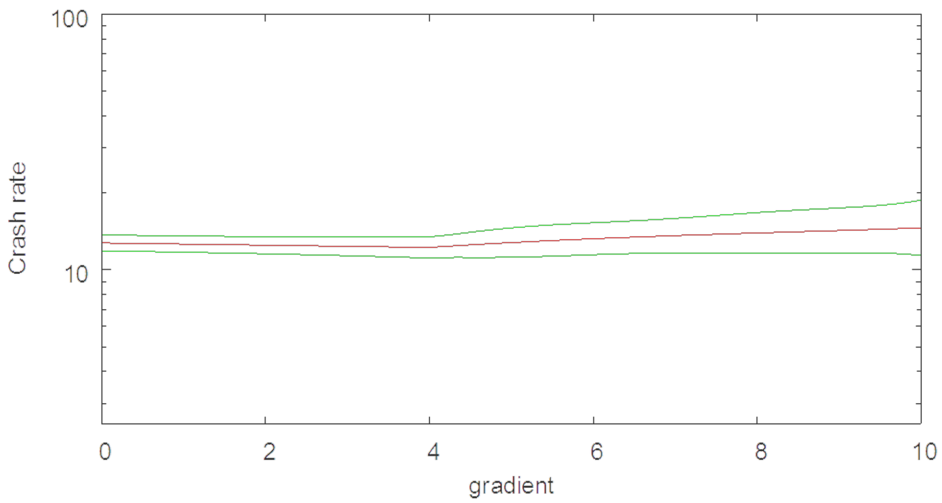


Figure 3.10 shows the crash rate increasing slightly as absolute gradient increases. Because the direction of travel of vehicles involved in crashes is unknown, it is not possible to distinguish between uphill and downhill gradient. Assuming uphill decreases risk and downhill increases it, the effects may cancel each other and there might not be a big effect. Values less than or equal to 4 are grouped, so a horizontal line results for these points.

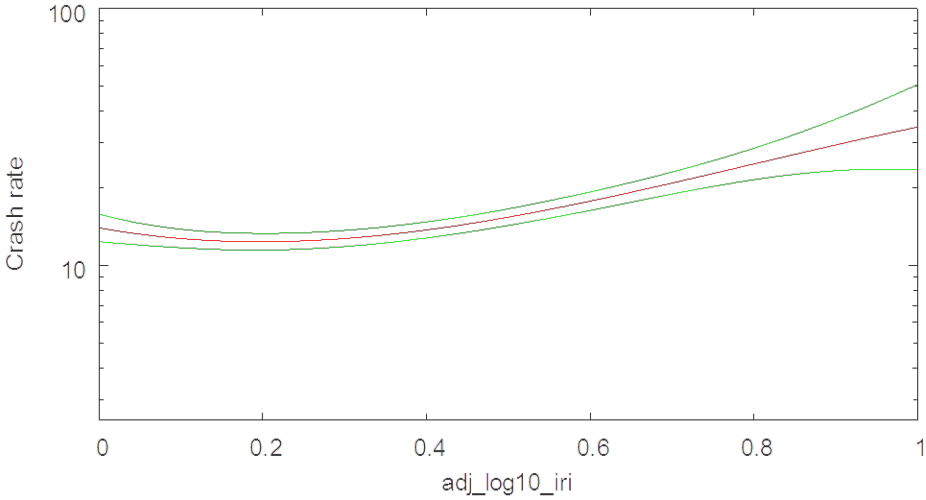
Figure 3.10 'All' casualty crash rate versus gradient



With reference to appendix sections A.3.9 and A.3.10, the IRI-based lane roughness data in the RAMM database appears to be affected by both horizontal curvature and gradient. Therefore, the adjustment as detailed in appendix D has been applied to the IRI data to reduce this measurement effect.

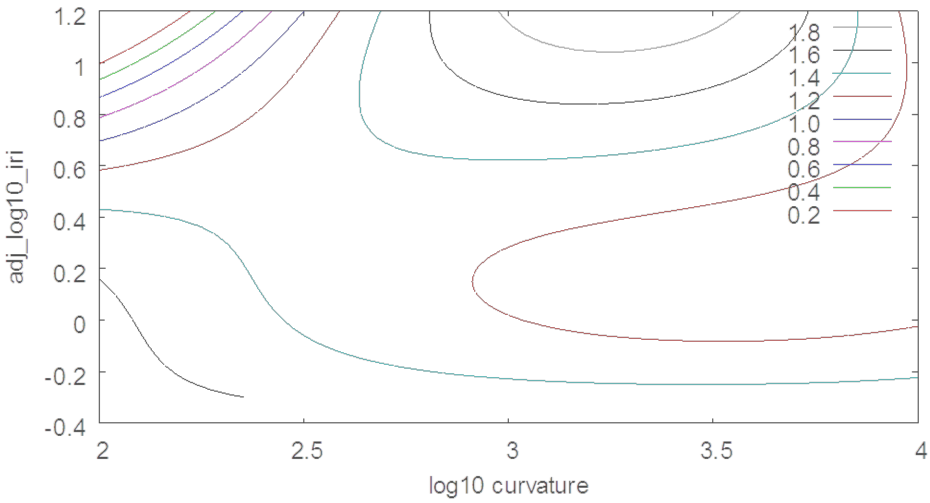
Figure 3.11 shows the relationship between crash rate and the logarithm (base 10) of the adjusted lane roughness. At low levels of roughness there is little effect. From an *adjusted* log_{10} *IRI* value of 0.3, corresponding to 2mm/m IRI or approximately 50 NAASRA counts/km, it becomes apparent that *adjusted IRI* has an increasing effect on crash rate.

Figure 3.11 'All' casualty crash rate versus lane roughness (*adjusted log_{10} IRI*)



To see the interaction between curvature and *adjusted IRI* both effects need to be looked at together, resulting in the contour plot shown as figure 3.12.

Figure 3.12 Contour plot of all casualty crash rate as curvature and *adjusted IRI* are varied



It is easier to understand the relationship if the crash rate is graphed against the *adjusted IRI* for a selection of values of curvature. In figure 3.13, crash rate is plotted against road sections with radii of curvature 10,000, 3000, 1000, 300, and 100, all with *OOCC* = 0.

In figure 3.13, the contour plots take the *adjusted log_{10} IRI* from -0.3, rather than 0 as in figure 3.11. There is very little data with the *adjusted log_{10} IRI* less than 0, so the increasing values here are possibly a quirk of the polynomial fit.

Figure 3.13 The relationship between 'all' casualty crash rate and the *adjusted IRI* for selected values of curvature

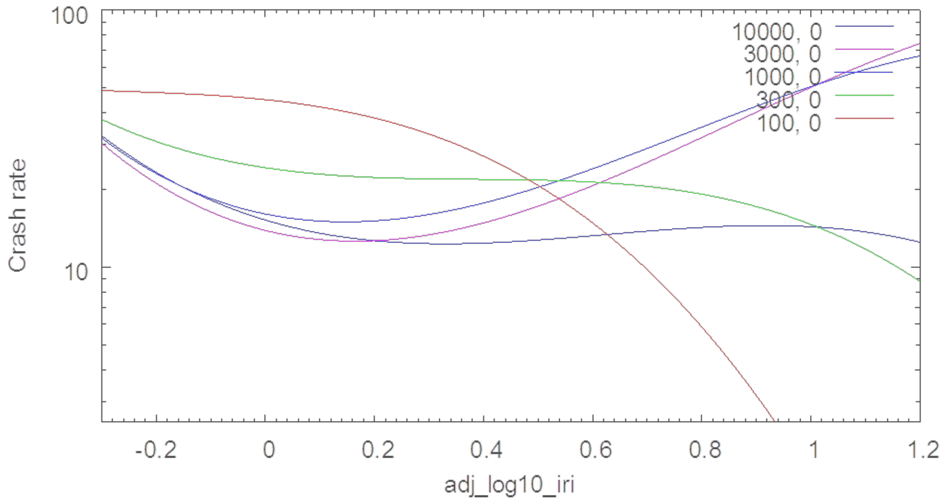


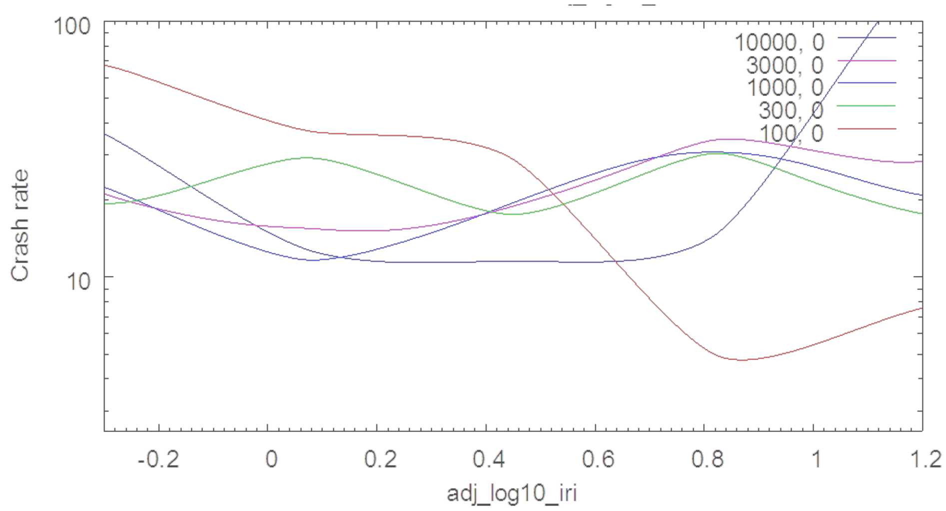
Figure 3.13 suggests that:

- for straight or near-straight roads, roughness is not very relevant
- for modestly curved roads, roughness increases crash risk – the contour plot suggests a range of radii of curvature of 500–5000m for roughness, causing increased crash risk
- for very curved roads, roughness is associated with decreased crash risk, which could be associated with lower speeds – but one should be very wary about interpreting this result
- possibly, very low roughness may be associated with an increase in crash risk (possibly low roughness encourages drivers to go faster, or possibly low roughness is associated with other hazards – for example, bridge decks).

To see whether the modelling of the effect of curvature and *adjusted IRI* was being constrained by the use of polynomial functions, the *curvature* and *adjusted IRI* terms have been replaced by a thin plate spline (tps) function of these two terms. A thin plate spline is like a higher-dimensional analogy to the one-dimensional spline used for fitting curves. A thin plate spline based on a 5×5 array of knots covering the range of *log_{10} curvature* from 2 to 4 and *adjusted log_{10} iri* from -0.3 to 1.2 is applied. This gives 24 degrees of freedom for the combined *log_{10} curvature* and *adjusted log_{10} IRI* term as opposed to the 9 assigned to *log_{10} curvature* and *adjusted log_{10} IRI* and their interaction as shown in figures 3.12 and 3.13. Following this approach provides more flexibility at the expense of more random error. Figure 3.14 shows the resulting contour plot.

With reference to figure 3.14, most of the *adjusted log_{10} IRI* data is between 0.0 and 0.8 (*adjusted IRI* between 1 and 6.3). In this range the results are similar to those using the polynomial functions, thereby confirming that the modelling has not been constrained by the adoption of polynomial functions.

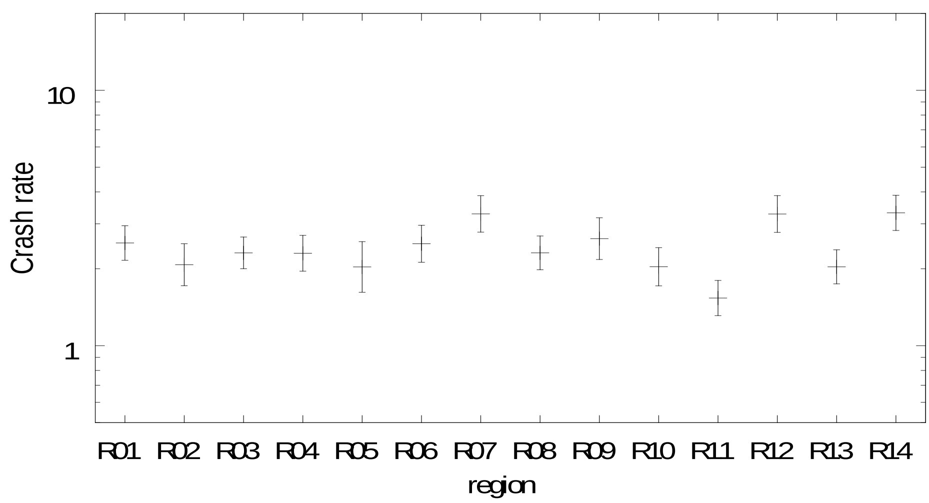
Figure 3.14 'All' casualty crash rate versus *adjusted iri* with thin plate spine interaction



3.5.1.1 'Wet' casualty crash rate relationships

Figure 3.15 shows that crash rates in wet conditions are similar to the corresponding plot for all crashes in figure 3.2. The confidence intervals are longer, reflecting the smaller number of crashes. The crash rate for R12 (West Coast) is somewhat higher, possibly a reflection of the higher rainfall in this region of New Zealand.

Figure 3.15 'Wet' casualty crash rate versus NZTA region



Comparing figure 3.16 with figure 3.5, it can be seen that the relationship between 'wet' casualty crash rates and OOCC is roughly similar to that for 'all' casualty crashes.

Figure 3.16 'Wet' casualty crash rate versus OOCC

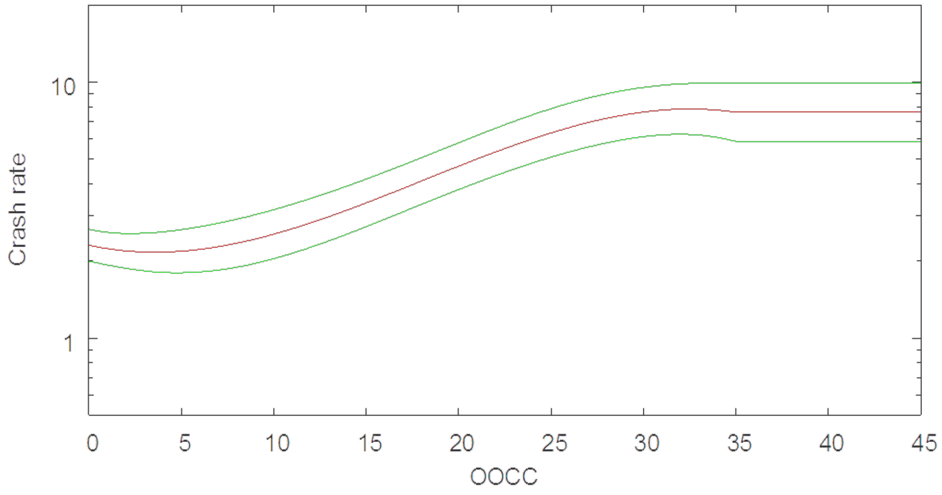
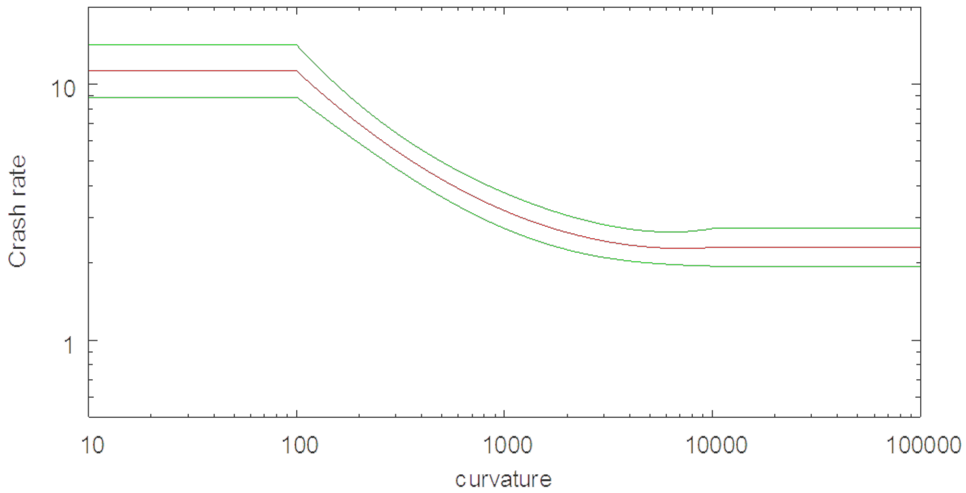


Figure 3.17 shows the effect of curvature on the 'wet' casualty crash rate to be only apparent within certain bounds. The effect is larger when compared with the 'all' casualty crash rates (refer to figure 3.6).

Figure 3.17 'Wet' casualty crash rate versus horizontal curvature



As expected, figure 3.18 shows a declining 'wet' casualty crash rate as the SCRIM skid resistance value increases. The effect is noticeably larger when compared with the 'all' casualty crash rate (refer to figure 3.9).

Figure 3.18 'Wet' casualty crash rate versus SCRIM skid resistance

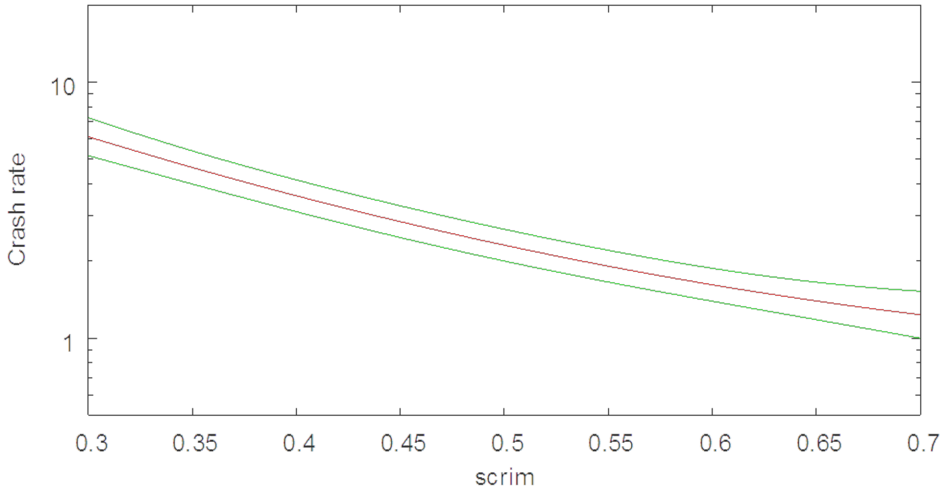
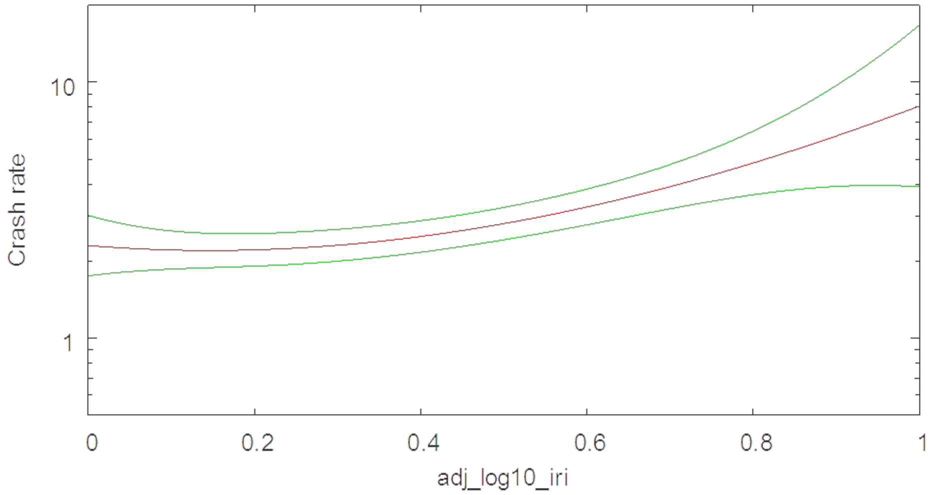


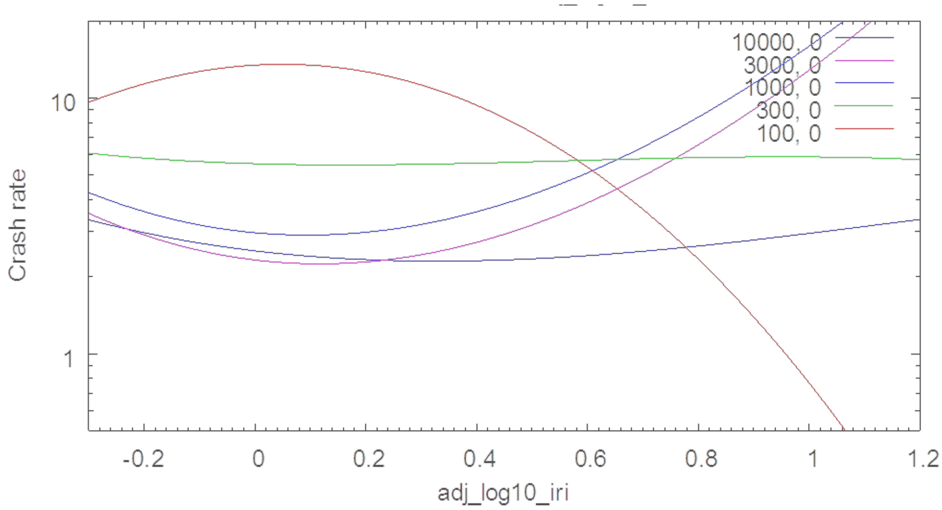
Figure 3.19 shows the effect of the *adjusted IRI* on the 'wet' casualty crash rate seems a little larger when compared with the 'all' casualty crash rate (refer to figure 3.11).

Figure 3.19 'Wet' casualty crash rate versus lane roughness (*adjusted log_{10} IRI*)



Comparing figure 3.20 with figure 3.13, the combined effects of roughness and curvature on the 'wet' casualty crash rate is similar to the 'all' casualty crash rate, except under wet conditions the effects seem to be a little larger. However these effects are only marginally statistically significant.

Figure 3.20 The relationship between the 'wet' casualty crash rate and the *adjusted IRI* for selected values of curvature



3.5.1.2 'Selected' casualty crash rate relationships

Figure 3.21 shows that the 'selected' casualty crash rates in urban areas is slightly lower than in rural areas. This could be possibly due to lower speed limits, but there might be other factors such as better roads and better policing.

Figure 3.21 'Selected' casualty crashes: urban/rural

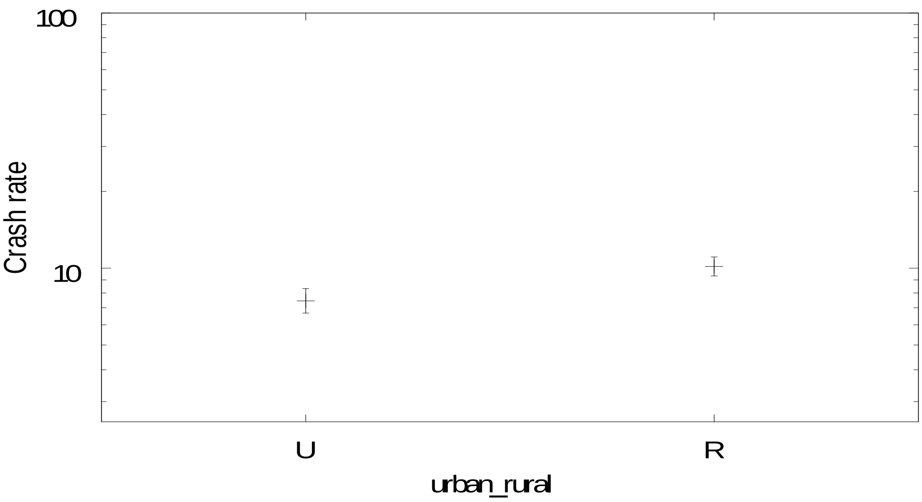


Figure 3.22 shows T10 skid sites to have a smaller effect on the 'selected' casualty crash rate than on the 'all' casualty crash rate (refer to figure 3.4), presumably because some types of intersection crashes are excluded by the selection criteria.

Figure 3.22 'Selected' casualty crash rate versus adjusted T10 skid site

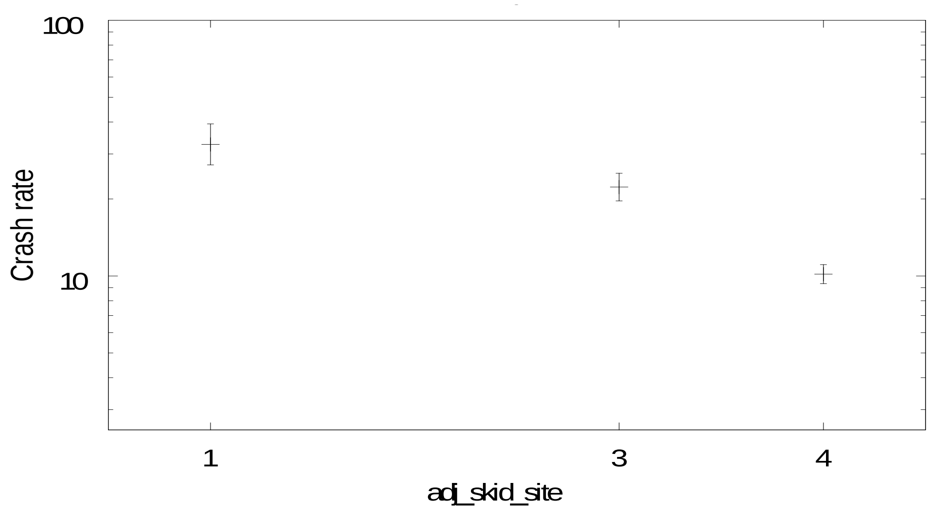
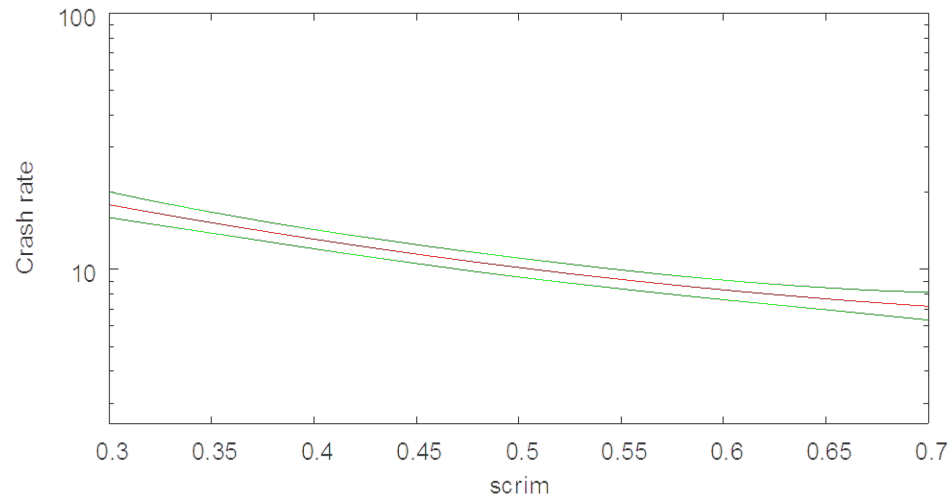


Figure 3.23 shows SCRIM skid resistance to have a larger effect on the 'selected' casualty crash rate than on the 'all' casualty crash rate (refer to figure 3.9), presumably because some types of non-skid-related crashes are excluded by the selection criteria.

Figure 3.23 'Selected' casualty crash rate versus SCRIM skid resistance



The effect of lane roughness, and lane roughness in combination with curvature, on the 'selected' casualty crash rate are shown in figures 3.24–3.26. These figures can be directly compared with figures 3.11–3.13, which show the corresponding 'all' casualty crash rate relationships. As can be seen, the relationships for the 'selected' casualty crash rate mirror those for the 'all' casualty crash rate.

Figure 3.24 'Selected' casualty crash rate versus lane roughness (*adjusted* $\log_{10}$ *IRI*)

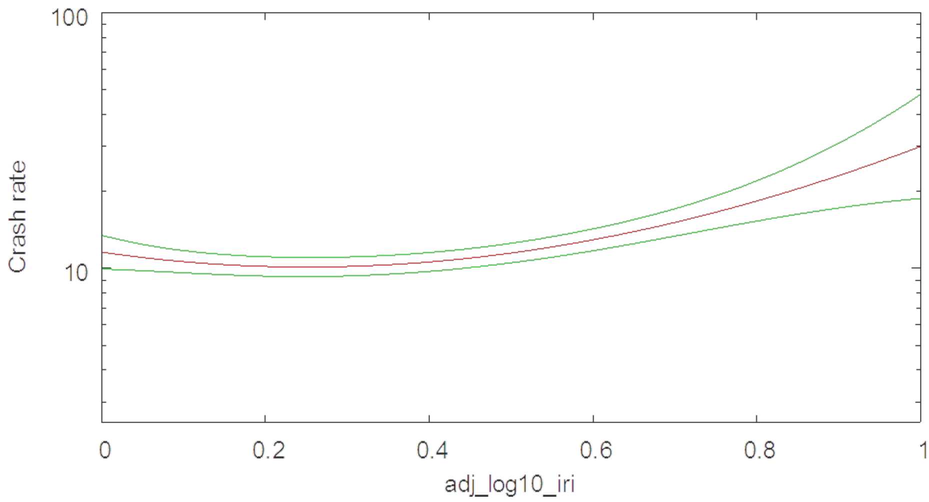


Figure 3.25 Contour plot of the 'selected' casualty crash rate as curvature and *adjusted IRI* are varied

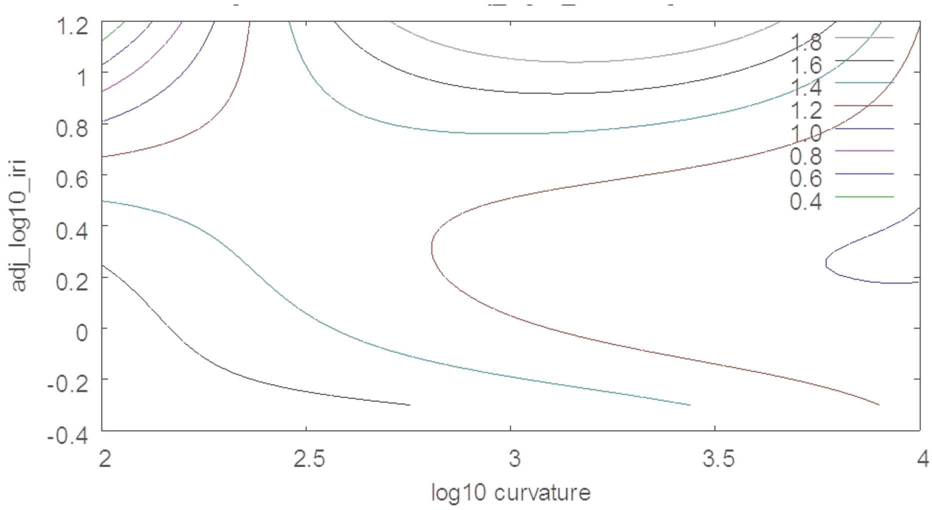
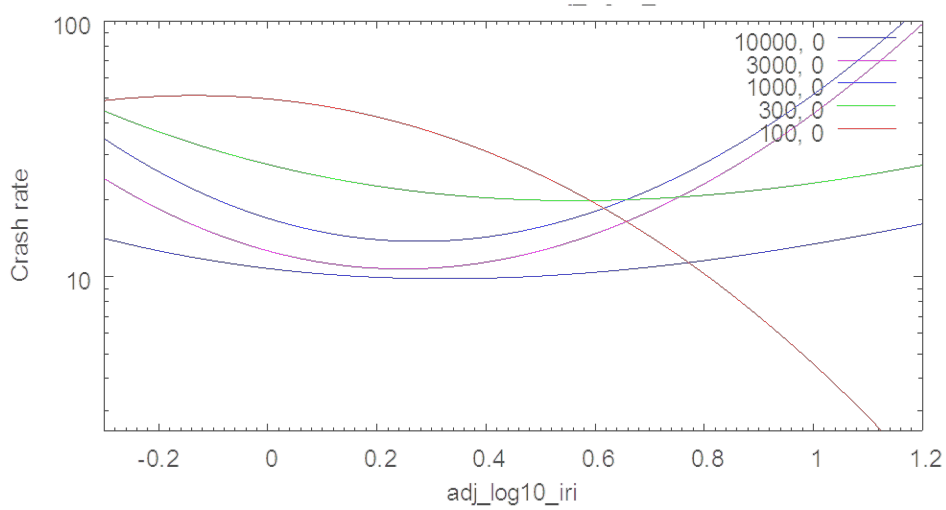
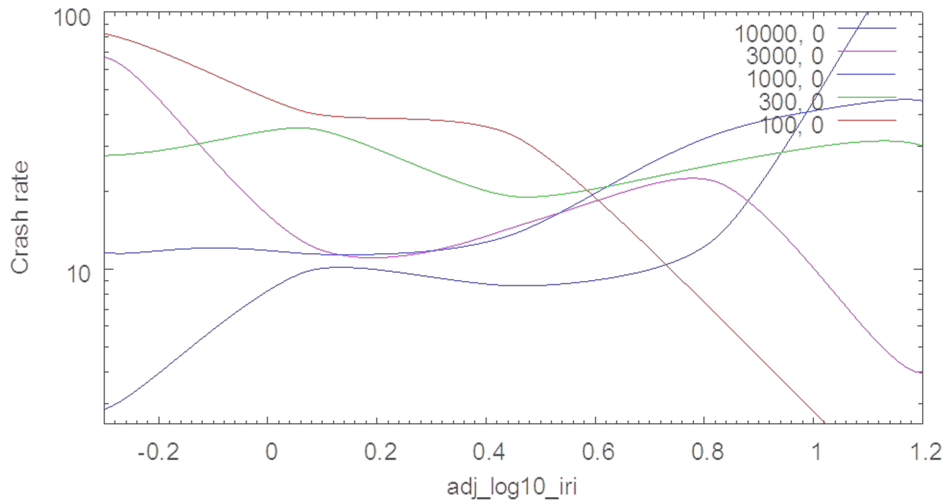


Figure 3.26 The relationship between the 'selected' casualty crash rate and the *adjusted IRI* for selected values of curvature



The analysis was repeated with the 5×5 thin plate spline and again the results, shown in figure 3.27, are similar to those for the 'all' casualty crash rate (refer to figure 3.14).

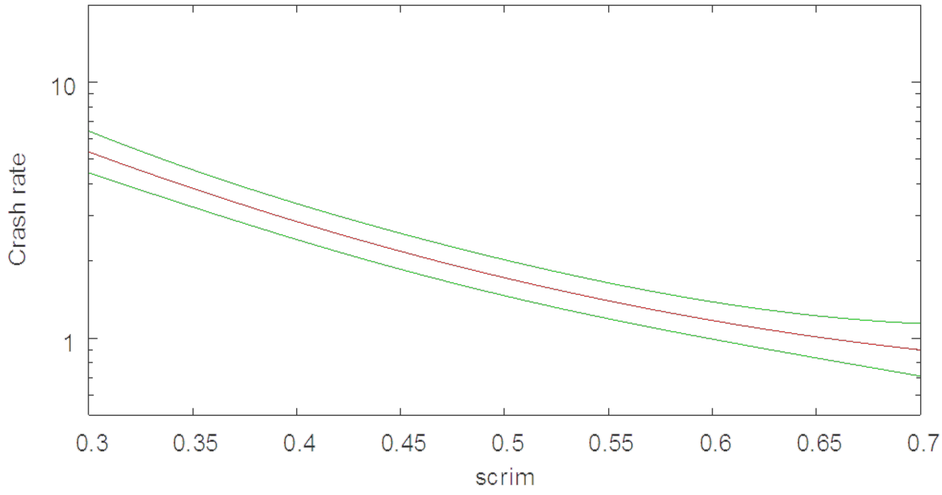
Figure 3.27 'Selected' casualty crashes versus *adjusted iri* with with thin plate spine interaction



3.5.1.3 'Wet selected' casualty crash rate relationships

Figure 3.28 shows that the effect of SCRIM skid resistance on the 'wet selected' casualty crash rate is similar to that for the 'wet' casualty crash rate (refer to figure 3.18).

Figure 3.28 'Wet' selected casualty crashes: SCRIM measurements



The effect of lane roughness, and lane roughness in combination with curvature, on 'wet selected' casualty crash rate are shown in figures 3.29 and 3.30. These figures can be directly compared with figures 3.19 to 3.20, which show the corresponding 'wet' casualty crash rate relationships. As can be seen, the relationships for the 'selected wet' casualty crash rate mirror those for the 'wet' casualty crash rate.

Figure 3.29 'Wet selected' casualty crash rate versus lane roughness (*adjusted log_{10} IRI*)

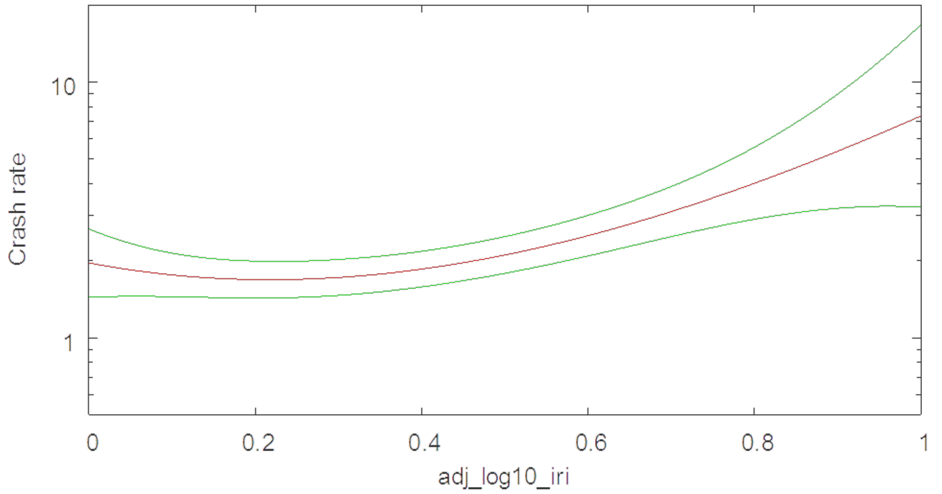
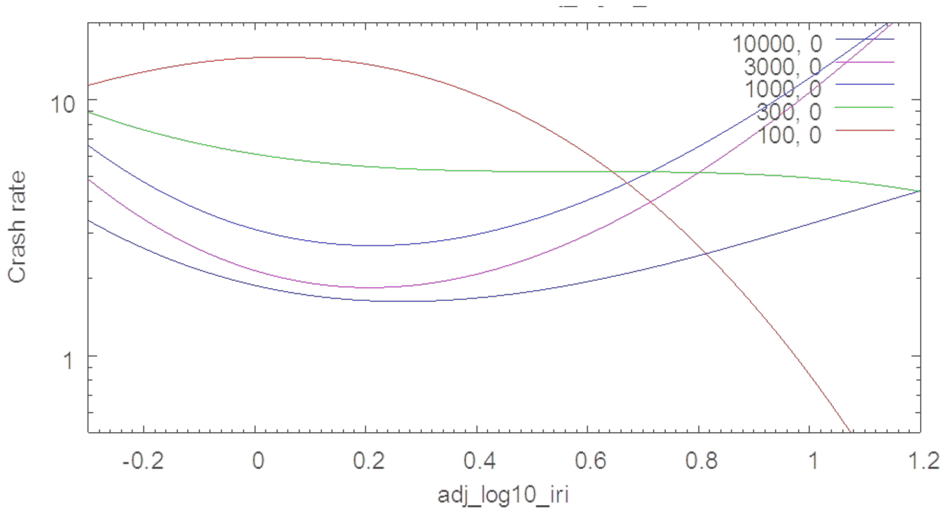


Figure 3.30 The relationship between the 'wet selected' casualty crash rate and the *adjusted IRI* for selected values of curvature



3.6 Calculation of predicted crash rates

The regression coefficients for the 'all', 'wet', 'selected' and 'wet selected' crash rate prediction models are given in appendix E, along with an example calculation using the 'all' casualty crash rate model. Furthermore, Microsoft Excel™ spreadsheets that automate the calculation procedure have been prepared. These spreadsheets can be downloaded from the NZTA website (www.nzta.govt.nz/resources/research/reports/477/index.html) and are named *fitted_all.xls*, *fitted_wet.xls*, *fitted_sel_xls* and *fitted_wet_sel.xls.* There are comments in the spreadsheets explaining the calculations.

The road data is entered into the yellow region in the worksheet 'Values'. If an invalid value is entered into cells C9 to C12, the word 'Error' will appear in column D. Remember that if the year 2009 is entered, low results will be obtained because there is only partial data for 2009. The value in cell C25 should be 'y' unless the user doesn't want to adjust the IRI value – for example if the adjusted IRI value has been entered into cell C18. The column 'Default' shows the default values used to generate the graphs presented in section 3.5.1.

The adjustment to the IRI is carried out in worksheet 'Adj IRI' and the results reported in rows 23 and 24 in the worksheet 'Values'. Cell C23 shows the amount to be subtracted from the $log_{10}(IRI)$ and C24 shows what *IRI* would be multiplied by.

Lines 29 to 40 in worksheet 'Values' calculate the transformed and bounded values of the predictor variables.

The actual calculation of the predicted crash rate is carried out in the worksheet 'Calculation'.

The result is reported in B42 on the 'Values' sheet.

Appendix B investigates the level of agreement between the number of predicted crashes and the number of actual crashes.

4 Examples of applications of the model

4.1 A 'what-if' study for skid resistance

This section provides an illustration of the use of the model for investigating the effect of improving skid resistance.

The 'what-if' study is for 2008. Table 4.1 summarises the actual road and crash data for 2008 for the road segments classified as T10 skid site category 2.

Table 4.1 Characteristics of T10 skid site category 2 sections of the state highway network as at 2008

Length of road sides (lane-km)	2056
Actual number of injury crashes	567
Actual number of 'wet' injury crashes	207

Let us consider increasing the skid resistance of the sections of state highway that are classified as T10 skid site category 2 (ie curves with less than 250m radius or gradient greater than 10%; investigatory SCRIM skid resistance level 0.5). One of the following minimum levels for SCRIM skid resistance is chosen: 0.4, 0.5, 0.6. If the actual SCRIM skid resistance value is below this, the road surface is upgraded to raise the SCRIM skid resistance value to this amount; otherwise it is left the same. This can be done for all the sections of state highway that fall short of the minimum SCRIM skid resistance level, or only those with ADT above some prescribed level.

Table 4.2 shows the reduction in the predicted number of crashes for 2008 if the upgrade had been done before 2008.

Table 4.2 Effect of different levels of SCRIM skid resistance on injury crashes occurring on sections of state highway classified as T10 skid site category 2

			'All' crashes		'Wet' crashes	
Minimum SCRIM	**Fix for ADT≥**	**Fix length (lane-km)**	**Predicted crashes**	**Saved crashes**	**Predicted crashes**	**Saved crashes**
0	0	0	565	0	219	0
0.4	0	440	547	18	194	24
0.5	0	1292	506	59	159	60
0.6	0	1911	452	113	124	95
0.4	1000	350	548	17	195	23
0.5	1000	955	509	56	162	57
0.6	1000	1336	460	105	130	88
0.4	5000	69	558	7	210	9
0.5	5000	179	541	24	196	22
0.6	5000	228	521	44	184	35

The first row of table 4.2 is for no upgrade; the rest are for the values of minimum SCRIM skid resistance value and ADT shown in the first two columns. It is supposed that the two sides of the road are handled independently and the column *fix length* shows the length of side that needs to be upgraded. The analysis is carried out using the model for 'all' injury crashes (refer to appendix table

E.1) and the model for all ‘wet’ injury crashes (refer to appendix table E.2). Table 4.2 shows the predicted number of crashes for each modelled situation and then the reduction in the number of crashes compared with the first line.

Table 4.2 suggests that raising the investigatory SCRIM skid resistance level from 0.5 to 0.6 on T10 skid site category 2 sections of state highway with ADT ≥1000 is likely to be a very effective safety intervention.

4.2 ‘What-if’ study for IRI

This section provides an illustration of the use of the model for investigating the effect of reducing roughness.

Let us consider decreasing roughness on sections of state highway with absolute radius of curvature greater than 500m and less than 5000m. One of the following maximum levels for IRI is chosen: 2.00, 3.98, 7.94. If the actual IRI value is above this, the road surface is upgraded to lower the IRI value to this amount; otherwise it is left the same. This can be done for all the roads that are above the maximum IRI level, or only those with ADT above some prescribed level.

Again, the what-if study is for 2008. Table 4.3 summarises the actual road and crash data for 2008 for state highway sections in the curvature range 500–5000m.

Table 4.3 Characteristics of state highway sections with horizontal curvature in the range 500–5000m as at 2008

Length of road sides (lane-km)	7348
Actual number of injury crashes	1229

Table 4.4 shows the reduction in the predicted number of crashes for 2008 if the upgrade had been done before 2008. The table is using adjusted IRI (refer to appendix D), but for this curvature range the adjustment has little effect.

Table 4.4 Effect on different IRI lane roughness on all injury crashes occurring on sections of state highway with horizontal curvature between 500m and 5000m

Maximum adjusted IRI	Fix for ADT≥	Fix length (lane-km)	Predicted crashes	Saved crashes
	0	0	1182	0
7.94	0	30	1180	2
3.98	0	815	1148	33
2.00	0	4212	1048	134
7.94	1000	20	1180	1
3.98	1000	602	1150	31
2.00	1000	3242	1055	126
7.94	5000	5	1181	1
3.98	5000	169	1163	19
2.00	5000	953	1106	76

The first row of table 4.4 is for no upgrade; the rest are for the values of maximum adjusted IRI value and ADT shown in the first two columns. It is supposed that the two sides of the road are handled independently and the column *fix length* shows the length of side that needs to be upgraded. The analysis is carried out using the model for all crashes (refer to appendix table E.1). Table 4.4 shows the predicted number of crashes for each modelled situation and then the reduction in the number of injury crashes compared with the first line.

Table 4.4 suggests that reducing lane roughness from 7.94mm/m IRI to 2mm/m IRI on relatively straight (horizontal curvature between 500m and 5000m) and heavily trafficked (ADT ≥ 5000) sections of state highway is also likely to be a very effective safety intervention.

5 Variations on the model

The following sections explore a number of minor variations to the crash risk model presented in section 3 and appendix E.

5.1 Interaction between curvature and SCRIM skid resistance

The results indicate that an interaction between curvature and the SCRIM skid resistance measurements might be present. The question is: Does the effect of the SCRIM skid resistance values change according to curvature? In the analysis presented, a very simple $\log_{10}$(curvature) × SCRIM skid resistance = interaction is used for both the 'all' crash data and the 'wet' crash data. In both cases, it is fairly marginal whether the results are statistically significant. In the analysis of the variance tables below (ie tables 5.1 and 5.2), the last line corresponds to the interaction term. The amount needed to inflate the significance levels is expected to be rather less than the 5.4 used earlier, since curvature is a rapidly changing predictor variable.

With reference to tables 5.1 and 5.2, the chi-squared value for the curvature × SCRIM term is 22.1 for 'all' casualty crashes and 28.4 for 'wet' casualty crashes. Therefore the effect should be regarded as statistically significant.

Table 5.1 Analysis of variance, curvature × SCRIM term included – 'all' casualty crashes

Predictor variable	df	1% pt	Chi-squared	
			Type III	Type I
year	9	21.70	514.67	533.09
region	13	27.70	301.76	673.14
urban_rural	1	6.63	27.54	493.38
adj_skid_site	2	9.21	4693.10	6298.40
poly3_bound_OOCC	3	11.30	455.22	5494.10
poly2_bound_log10_abs_curvature	2	9.21	107.40	470.42
poly2_log10_ADT	2	9.21	575.29	519.67
poly2_scrim-0.5000	2	9.21	61.69	262.21
poly3_bound_abs_gradient	3	11.30	49.31	66.26
poly3_bound_adj_log10_iri	3	11.30	82.23	109.61
poly2_bound_log10_abs_curvature × poly2_bound_adj_log10_iri	4	13.30	110.42	106.86
bound_log10_abs_curvature × scrim-0.5000	1	6.63	22.11	22.11

Table 5.2 Analysis of variance, curvature × SCRIM term included – 'wet' casualty crashes

Predictor variable	df	1% pt	Chi-squared	
			Type III	Type I
year	9	21.70	191.62	138.25
region	13	27.70	228.46	508.26
urban_rural	1	6.63	39.01	6.780
adj_skid_site	2	9.21	687.16	1014.30
poly3_bound_OOCC	3	11.30	246.82	4436.60
poly2_bound_log10_abs_curvature	2	9.21	70.93	464.02
poly2_log10_ADT	2	9.21	137.65	104.25
poly2_scrim-0.5000	2	9.21	104.52	416.16
poly3_bound_abs_gradient	3	11.30	73.61	84.39
poly3_bound_adj_log10_iri	3	11.30	32.22	42.28
poly2_bound_log10_abs_curvature × poly2_bound_adj_log10_iri	4	13.30	46.79	42.72
bound_log10_abs_curvature × scrim-0.5000	1	6.63	28.37	28.37

Figures 5.1 and 5.2 show the the effect of SCRIM skid resistance values on the 'all' and 'wet' casualty crash rates respectively for various specified values of horizontal radius of curvature ranging from 100m to 10,000m. In all cases the OOCC effect has been set to 0. These figures suggest the effect of the SCRIM skid resistance values is higher on curves than on straight or near-straight roads.

Figure 5.1 Effect of skid resistance on the 'all' casualty crash rate as a function of road curvature

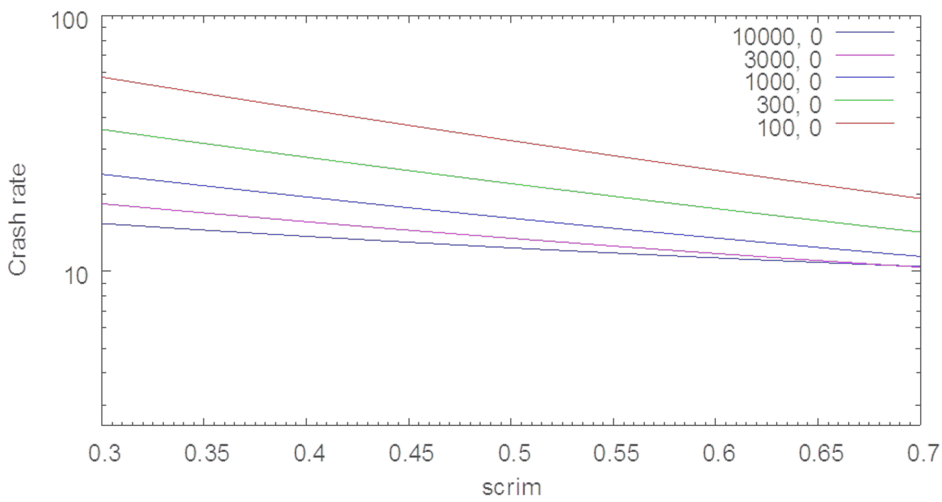
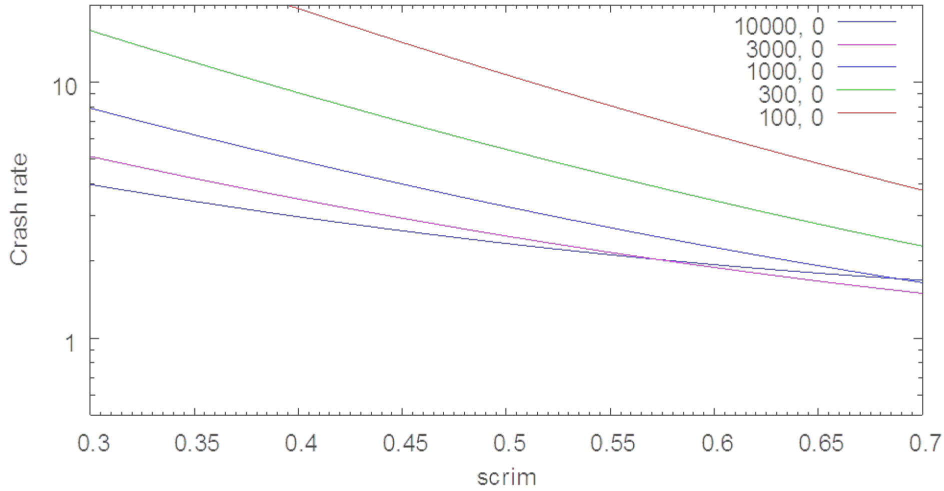


Figure 5.2 Effect of skid resistance on the 'wet' casualty crash rate as a function of road curvature



5.2 Interaction between curvature and texture

The analysis for 'all' casualty crashes with additional terms for mean texture depth and its interaction with curvature are repeated. Table 5.3 summarises the results of the analysis of variance for only the terms involving mean texture depth.

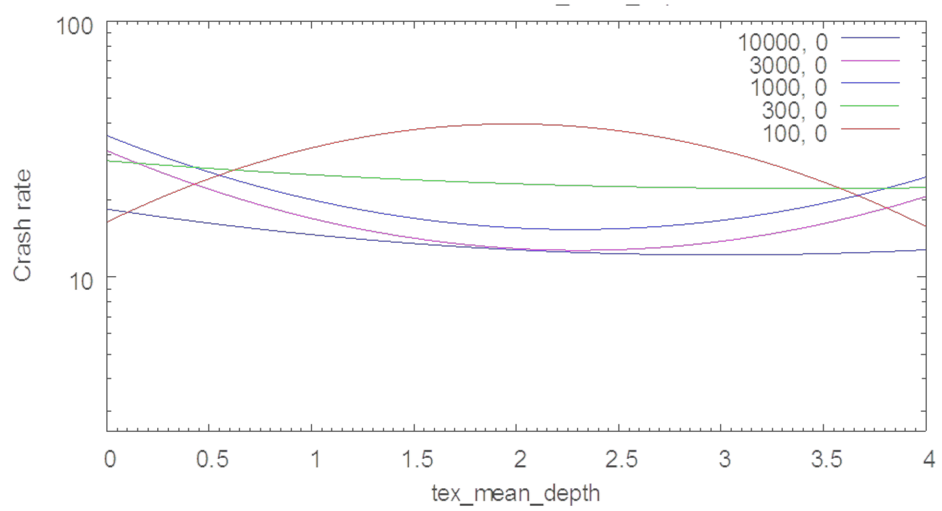
Table 5.3 Analysis of variance – 'all' causalty crashes: mean texture depth terms

Predictor variable	df	1% pt	Chi-squared	
			Type III	Type I
poly2_bound_tex_mean_depth	2	9.21	29.01	75.17
poly2_bound_log10_abs_curvature × poly2_bound_tex_mean_depth	4	13.30	58.92	58.92

With reference to table 5.3, both texture terms appear to be statistically significant.

Figure 5.3 shows plots of the 'all' casualty crash rate versus mean texture depth for various horizontal curvatures.

Figure 5.3 Effect of mean texture depth on the 'all' casualty crash rate as a function of curve radius



With reference to figure 5.3, the plot for 100m horizontal curvature, and the upwards slope of the plots for the other curvatures at high values of mean texture depth, may be a result of polynomial fit.

These results concerning texture should be treated with caution for the reasons stated in appendix section A.3.16.

5.3 Interaction between curvature and mean rut depth

The analysis for 'all' casualty crashes with additional terms for the mean rut depth and its interaction with curvature are repeated. Table 5.4 summarises the results of the analysis of variance for only the terms involving mean rut depth.

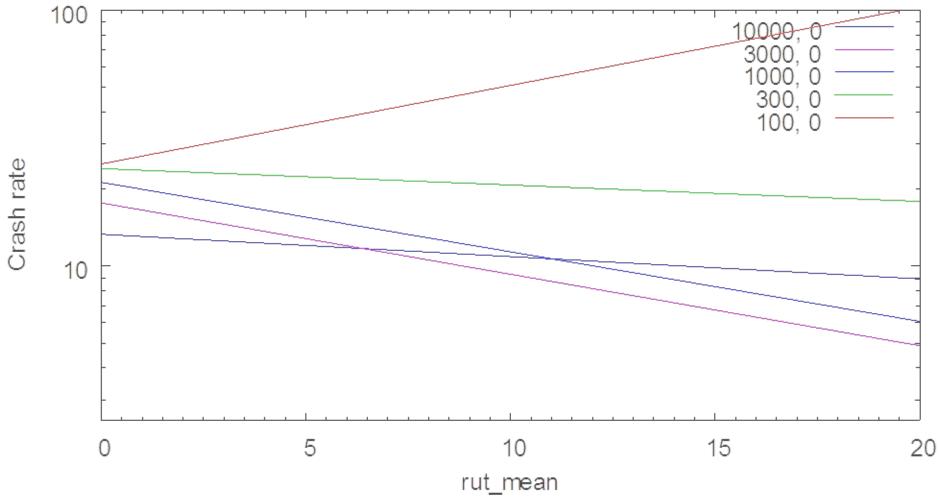
Table 5.4 Analysis of variance – 'all' casualty crashes: mean rut depth terms

Predictor variable	df	1% pt	Chi-squared	
			Type III	Type I
poly2_sqrt_bound_rut_mean	1	6.63	37.16	39.66
poly2_bound_log10_abs_curvature × bound_rut_mean	2	9.21	58.43	58.43

With reference to table 5.4, both mean rut depth terms appear to be statistically significant.

Figure 5.4 shows plots of the 'all' casualty crash rate versus mean rut depth for various horizontal curvatures. These plots don't make much sense, highlighting the need for additional investigation to determine what exactly is going on in terms of the interaction between mean rut depth and road curvature.

Figure 5.4 Effect of mean rut depth on the 'all' casualty crash rate as a function of curve radius



5.4 Interaction between year and region

The analyses detailed in section 3 are robust against variations in crash rate due to changes in reporting rates, policing, changes in traffic volume, and weather, provided they were the same across the whole network. Likewise, variations from region to region would not be a problem provided they remained the same from year to year. However, if the year-to-year changes varied from region to region, then there could be an issue. Year-to-year changes that vary from region to region can be partially compensated by including a year-×-region interaction term in the analysis. This substantially increases the computer time required, but is feasible.

This analysis has been run for 'all' crashes and 'wet' crashes. The estimates are almost identical to the corresponding ones in section 3.5. In support of this statement, the analysis of variance for 'all' crashes is given in table 5.5, and this should be compared with table 3.3. As can be seen, apart from the year and region, the chi-squared entries are almost the same as in table 3.3.

Table 5.5 Analysis of variance, year-×-region term included – 'all' casualty crashes

Predictor variable	df	1% pt	Chi-squared	
			Type III	Type I
year	9	21.70	96.85	521.06
region	13	27.70	44.89	675.18
urban_rural	1	6.63	27.95	484.56
adj_skid_site	2	9.21	4705.60	6291.90
poly3_bound_OOCC	3	11.30	452.66	5402.20
poly2_bound_log10_abs_curvature	2	9.21	110.14	460.85
poly2_log10_ADT	2	9.21	579.81	517.45
poly2_scrim-0.5000	2	9.21	225.50	266.24
poly3_bound_abs_gradient	3	11.30	45.92	65.85
poly3_bound_adj_log10_iri	3	11.30	81.86	108.67
poly2_bound_log10_abs_curvature × poly2_bound_adj_log10_iri	4	13.30	105.56	107.61
year × region	117	155.50	274.61	274.61

The results of a fit test similar to that detailed in appendix section B.1 are given in table 5.6.

Table 5.6 Sum of squares of normalised residuals for fits to 'all' and 'wet' casualty crashes

	'All' crashes	'Wet' crashes
Number of crashes	22870	6476
Chi-squared value	675	624
Chi-squared value with inclusion of year×region term	672	619

With reference to table 5.6, the results are almost unchanged with the inclusion of the year-×-region term. The data indicates that changing patterns of reporting rates, policing, traffic volume and weather do not seem to be a problem. There could, of course, be more subtle changes that these analyses would not detect (refer to appendix B).

5.5 Serious and fatal crashes only

The analyses described in section 3 of this report have used all reported crashes that have involved at least one fatality, serious injury or minor injury. Minor-injury crashes have a low reporting rate and this may be variable over time and over the country. With 10 years' data it is possible to carry out some of the analyses with only fatal and serious-injury crash data, to be consistent with the focus of the New Zealand government's 'Safer Journeys' initiative, where the vision is for *'a safe road system increasingly free of death and serious injury'*. Crash numbers are given in appendix section A.3.4.

In this section, the analysis is repeated for 'all' crashes and 'wet' crashes, using only the serious and fatal crash data.

5.5.1 'All' fatal and serious-injury crashes

The resulting variance table and predictor graphs for SCRIM skid resistance and lane IRI roughness for 'all' fatal and serious-injury crashes are given in table 5.7 and figures 5.5 and 5.6 respectively.

With reference to table 5.6, the gradient and IRI terms are now barely statistically significant.

Table 5.7 Analysis of variance – 'all' fatal and serious-injury crashes

Predictor variable	df	1% pt	Chi-squared	
			Type III	Type I
year	9	21.70	154.58	127.38
region	13	27.70	69.34	196.97
urban_rural	1	6.63	64.97	1.3263
adj_skid_site	2	9.21	1167.70	1379.80
poly3_bound_OOCC	3	11.30	147.83	1835.60
poly2_bound_log10_abs_curvature	2	9.21	22.54	153.84
poly2_log10_ADT	2	9.21	275.36	250.43
poly2_scrim-0.5000	2	9.21	61.67	67.46
poly3_bound_abs_gradient	3	11.30	9.10	11.75
poly3_bound_adj_log10_iri	3	11.30	13.21	8.72
poly2_bound_log10_abs_curvature × poly2_bound_adj_log10_iri	4	13.30	21.35	21.35

Figure 5.5 'All' fatal and serious-injury casualty crash rate versus SCRIM skid resistance

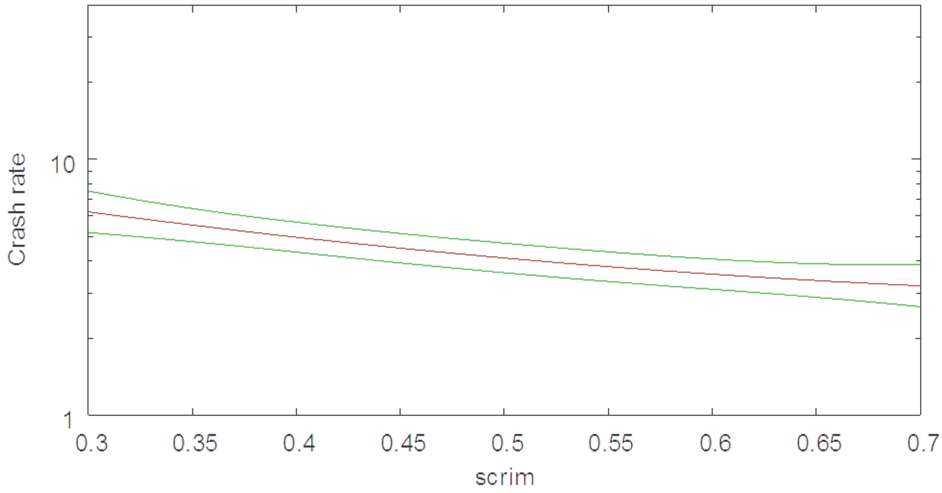
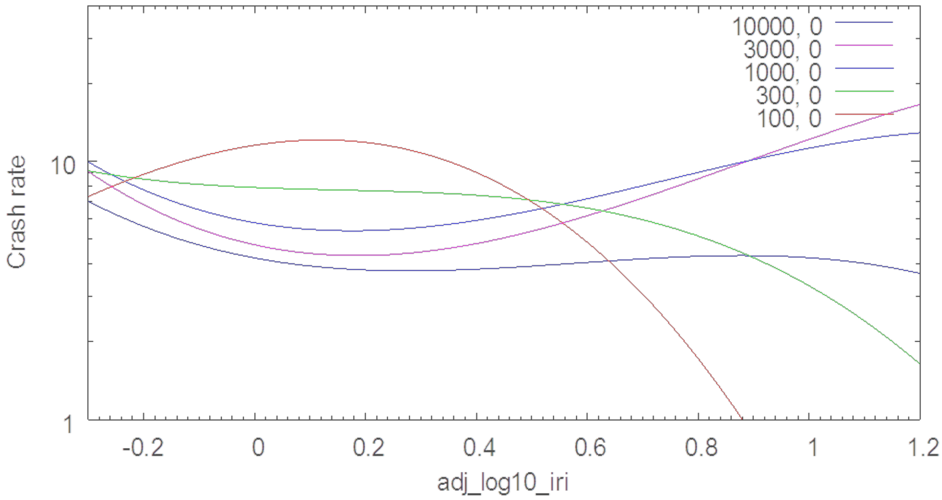


Figure 5.6 'All' fatal and serious-injury casualty crash rate versus lane roughness (*adjusted log_{10} IRI*) for selected values of curvature



The predictor graphs shown in figures 5.5 and 5.6 are remarkably similar to those for 'all' crashes in section 3.5.1 (refer to figures 3.9 and 3.13). The other predictor graphs, not shown here, are also similar.

5.5.2 'Wet' fatal and serious-injury crashes

The resulting variance table and predictor graphs for SCRIM skid resistance and lane IRI roughness for 'wet' fatal and serious-injury crashes are given in table 5.8 and figures 5.7 and 5.8 respectively.

With reference to table 5.8, the lane IRI roughness terms are now not statistically significant.

Table 5.8 Analysis of variance – 'wet' fatal and serious-injury crashes

Predictor variable	df	1% pt	Chi-squared	
			Type III	Type I
year	9	21.7	82.281	47.539
region	13	27.7	52.372	132.35
urban_rural	1	6.63	39.751	33.754
adj_skid_site	2	9.21	196.88	235.06
poly3_bound_OOCC	3	11.3	66.906	1160.8
poly2_bound_log10_abs_curvature	2	9.21	11.109	111
poly2_log10_ADT	2	9.21	43.878	26.218
poly2_scrim-0.5000	2	9.21	111.01	115.24
poly3_bound_abs_gradient	3	11.3	13.693	15.982
poly3_bound_adj_log10_iri	3	11.3	4.6432	2.0679
poly2_bound_log10_abs_curvature × poly2_bound_adj_log10_iri	4	13.3	12.69	12.69

The predictor graphs shown in figures 5.7 and 5.8 are again very similar to those for 'wet' crashes in section 3.5.1.1 (refer to figures 3.18 and 3.20).

Figure 5.7 'Wet' fatal and serious-injury casualty crash rate versus SCRIM skid resistance

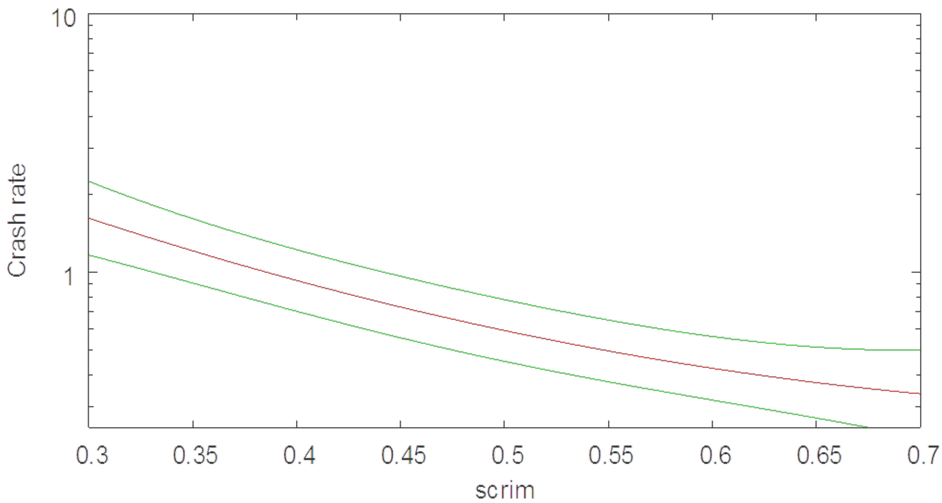
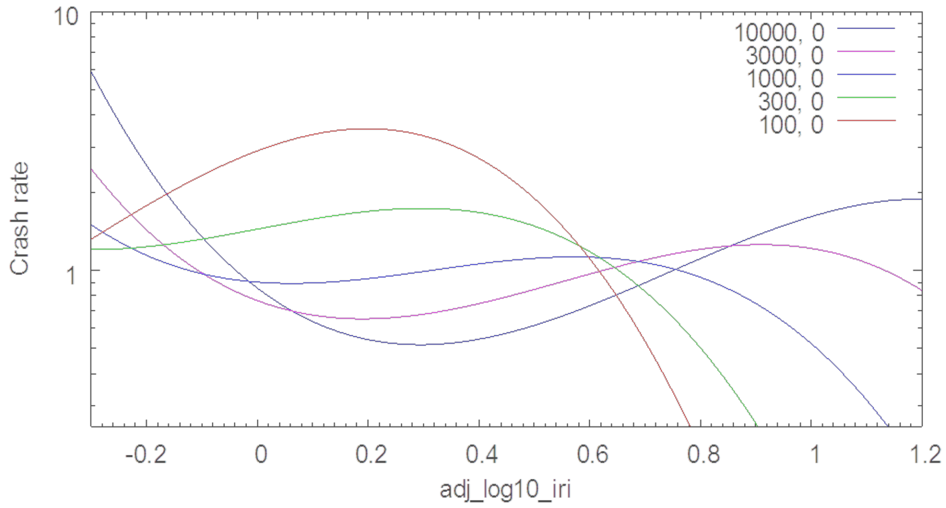


Figure 5.8 'Wet' fatal and serious-injury casualty crash rate versus lane roughness (*adjusted log_{10} IRI*) for selected values of curvature



5.5.3 Model fit

The results of the fit test to the reduced dataset of only fatal and serious-injury crashes is summarised in table 5.9. In comparison with table 5.6, the chi-squared values are now substantially reduced. The difference is likely due to the random variation of the reduced crash data partially masking whatever is causing the lack of fit. Table 5.9 suggests an adjustment factor of 2.3 for 'all' fatal and serious-injury crashes and a slightly higher value for 'wet' fatal and serious-injury crashes.

Table 5.9 Sum of squares of normalised residuals for fits to 'all' and 'wet' fatal and serious-injury crashes

	'All' fatal & serious-injury crashes	'Wet' fatal & serious-injury crashes
Number of crashes	7088	1853
Chi-squared value	287	332

6 Implications for models in present use

6.1 Background

Two crash prediction models derived from statistical modelling exercises performed on 1997–2002 state highway casualty crash and road condition and road geometry data in RAMM are currently being used by NZTA for the safety management of rural state highways. Both these models incorporate the out-of-context-curve (OOCC) effect. Details of the 1997–2002 dataset can be found in Davies et al (2005) for comparison with appendix A.

One model is used to automatically calculate the horizontal alignment road protection score from road geometry parameters stored in RAMM. The horizontal alignment road protection score is one of the inputs used for generating 'STAR RATINGS' for New Zealand's rural state highways as part of the KiwiRAP initiative (refer to www.kiwirap.org.nz). Star Rating a road is a proactive approach to road safety. It enables sections of road with a relatively high level of risk to be identified before a crash occurs. The Star Ratings make drivers aware of the relative safety of the roads they use, as well as help to identify the roads that will benefit from safety improvements. Details of the KiwiRAP crash prediction model is given in appendix F.

The other model is used to calculate the expected personal risk of rural state highway curves with a horizontal radius of curvature less than 400m, for the purposes of assigning appropriate skid resistance investigatory levels (refer to Cenek et al 2011). For ready reference, the model used to generate expected personal and collective risk ratings provided in the 'Curve-context' table in RAMM is provided in appendix G.

This section investigates whether the increased 2000–2009 dataset, described in appendix A, has resulted in any significant changes to the model forms. Particular emphasis has been placed on the 'Curve-context' model, as the current model does not include a term to account for lane roughness.

6.2 KiwiRAP Model

Comparing the results of the analysis of variance of the model fit to 'all' casualty crashes for both models (refer to tables 3.3 and F.1), it can be seen that inclusion of the interaction term *curvature and adjusted IRI* had no impact on the original model form. Therefore, the results from the statistical modelling utilising the reduced 1997–2002 dataset remain valid, and so the KiwiRAP model can continue being used to generate horizontal alignment road protection scores without the need to update to the newer model.

6.3 Curve-context model

The model detailed in appendix G was rerun with the 2000–2009 dataset, and with terms for roughness plus its interaction with advisory speed, included. Table 6.1 is the resulting analysis of the variance table, which should be compared with table G.2. Again, the results are almost the same as in the previous study, which ustilised the 1997–2002 dataset.

Table 6.1 Analysis of variance table

Predictor variable	df	1% pt	Chi-squared	
			Type III	Type I
year	9	21.7	151.33	150.66
region	13	27.7	140.91	201.76
poly3_OOCC-30.0000	3	11.3	337.62	1127.1
poly3_AS-50.0000	3	11.3	4.6604	9.327
poly2_scrim-0.5000	2	9.21	107.78	75.253
poly3_log10_ADT-3.0000	3	11.3	129.19	122.62
poly2_gradient_app	2	9.21	17.856	19.841
poly3_adj_log10_iri	3	11.3	37.886	32.753
poly2_adj_log10_iri × poly2_AS-50.0000	4	13.3	39.654	39.654
poly2_sqrt_lengthR-15.0000	2	9.21	36.573	36.573

Figures 6.1–6.7 show the relationship between the predictor variable and the crash rate, which in this case is in terms of expected casualty crashes per year per 100 million vehicles entering the curve.

With reference to figures 6.1–6.7, the vertical scale is linear rather than logarithmic.

Figure 6.1 'All' casualty crash rate versus ADT

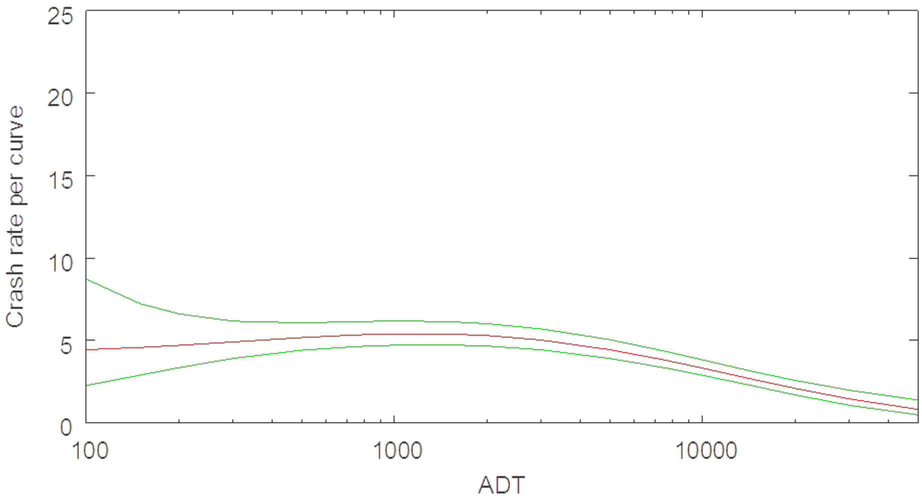


Figure 6.2 'All' casualty crash rate versus skid resistance

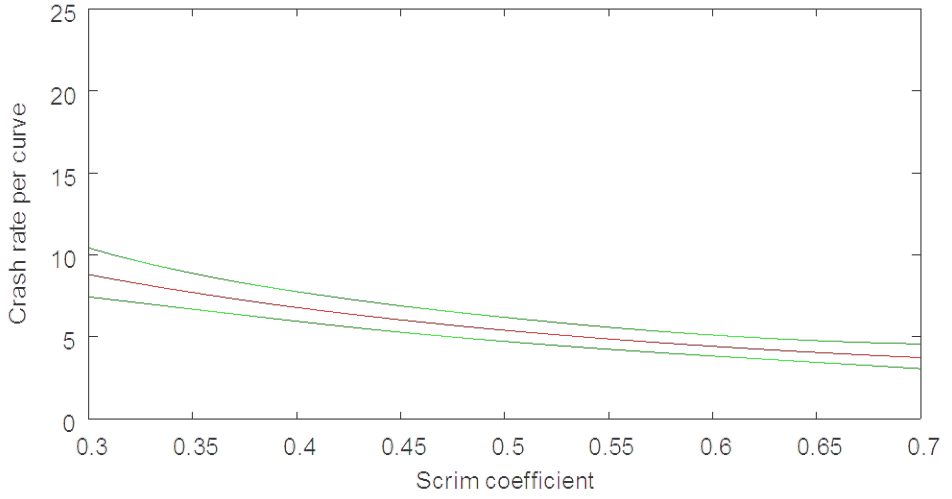


Figure 6.3 'All' casualty crash rate versus length of curve

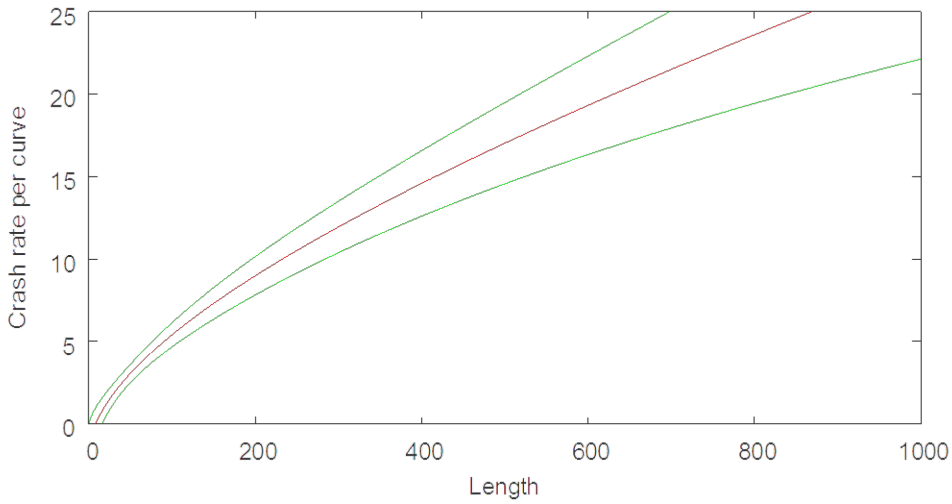


Figure 6.4 'All' casualty crash rate versus approach gradient

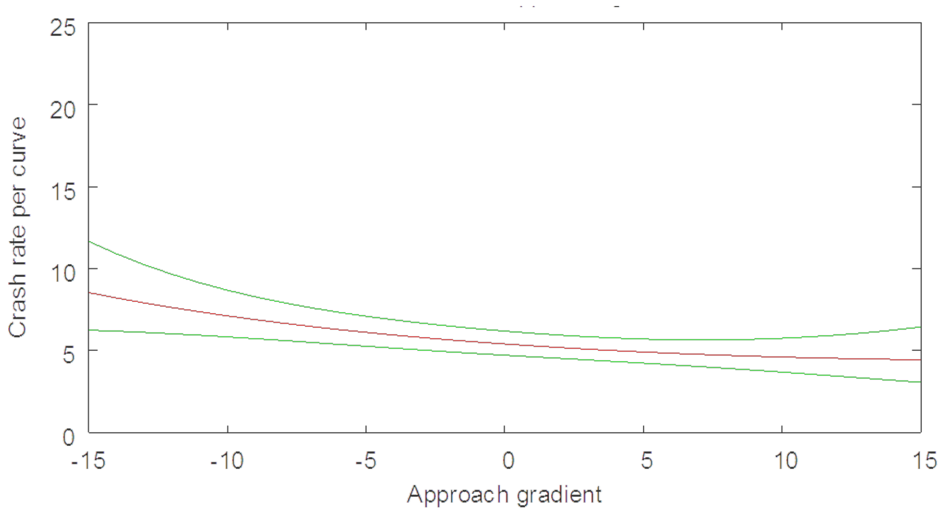


Figure 6.5 'All' casualty crash rate versus the difference between approach and curve speeds (OOCC effect)

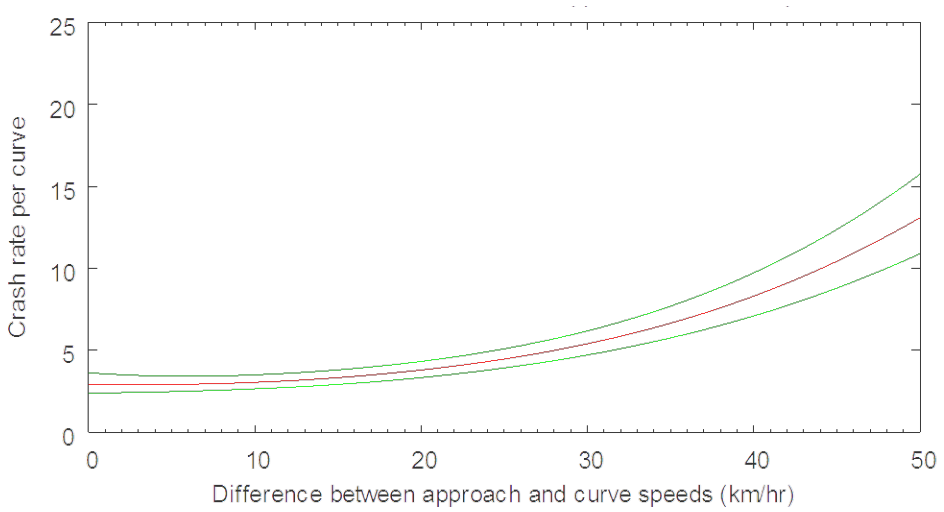


Figure 6.6 'All' casualty crash rate versus curve advisory speed

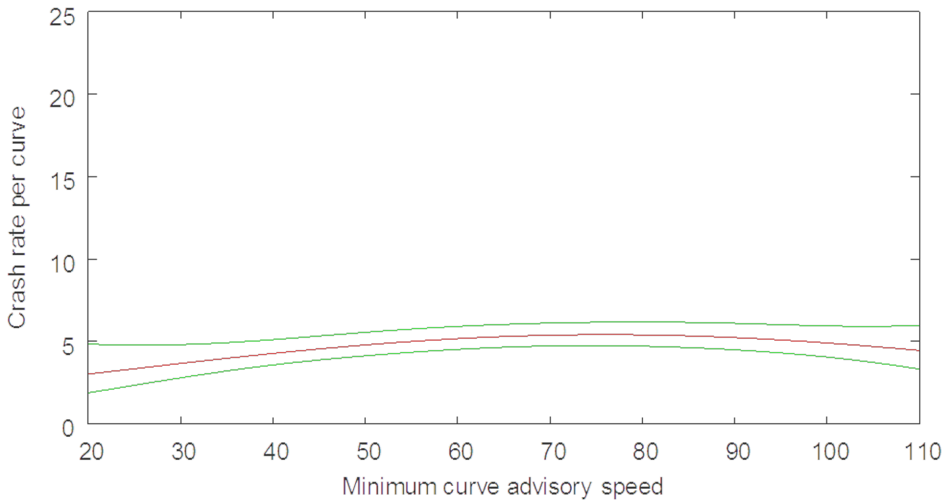
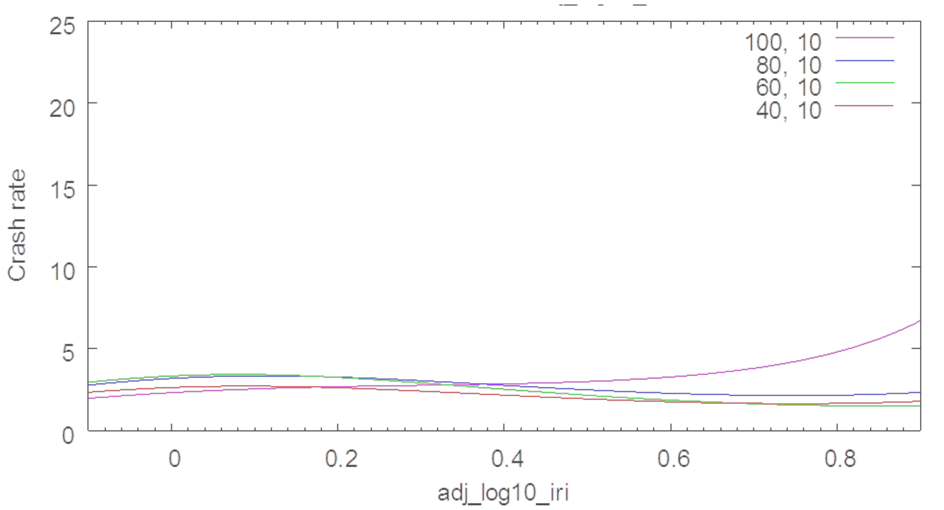


Figure 6.7 'All' casualty crash rate versus adjusted iri



With reference to figure 6.7, the ‘all’ casualty crash rate is graphed against the *adjusted IRI lane roughness* for curve advisory speeds of 40, 60, 80 and 100km/h and OOCC=10.

It appears that roughness is important only for the curves with high minimum-advisory speeds, though it is somewhat marginal whether the IRI should have been included. This result is in line with the findings throughout this report.

Therefore, the results from the statistical modelling utilising the reduced 1997–2002 dataset remain valid and so the model detailed in appendix G can continue being used for assigning skid resistance investigatory levels to curves with horizontal radius of curvature 400m or less.

7 Concluding remarks

The most striking result of this research has been the close agreement with previous statistical modelling using the 1997–2002 data described in Cenek and Davies (2006).

The inclusion of the term for interaction between roughness and curvature suggests that roughness is a factor for curves where traffic is going at close to full speed but there still is some curvature.

There is a suggestion that skid resistance is more important on curves than on straight roads. This makes sense but as yet, the dependence has not been precisely quantified. The agreement between the analyses when we look at all casualty crashes and those when we consider only serious/fatal crashes suggests that the low reporting rates associated with minor-injury crashes is not a serious problem.

Similarly there is little change when the year-x-region interaction is included. This also suggests that reporting rates are sufficiently consistent for the analyses to be valid.

There is still more variability in the data than the Poisson model would predict. In this case, the model is unlikely to fit exactly, as there are numerous things not included and the fit might be the best that one can reasonably expect. However, it is possible that the problem lies in the estimates of ADT and this might be worth investigating further. Ideally the analysis should be improved to take account of this variability. One approach is through adding an additional 'random effects' term. Another is using a statistical technique known as *the jack-knife*[6] to find improved significance tests and confidence intervals.

The presence of the ADT term in the regression part of the model is also slightly worrying. It suggests that there are other road characteristics that are not included in the model that are present in low-ADT roads. It is easy to suggest a number of these – eg lower standard of driving, less policing, poorer signage, more roadside hazards, and so on. It is also possible that there is a bias in the measurement of ADT on low-ADT roads, but this is unlikely to be sufficient to cause the effect seen.

The analysis that has been done here is a *retrospective analysis* as opposed to a *designed experiment*. So it is not possible to be sure that the predictor variables used in the regression analysis are really the ones affecting the crash rates. It has already been suggested that this is the case with the ADT effect. Roughness, in particular, could be a surrogate for a number of characteristics associated with roads in need of repair.

However, the overall results are sufficiently encouraging to suggest that the revised model will be a useful addition to the tools already being used in the safety management of New Zealand state highways. It allows proactive identification of existing engineering-related road safety deficiencies and more importantly, the ability to better quantify the potential for improvement.

6 Jun Shao and Dongsheng Tu (1995) *The jack-knife and bootstrap*.

8 Bibliography

Affum, J and R Goudens (2008) Application of NetRISK road network safety assessment tool: the North Coast Hinterland District case study. *23rd ARRB Conference – Research Partnering with Practitioners*, 30 July–1 August 2008, Adelaide.

Bansen, J and K Passetti (2005) Application of the IHSDM: a case study. *3rd International Symposium on Highway Geometric Design*, June 2005, Chicago, US.

Bobevski, I, S Hosking, P Oxley and P Cameron (2007) Generalised linear modelling of crashes and injury severity in the context of the speed-related initiatives in Victoria during 2000–2002. *Monash University Accident Research Centre: report no.268*.

Cenek, PD, RB Davies, MW McLarin, G Griffith-Jones and NJ Locke (1997) Road environment and traffic crashes. *Transfund New Zealand research report 79.* 132pp.

Cenek, PD and RB Davies (2006) Modelling and analysis of crash densities for Karangahake Gorge, New Zealand. Australasian Road Safety Research, Policing and Education Conference, 25–27 October 2006, Surfers Paradise, Queensland. Accessed 2 June 2010. *www.robertnz.net/pdf/Karangahake.pdf*

Chen, S, A Rakotonirainy, M Sheehan, S Krishnaswamy and SW Loke (2006) Assessing crash risks on curves. Australasian Road Safety Research, Policing and Education Conference, 2006, Surfers Paradise, Queensland. Accessed 7 June 2010. *www.rsconference.com/pdf/RS060010.pdf*

Cossens, I and P Cairney (2008) Road surface characteristics and crash occurrence: a literature review. Austroads project no.AT1170. *Austroads Publication No. AP–T96/08*. 53pp.

Davies, RB, PD Cenek and RJ Henderson (2005) The effect of skid resistance and texture on crash risk. Surface Friction Conference, May 2005, Christchurch. Accessed 22 August 2010. *http://sh20mountroskill.co.nz/technical/surfacefriction/pdf/Effect-Skid-Resistance-Texture-Crash-Risk.pdf*

Easa, SM and QC You (2009) Collision prediction models for three-dimensional two-lane highways: horizontal curves. *Transportation Research Record 2092*: 48–56.

Elvik, R (2008) Comparative analysis techniques for identifying locations of hazardous roads. *Transportation Research Record 2083*: 72–75.

Federal Highway Administration (FHWA) (2006) Interactive highway safety design model (IHSDM). Website Home Page & Overview. Accessed 7 June 2010. *www.tfhrc.gov/safety/ihsdm/ihsdm.htm*

Harper, NJ and RCM Dunn (2005) Accident prediction models at roundabouts. *Institute of Transportation Engineers (ITE) Annual Meeting*, August 2005, Melbourne.

Hildebrand, E, K Robichaud and H Ye (2008) Evaluation of accident prediction for rural highways. *Canadian Journal of Civil Engineering 35*: 647–651.

Koorey, GF and FN Tate (1997) Review of accident analysis procedures for project evaluation manual. *Transfund research report 85*: 54pp.

Koorey, GF (2006) Application of IHSDM highway safety modelling to New Zealand. Australasian Road Safety Research, Policing and Education Conference, 25–27 Oct 2006, Surfers Paradise, Australia.12pp. Accessed 2 June 2010. *http://ir.canterbury.ac.nz/bitstream/10092/319/1/12604186_RSPE-Paper-Revised-GKoorey.pdf*

Koorey, G (2010) Calibration of overseas highway crash prediction models for New Zealand – a case study with IHSDM. IPENZ Transportation Group Conference, March 2010, Christchurch. Accessed 2 June 2010. *http://ir.canterbury.ac.nz/bitstream/10092/3896/1/12624225_IPENZTG10-Koorey.pdf*

Liu, C and R Subramanian (2009) Factors related to fatal single-vehicle run-off-road crashes. *National Highway Traffic Safety Administration (NHTSA) Technical Report DOT HS 811 232*. 30pp.

McInerney, R, J Harkness, J Affum, K Armstrong, J Metcalfe and G Smith (2008) Comparing risk maps and star ratings. Australian Automobile Association. Technical Working Paper, 2008-04. 28pp. Canberra ACT. *www.aaa.asn.au/documents/reports%2F2008%2FComparing_Risk_Maps_and_Star_Ratings_FINAL_%28web%29.pdf Viewed 02/06/2010*

Montella, A (2010) A comparative analysis of hotspot identification methods. *Accident Analysis and Prevention 42*: 571–581.

Montella, A, L Colantuoni and R Lamberti (2008) Crash prediction models for rural motorways. *Transportation Research Record 2083*: 180–189.

Ministry of Transport (MOT) (2010) Motor vehicle crashes in New Zealand 2009. Accessed 12 March 2012. *www.transport.govt.nz/research/Motor-Vehicle-Crashes-in-New-Zealand-2009/*

NZ Transport Agency (NZTA) (2010) Economic evaluation manual (vol 1). 1st edition, amendment 0. Effective from January 2010. Accessed 17 June 2010. *www.nzta.govt.nz/resources/economic-evaluation-manual/volume-1/index.html*

Pigman, JG and KR Agent (2007) *NCHRP SYNTHESIS 369: State DOT crash reconstruction practices: a synthesis of highway practice*. Transportation Research Board National Cooperative Highway Research Program. Washington DC: Federal Highway Administration.

Prinsloo, B and D Chee (2005) Development of a crash prediction model for rural roads in NSW, Australia. *Institute of Transportation Engineers (ITE) Annual Meeting*, August 2005, Melbourne.

Prinsloo, B and C Goudanas (2003) Development of a crash prediction model for rural roads in NSW. *21st ARRB Transport Research Conference*, 18–23 May 2003, Queensland.

Roozenburg, A and S Turner (2005) Accident prediction models for signalized intersections. IPENZ Transportation Group Conference, 7 September 2005, Auckland. Accessed 2 June 2010. *www.ipenz.org.nz/ipenztg/ipenztg_cd/cd/2005_pdf/03_Roozenburg.pdf*

Tate, FN and SA Turner (2007) Road geometry and drivers' speed choice. *Road and Transport Research 4*, vol 16.

Transfund NZ (1997) *Project evaluation manual September 2002: revision*. Wellington: Transfund NZ.

Transit NZ (2002) TNZ T10:2002 Specification for skid resistance investigation and treatment selection. Accessed 27 February 2012. *www.nzta.govt.nz/resources/skid-resistance-investigation-treatment-selection/*

Turner, B (2007) Research to improve the accuracy of economic evaluations in road safety. *Australasian Road Safety Research, Policing and Education Conference*, 17–19 October 2007, Melbourne.

Turner, B (2008a) Review of best practice in road crash database and analysis system design. *Australasian Road Safety Research, Policing and Education Conference,* 9–12 November 2008, Adelaide.

Turner, B (2008b) Road safety engineering risk assessment: findings from a program of research. *23rd ARRB Conference,* 30 July–1 August, 2008, Adelaide.

Turner, B and D Bennett (2005) Road safety engineering risk assessment: update on Austroads research program. *Australasian Road Safety Research, Policing and Education Conference*, 14–16 November 2005, Wellington.

Turner, B and L Comport (2007) Automatic collection of safety related road and roadside data. *Australasian Road Safety Research, Policing and Education Conference,* 17–19 October 2007, Melbourne.

Turner, B and K Imberger (2005) Investigation of risk reduction for various safety treatments in different environments. *Australasian Road Safety Research, Policing and Education Conference,* 14–16 November 2005, Wellington.

Turner, B and C Jurewicz (2008): Austroads road safety engineering risk assessment program. Australian Institute of Traffic Planning and Management. Accessed 16 June 2010. *www.aitpm.org.au/d3/AITPM_Vic_TechForum_20090325(RoadSafetyRiskAssesment).pdf*

Turner, B and M Tziotis (2006) Road safety engineering risk assessment: recent and future research. *Australasian Road Safety Research, Policing and Education Conference,* 25–27 October 2006, Surfers Paradise, Queensland.

Turner, B and M Tziotis (2008) Road safety engineering risk assessment: recent and future research. *Journal of the Australasian College of Road Safety*: 29–35.

Turner, B, M Tziotis, P Cairney and C Jurewicz (2009) Safe system infrastructure: national roundtable report. *Research report ARR 370*: 35pp.

Turner, S (date unknown): Road safety trends, targets and safety programs in New Zealand. Accessed 2 June 2010. *www.hsrc.unc.edu/directions/2007/winter/RoadSafetyTrendsinNewZealandS.Turner.pdf*

Turner, S and G Wood (2009a) Crash prediction modelling down-under: some key findings. TRB 88th Annual Meeting Compendium of Papers DVD. Accessed 2 June 2010. *http://pubsindex.trb.org/view.aspx?id=881349*

Turner, S and G Wood (2009b) Crash prediction modelling at intersections in New Zealand: 1990 to 2009. Australian Transport Research Forum, 29 September–1 October 2009, Auckland. Accessed 2 June 2010. *www.patrec.org/web_docs/atrf/papers/2009/1793_paper131-Turner.pdf*

Turner, S and G Wood (2009c) Crash prediction modelling down-under: a literature review. *TRB 2009 Annual Meeting CD-ROM.*

Turner, S, A Dixon and G Wood (2004) Assessing the crash risk implications of roadside hazards. IPENZ Transportation Conference, 8 September, 2004, Wellington. Accessed 2 June 2010. *www.ipenz.org.nz/ipenztg/ipenztg_cd/cd/2004_pdf/13_Turner_Dixon_Wood.pdf*

Turner, S, B Persaud and M Chou (2007) Transferability of overseas crash prediction models to New Zealand. IPENZ Transportation Group Conference, 7–10 October 2007, Tauranga. Accessed 2 June 2010. *www.ipenz.org.nz/ipenztg/papers/2007_pdf/A3a_TurnerPersaudChou.pdf*

Turner, SA, AP Roozenburg and T Francis (2006) Predicting accident rates for cyclists and pedestrians. *NZ Transport Agency research report 289*. 180pp.

Turner, S, AP Roozenburg and AW Smith (2009) Roundabout crash prediction models. *NZ Transport Agency research report 386.* 102pp.

Turner, S, F Tate and G Koorey (2007) A 'sidra' for road safety. IPENZ Transportation Group Conference, 7–10 October 2007, Tauranga. Accessed 2 June 2010. *http://ir.canterbury.ac.nz/bitstream/10092/692/1/12606057_A3b_TurnerTateKoorey.pdf*

Turner, S, B Turner and G Wood (2008) Accident prediction models for traffic signals. *23rd ARRB Conference – Research Partnering with Practitioners*, 30 July–1 August 2008, Adelaide, Australia.

Turner, SA, G Wood and A Roozenburg (2006) Accident prediction models for high speed intersections (both rural and urban). *22nd ARRB Conference*, 22–26 October 2006, Canberra, Australia.

Turner, S, P Durdin, I Bone and M Jackett (2003) New Zealand accident prediction models and their applications. *21st ARRB/REAAA Conference,*18–23 May 2003, Cairns.

Turner, S, BN Persaud, M Chou, C Lyon and A Roozenburg (2007) International crash experience comparisons using prediction models. Transportation Research Board 86th Annual Meeting, 21–25 January 2007, Washington DC. Accessed 2 June 2010. *http://pubsindex.trb.org/view.aspx?id=802593*

Appendix A More detailed description of the data and its processing

A.1 Initial processing

Initial processing was carried out using the MySQL database program.

Each line in each RAMM table has a unique identity value and the following variables: *survey_number*; *road_id*; *start_m*; *end_m*; and *lane,* which identify the segment of road being surveyed. For the geometry, SCRIM skid resistance and texture tables, the procedure is to have *end_m – start_m = 10*, ensuring *start_m* is a multiple of 10. Initial processing consisted of using the survey-number-to-year correspondence table to establish the *survey_year* of each survey, rounding *start_m* and *end_m* to be multiples of 10 and rejecting lines where the rounded values didn't differ by exactly 10. Where there were duplicate measurements, the one with the highest identity value was chosen.

For the roughness and rutting tables, the procedure is to have *end_m – start_m = 20*, ensuring *start_m* is a multiple of 20. Initial processing consisted of using the survey-number-to-year correspondence table to establish the *survey_year* of each survey, rounding *start_m* and *end_m* to be multiples of 10 and rejecting lines where the rounded values didn't differ by exactly either 10 or 20, or where *end_m – start_m* < 6. Where there were duplicate measurements, the one with the highest identity value was chosen.

The subsequent processing was as follows:

- Create a base set of 10m road segments. This consists of all values of *survey_year*, *road_id* and the rounded version of *start_m* that appear in any of the 10m SCRIM+ variables. This is joined with the road names table. A sort can be carried out using the *road_name* variable to ensure that the sections of road identified by each *road_id* are in consecutive order along each state highway.
- The 10m data is joined to this base set with separate columns for the left and right lanes – only the data with lane = *L1* or R1is chosen. Locations of the beginning and end of each *road* as identified by the *road_id* are found.
- Now considering the roughness and rutting data, table columns are made showing where, for each 10m segment in the base set, the corresponding roughness and rutting data values occur in the roughness and rutting tables. Where there are 20m roughness or rutting segments, each value will typically be referenced twice in these columns. Where there are both 10m and 20m roughness or rutting segments that could correspond to a 10m segment in the base set, the 10m one is chosen preferentially. The actual combining of the roughness and rutting data into the rest of the data is done in the subsequent analysis carried out by the C++ programs.

Vehicle crash data and causes for each crash are each amalgamated into single fields and joined in to the crash table. The linking of the crashes to the 10m data is carried out by the C++ programs.

A.2 Final assembly of the data

The C++ programs set up the data structures needed for carrying out the analyses described in this report, read in the data generated by MySQL, carried out some checking, and generated the transformed data where required.

In particular, they linked in the roughness, rutting, crash and road data, and calculated the adjusted skid site, adjusted IRI, and the OOCC variables. They checked for isolated missing values in the predictor variables and attempted to estimate these from neighbouring variables.

As noted in the previous section, the state highways are divided into lengths of roads identified by their *road_id*s. It is important that any gaps between these lengths of roads are identified. The GPS data collected in the 2009–2010 as part of the geometry data was used to identify these gaps.

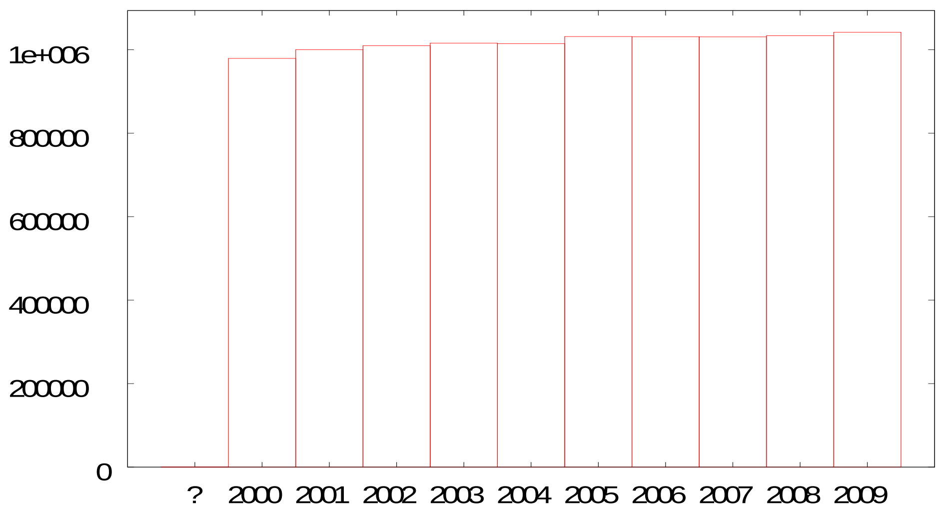
A.3 The data and calculated values used in the analysis

The following sections show histograms of the 10m data. The y-axis in each plot represents the number of 10m state highway sections under each class interval of the parameter being plotted.

A.3.1 Year

Figure A.1 shows the number of 10m segments surveyed each year – around 1 million, and increasing slightly from 2000 to 2009. This could indicate an increase in the length of road surveyed or a reduction in the number of missing values.

Figure A.1 Histogram of analysis year



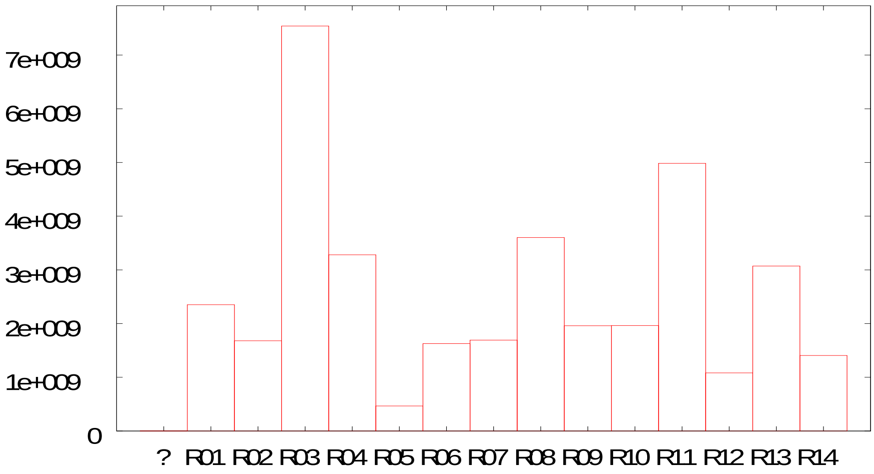
A.3.2 Region

The state highway network is divided into 14 road regions. In the 2010 survey, GPS locations of the roads were reported, enabling an accurate location plot of the of the road regions.

Figure A.2 shows the number of 10m segments in each region. Since we were considering 10 years of data (ie 2000–2009), most segments were counted 10 times. This histogram plot was weighted by the estimated ADT.

With reference to figure A.2, there is quite a lot of variability in the size of the road regions when weighted by ADT. Nevertheless, this provides a convenient way of breaking the total state highway network up into smaller subsets and the statistical modelling was set up to cope adequately with their varied sizes.

Figure A.2 Histogram of NZTA region weighted by estimated traffic (ADT)



A.3.3 Urban/rural

The roads are classified as *urban* or *rural*. *Rural* means a speed limit of more than 70km/h, which is a very rough classification. For this report *urban* means those parts of the state highway network where the speed limit is less than or equal to 70km/h. Therefore, the results, in general, do not have a lot of relevance to urban roads as a very small proportion of urban roads are state highways. Most of the state highway network is *rural*.

A.3.4 The crash data

The analyses were applied to each of four subsets of the crash dataset described in table A.1 below and the MVMT_IDA codes are described in table A.2.

Table A.1 Criteria for selecting subsets of crash data

Group	Criteria
All	All casualty crashes
Wet	All casualty crashes with the road wet field being W or the cause code was 801, 823 or 901.
Selected	All casualty crashes with MVMT_IDA being one of A, B, C, D, F
Wet & selected	Satisfying both the wet and selected criteria

Table A.2 Crash movement codes

A	Overtaking and lane change
B	Head on
C	Lost control or off road (straight roads)
D	Cornering
E	Collision with obstruction
F	Rear end
G	Turning versus same direction
H	Crossing (no turns)
J	Crossing (vehicle turning)

K	Merging
L	Right turn against
M	Manoeuvring
N	Pedestrians crossing road
P	Pedestrians other
Q	Miscellaneous

The definitions of the relevant cause codes for the ‘wet’ crashes are summarised in table A.3.

Table A.3 Relevant crash cause codes

Cause code	Contributing factor
801	Rain
823	Flood waters, large puddles, ford
901	Heavy rain

Table A.4 describes the numbers of crashes, by year, for the crash subcategories used in the analyses. The numbers are low for 2009 because complete data was not available for that year at the time of the analysis.

Table A.4 The number of crashes in each crash subcategory modelled

Year	Crash subcategory			
	‘All’	‘Wet’	‘Selected’	‘Wet selected’
2000	1800	525	1307	421
2001	2023	642	1425	506
2002	2274	654	1607	505
2003	2391	641	1726	518
2004	2379	740	1699	596
2005	2463	690	1809	569
2006	2619	728	1922	587
2007	2801	771	2079	631
2008	2554	658	1816	525
2009	1566	427	1126	349
Total	**22870**	**6476**	**16516**	**5207**

Most of the crashes were assigned locations, and the number of crashes that could not be located was small enough to not be an issue.

The number of crashes resulting in fatalities or serious injuries, for ‘all’ and ‘wet’ crash subcategories are presented in table A.5.

Table A.5 Fatal or serious crash numbers by year

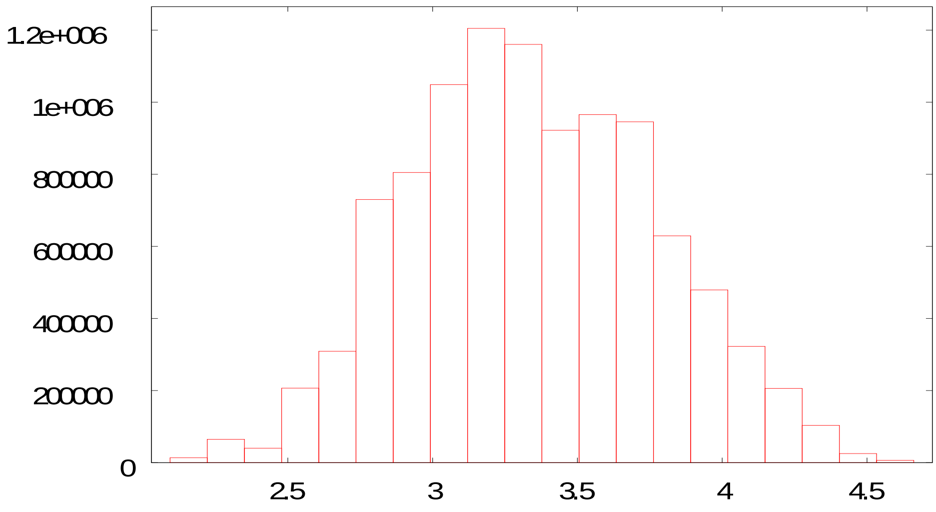
Year	Crash subcategory	
	‘All’	‘Wet’
2000	729	201
2001	731	209
2002	742	202
2003	755	177
2004	735	219
2005	751	201
2006	760	197
2007	759	182
2008	691	154
2009	435	111
Total	**7088**	**1853**

A.3.5 Estimated average daily traffic (ADT)

The ADT estimate for 2009 was used in the analysis. This would not be a problem for the analyses provided that the traffic had changed by the same relative amount for each road over the 10 years of the study. This is unlikely to be exactly true and ideally we should be using the estimate for each year. The analysis omitted roads where the ADT was less than 100.

Figure A.3 shows that a normal distribution results when ADT is transformed using $\log_{10}$.

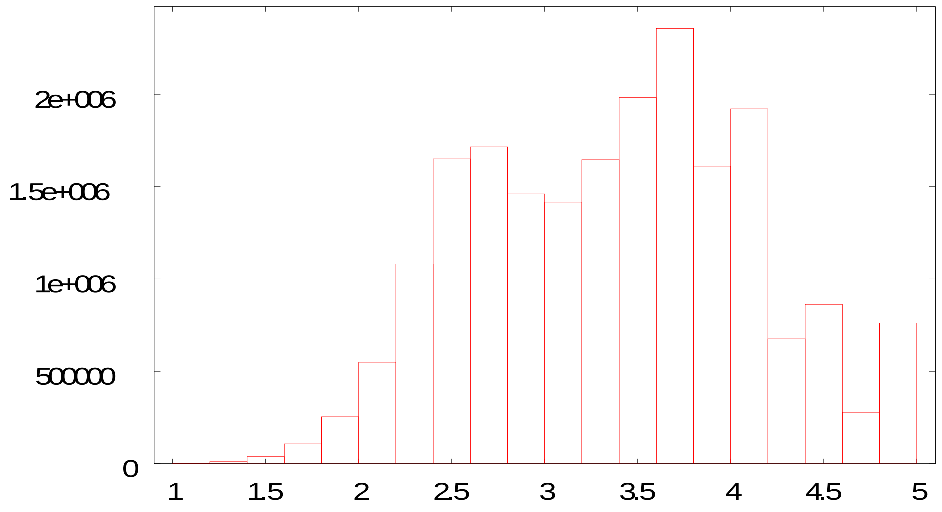
Figure A.3 Histogram of $\log_{10}$ ADT



A.3.6 Curvature

Figure A.4 is a histogram of *$\log_{10}$ absolute radius of curvature*. Straighter roads are to the right-hand side of the plot, sharp bends to the left-hand side. By convention, straight roads are assigned a radius of curvature of 100,000m corresponding to a *$\log_{10}$ absolute radius of curvature value* of 5.

Figure A.4 Histogram of $\log_{10}$ absolute radius of curvature



Figures A.5 and A.6 show the histogram of *log_{10} absolute curvature* weighted by *traffic adt est* and by the number of crashes.

Figure A.5 Histogram of log absolute curvature weighted by traffic

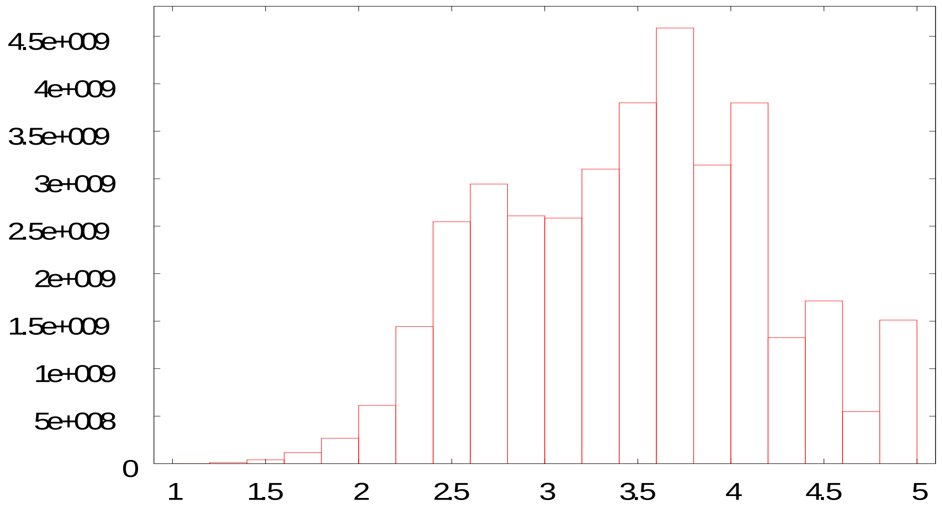
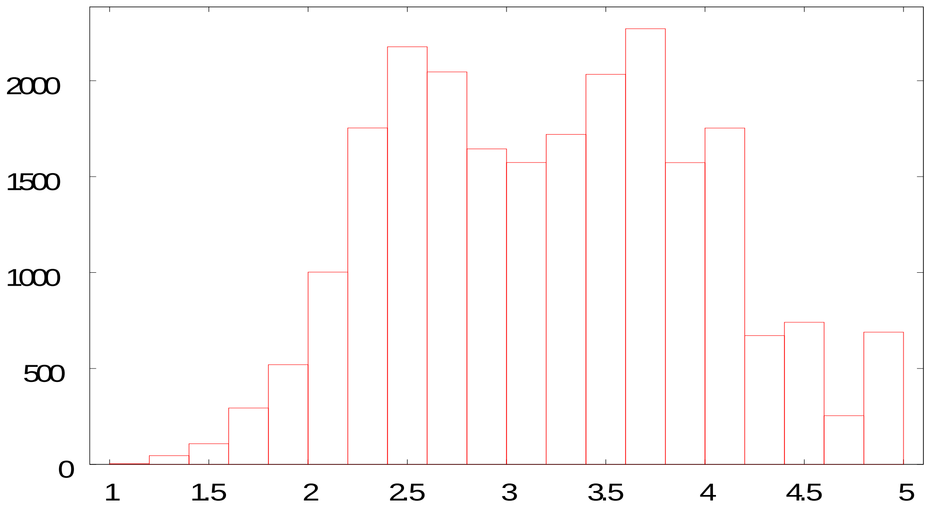


Figure A.6 Histogram of log absolute curvature weighted by crashes

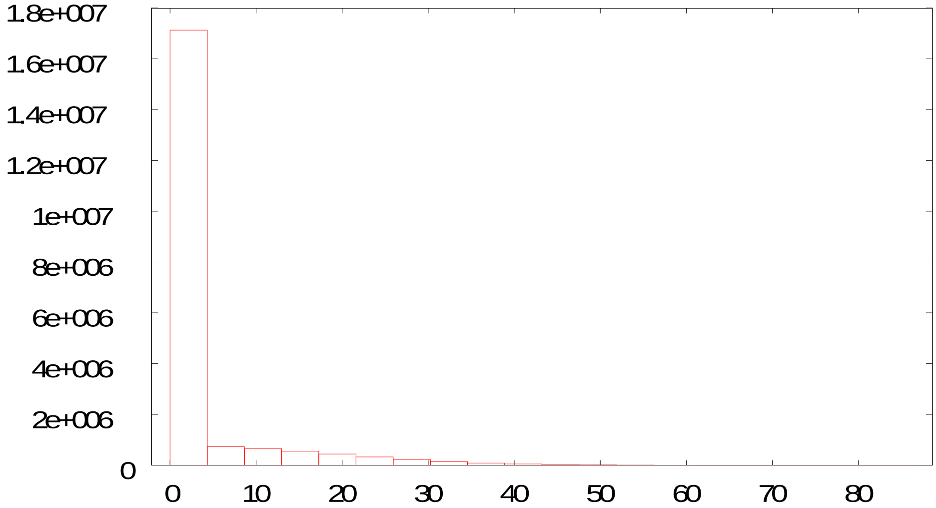


Weighting by *estimated adt* didn't make a lot of difference to the general height distribution of the histogram. However, weighting by the number of crashes showed increased heights in the histogram corresponding to low radii of curvature.

A.3.7 Out-of-context-curve indicator (OOCC)

Figure A.7 is a histogram plot of the OOCC indicator, where OOCC is defined in appendix C. Figure A.7 shows that very few 10m state highway segments have an OOCC value greater than 10km/h.

Figure A.7 Histogram of OOCC



A.3.8 SCRIM skid resistance

Figure A.8 is a histogram plot of SCRIM skid resistance. The distribution appears normal, centred around a value of about 0.52.

Figure A.8 Histogram of SCRIM skid resistance

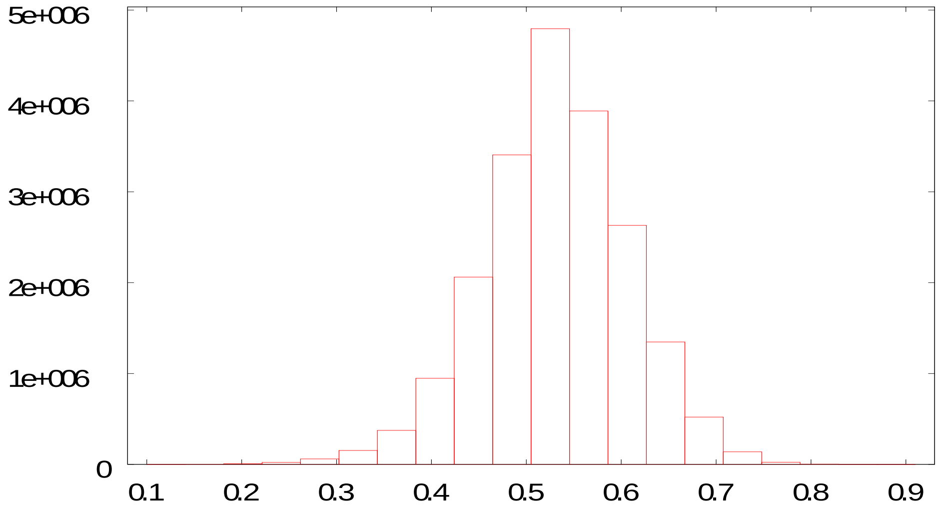
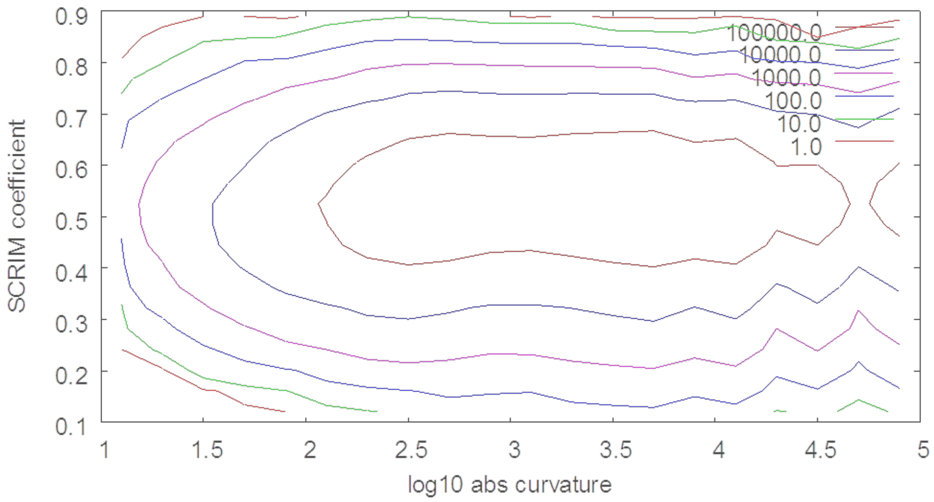


Figure A.9 is a two-dimensional histogram of *SCRIM* and *log*$_{10}$*(absolute curvature)*. This shows that there is only a slight, if any, relationship between *curvature* and *SCRIM*.

Figure A.9 Histogram of log curvature and SCRIM skid resistance



A.3.9 IRI (roughness)

IRI is the measure of lane roughness. Figure A.10 is a histogram of values of *log*$_{10}$*(IRI)*. Figure A.11 is a two-dimensional histogram of *log*$_{10}$*(IRI)* and *log*$_{10}$*(absolute curvature)*. Figure A.9 shows a slight curling up on the left-hand side of the plot, indicating that highly curved roads tend to have higher roughness. This is, at least partly, a measurement effect. It is considered advisable to try and reduce this effect and this is detailed in section A.3.10 below.

Figure A.10 Histogram of log_{10}(IRI)

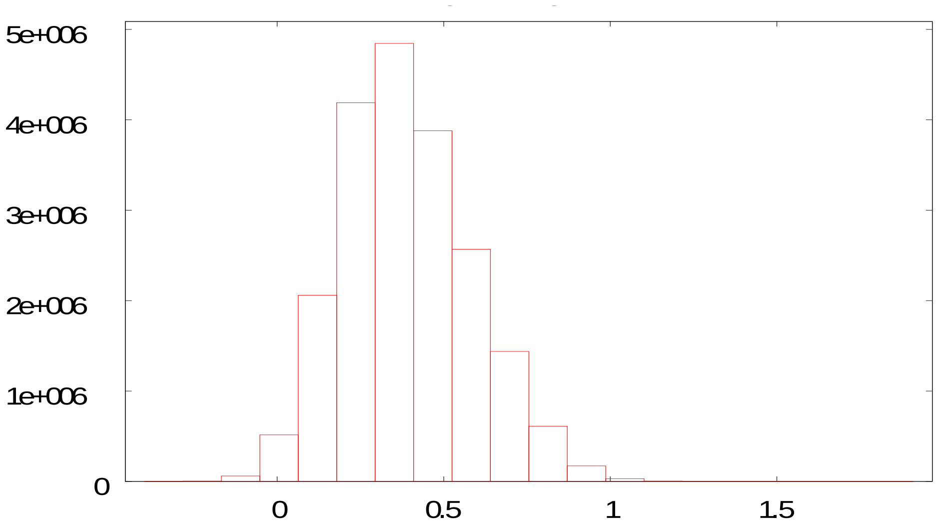
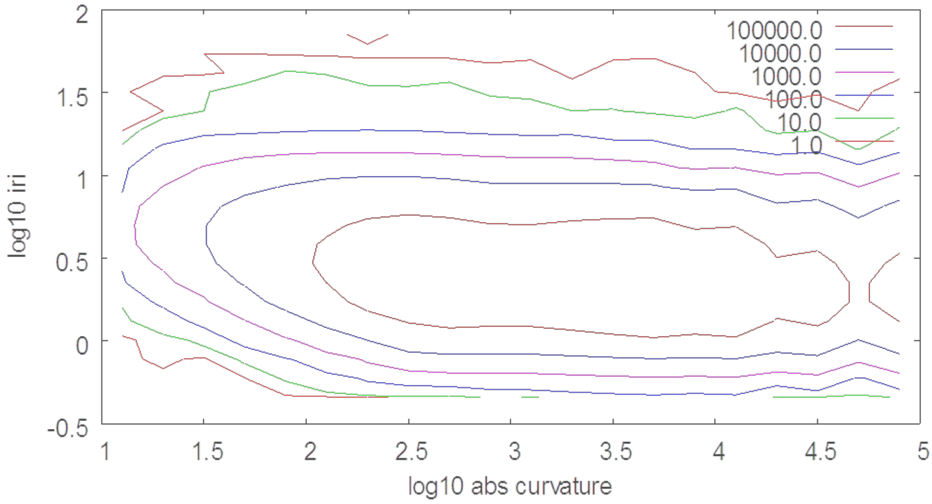


Figure A.11 Histogram of log_{10}(IRI) and log_{10}(absolute curvature)



A.3.10 Adjusted IRI

A regression analysis was carried out predicting *log_{10}(IRI)* with a fifth-degree polynomial of *log_{10}(absolute curvature)* and second-degree polynomial of *gradient*. The predicted values are shown in Figures A.12 and A.13 respectively.

The regression can be used to adjust the *IRI* value to remove the effect of *curvature* and *gradient*. The adjustment has been set so that there is no adjustment for *curvature* = 10,000 and *gradient* = 0. The adjustment reduces the *IRI* by a factor of about 2 when the *curvature* is 10 and the *gradient* 0, or by a factor of about 1.2 when the *curvature* is 10,000 and the *gradient* is 10.

The regression equations and an example calculation are provided in appendix D.

Figure A.12 IRI lane roughness versus curvature

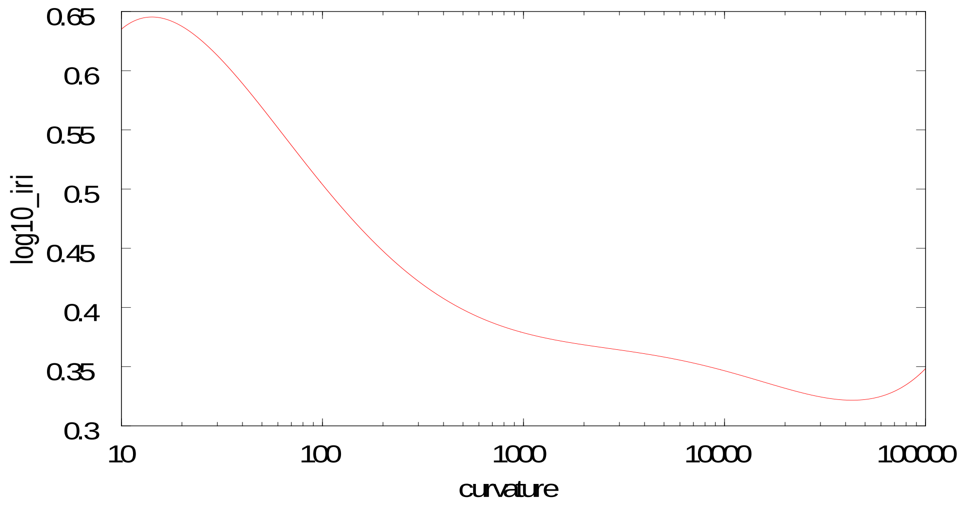
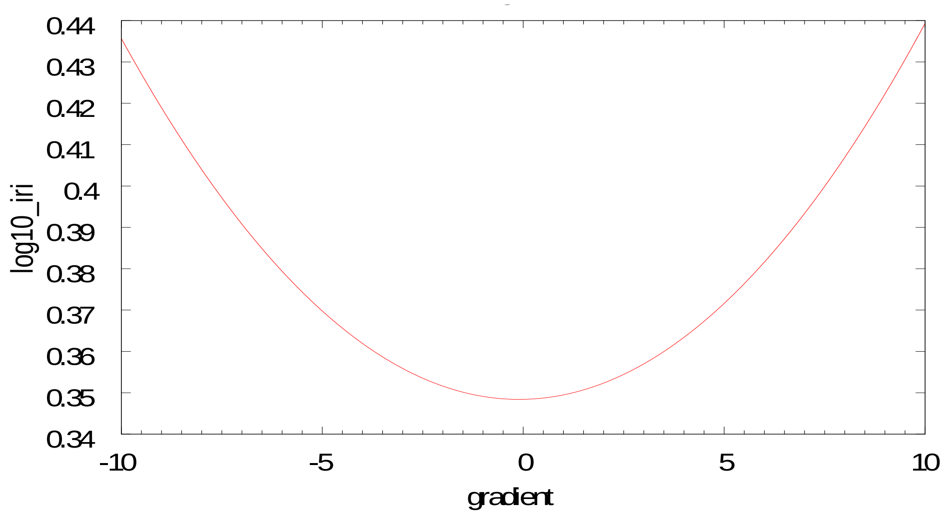


Figure A.13 IRI lane roughness versus gradient

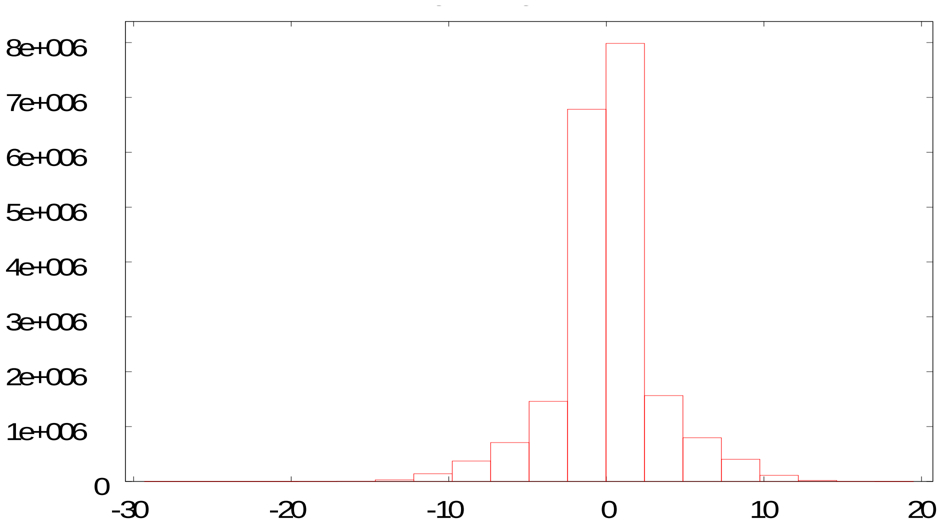


A.3.11 Gradient

Figure A.14 is a histogram of gradient. A positive gradient means uphill, whereas a negative gradient means downhill.

With reference to figure A.14, most values are between -10% and 10% gradient, but there are more extreme values in both directions.

Figure A.14 Histogram of gradient



A.3.12 Skid site and adjusted skid site categories

T10:2002 skid site categories are assigned to each 10m segment of state highway. Definitions of the categories are given in table A.6.

Table A.6 T10:2002 skid site categories

Skid site	Description
1	Railway level crossings, traffic signals, pedestrian crossings, stop and give-way signs, roundabout approaches
2	Urban curve <250m radius, rural curve <400m radius, down gradient >10%, on-ramps with ramp metering
3	Approaches to road junctions, down gradient 5–10%, motorway junction area
4	Normal roads (event free)
5	Divided carriageway (event free)

Divided roads were not considered in this study, so T10 skid site category 5 in table A.6 is not relevant.

Because we wanted curvature and gradient to be handled by the curvature and gradient predictors rather than skid site, there was a need to define an adjusted skid site category variable as defined in table A.7.

Table A.7 Adjusted skid site categories

Adjusted skid site	Description
4	Normal roads (event free)
3	Approaches to road junctions
2	Not used
1	Railway level crossings, traffic signals, pedestrian crossings, stop and give-way signs, roundabout approaches
Missing	Divided carriageway

A.3.13 Lane width

The carriageway table includes a lane-width variable. This appears to be incomplete and has not been used in this statistical modelling exercise.

A.3.14 Rut mean and rut standard deviation

These were shown not to be significant in an earlier analysis and so were not included in this statistical modelling exercise.

A.3.15 Crossfall

An analysis indicated that it was fairly marginal whether crossfall had a significant effect and so it was not included in this statistical modelling exercise.

A.3.16 Texture depth

There is a problem with including texture depth in this statistical modelling exercise. Asphaltic concrete surfaces (as opposed to chipseal) tend to be used in more populated areas. Roads with asphaltic concrete surfaces tend to have a lower texture, but being in more populated areas, may have a higher crash risk. Therefore, there is the possibility of a spurious correlation. By fitting both the urban-rural effect and the skid site categories in the model, one can reduce this spurious correlation, but it is necessary to include the surface type as a predictor variable. The information is available in RAMM, but it was outside the scope of the study undertaken.

Appendix B Fit of model

B.1 Comparison of fitted and observed counts

The state highway network was divided into segments by partitioning by carriageway area (the network is divided into 24 carriageway areas) and state highway number. Therefore, two sections of road were in the same partition if they were in the same carriageway area and on the same state highway. This gave 148 partitions. The model was then used to predict the number of crashes in each of these. The observed numbers were compared with the predicted number.

Figure B.1 shows the comparison for 'all' crashes and figure B.2 shows the normalised residual defined by equation B.1 in terms of *predicted*.

$$\frac{Observed\ Crash\ No. - Predicted\ Crash\ No.}{\sqrt{Predicted\ Crash\ No.}} \quad \text{(Equation B.1)}$$

Figure B.1 'All' crashes: observed versus predicted for 148 partitions of the state highway network

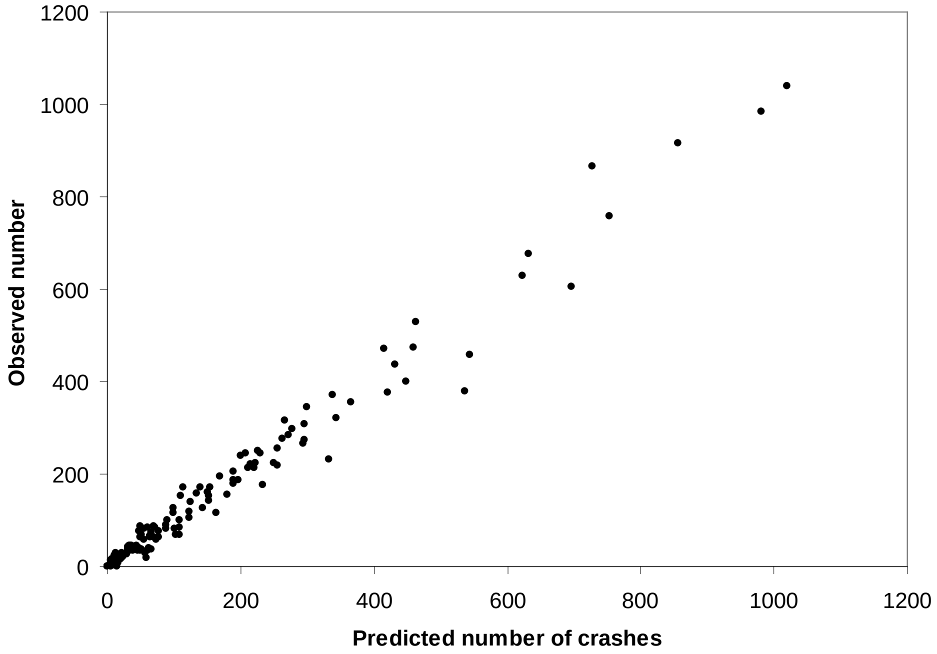
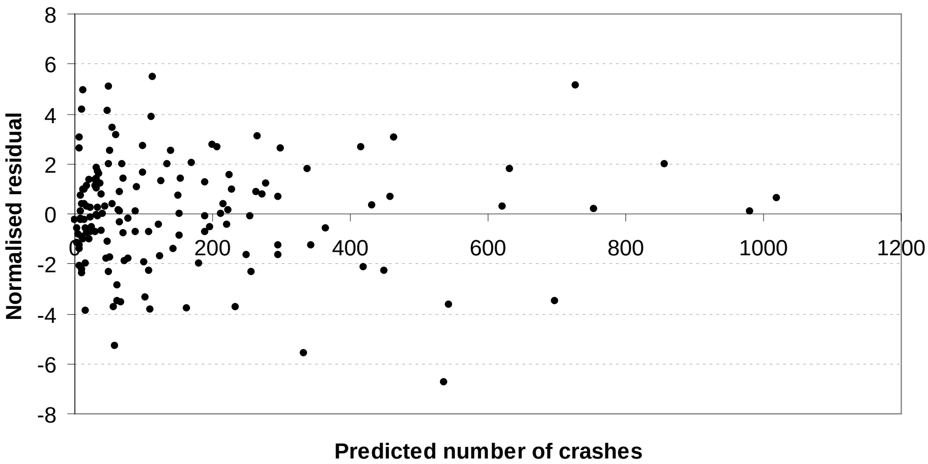


Figure B.2 'All' crashes: residual versus predicted



With reference to figure B.2, if the model was fitting perfectly there would be few points outside the range -2 to 2. The actual range of points is more like -4 to 4, with a few outside this range (particularly to the left of the graph). Therefore the model doesn't fit perfectly.

The corresponding plots for the 'wet' and 'selected' casualty crashes are given in figures B.3–B.6. The quality of the fit in both cases is about the same as for 'all' crashes.

Figure B.3 'Wet' crashes: observed versus predicted for 148 partitions of the state highway network

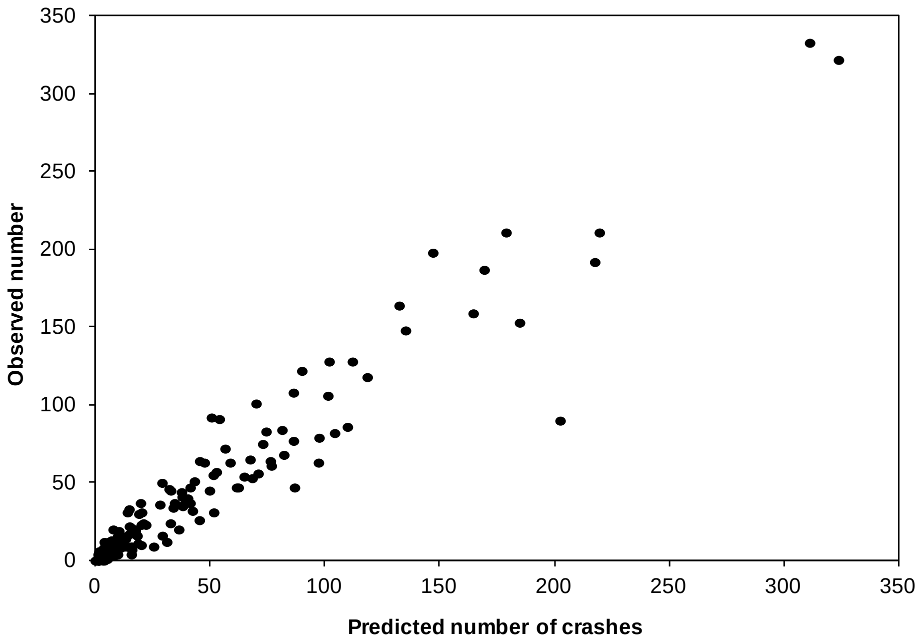


Figure B.4 'Wet' crashes: residual versus predicted

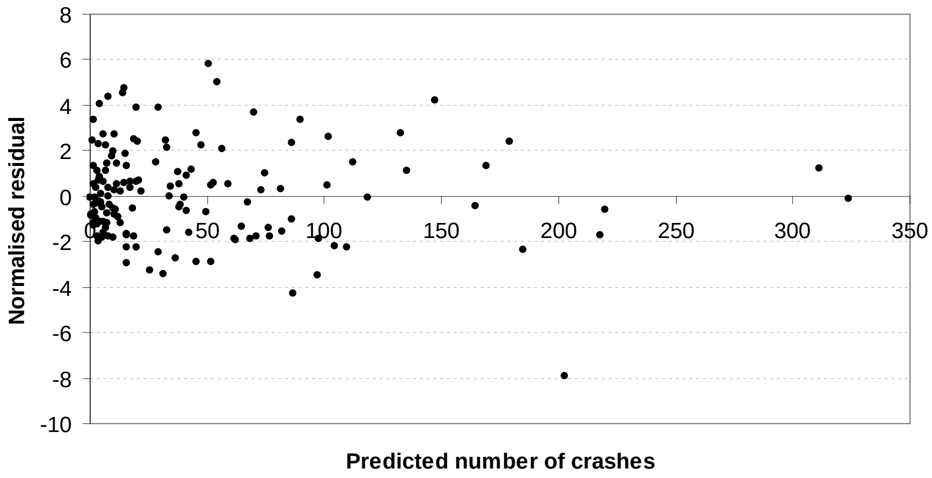


Figure B.5 'Selected' crashes: observed versus predicted for 148 partitions of the state highway network

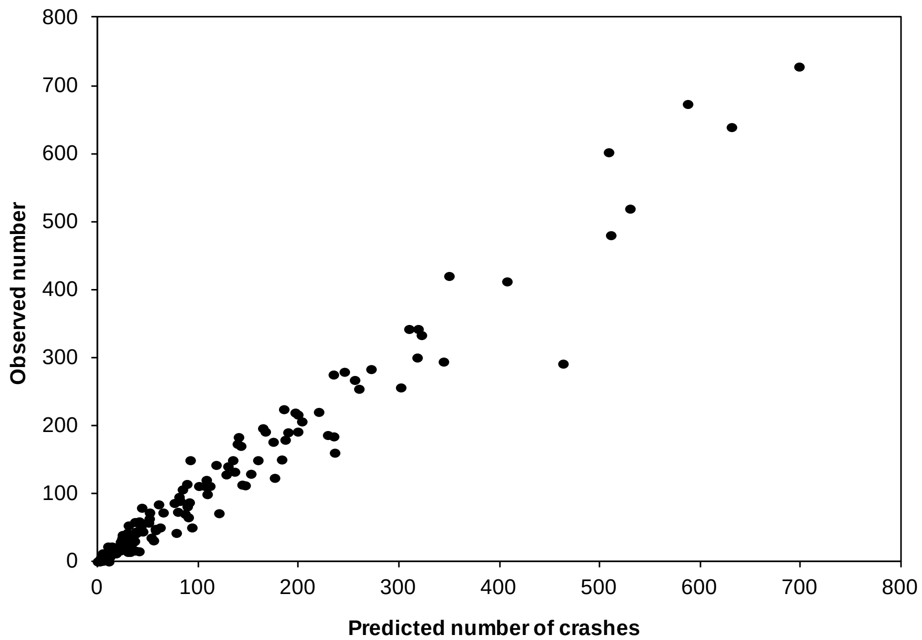
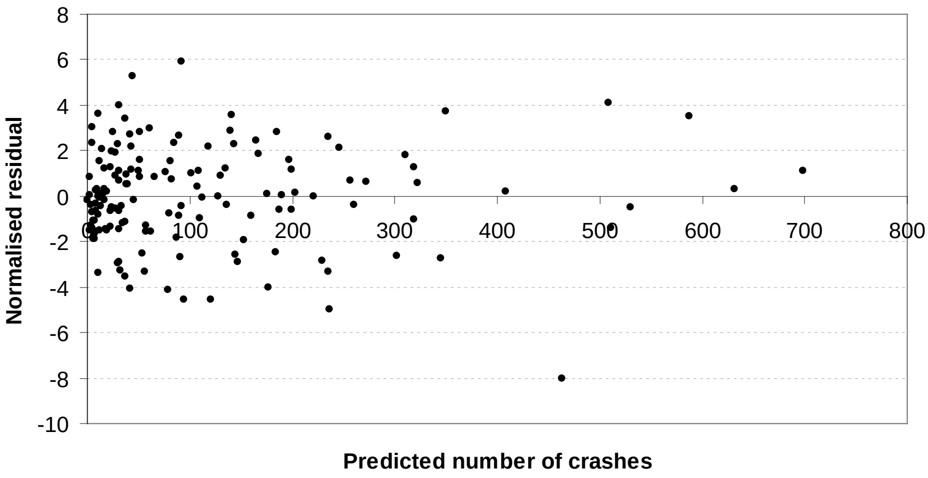


Figure B.6 'Selected' crashes: residual versus predicted



We could identify the particular points where the fit was bad and see if there was a data problem or if there were special risk factors. There was one point that possibly stood out as being low in the 'wet' and 'selected' graphs and, to a lesser extent in the 'all crashes' graph. This corresponded to SH25 in the East Waikato carriageway area. The possible reasons for this low value were not investigated.

Table B.1shows the numbers of crashes in each of the three categories of crashes that have just been considered and the sum of squares of the normalised residuals. If nothing had been fitted and the Poisson model was true, these would have chi-squared distributions with 148 (the number of categories) degrees of freedom. So the values would be close to 148. Because we have fitted parameters, the number of degrees of freedom needed to be reduced. We fitted 45 parameters (including the constant), but it would be wrong to reduce the number by 45 since there has been a lot of amalgamation of data. We chose 23 as a ball-park figure – so we had 125 degrees of freedom. So in the case of 'all crashes', we had a value that was too large by a factor of 5.4. This is where the suggested adjustment to the significance tests came from.

Table B.1 Sum of squares of normalised residuals for fits to 'all', 'wet' and 'selected' crashes

	'All' crashes	'Wet' crashes	'Selected' crashes
Number of crashes	22,870	6476	16,516
Chi-squared value	675	624	693

It is a little surprising that the chi-squared values remain high for the 'wet' crashes. We have a lot fewer crashes, so we would expect the randomness in the crashes to begin to mask whatever is causing the lack of fit.

The situation changes very little when we include a year-×-region interaction (refer to section 5.4), so the problem is probably not one of fluctuation of weather patterns in different regions. One possibility is that traffic patterns over a day and over a year are rather different from rainfall patterns over a day or year, and this might introduce additional variation into the 'wet' crash data.

B.2 Effect of the averaging

In applying the statistical model, the crash prediction from the log Poisson model was averaged over 21 adjacent 10m segments, ie the segments within a range of 10 segments of the one in which we were actually making the estimate. This section investigates the effect of altering the number of segments being averaged. Table B.2 shows the log-likelihood and the type I chi-squared values for the 'all' crashes model for various values of averaging length. The maximum values in each line of the table are shown in bold.

Table B.2 Effect of averaging length on log-likelihood and type 1 chi-squared values

Effect	Averaging length (metres)				
	410	210	110	50	30
Log-likelihood	-152,161	-151,293	**-150,680**	-150,803	-150,917
year	**528**	526	525	521	519
region	**678**	665	655	657	657
urban rural	383	486	**552**	546	523
adjusted skid site	4318	6289	**7319**	6541	6053
OOCC	**5871**	5400	4419	3401	3108
curvature	263	460	752	1125	**1228**
ADT	430	518	590	632	**640**
SCRIM skid resistance	250	**265**	237	249	256
gradient	62	66	**69**	64	62
adjusted IRI	40	109	160	**277**	275
curvature-x-adj IRI	85	107	102	117	**120**

The analyses in the report use an averaging length of 210m. If we just look at log-likelihood, then we might decide that an averaging length of 110m is more appropriate. However, the different chi-squared values don't show any consistent pattern and it is hard to make much sense of them.

Appendix C Calculation of OOCC effect

C.1 85 percentile speed calculation

The 85 percentile speed can be reasonably determined by inputting 10m radius and crossfall data from the geometry table in the NZTA's RAMM database into the 'advisory speed' formula below:

$$AS = -\left(\frac{107.95}{H}\right) + \sqrt{\frac{107.95^2}{H} + \left(\frac{127{,}000}{H}\right) \times \left(0.3 + \frac{X}{100}\right)} \qquad \text{Equation C.1}$$

where:

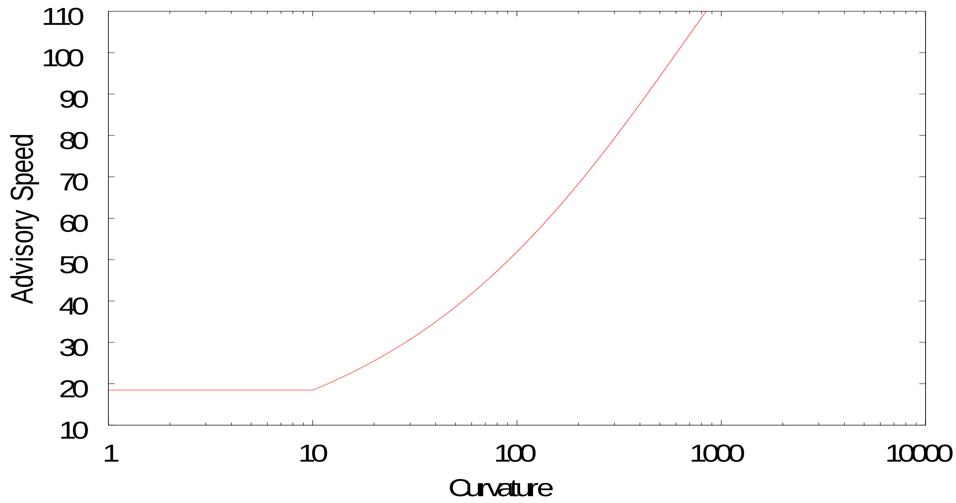
AS	=	advisory speed (km/h)≈85 percentile speed (km/h)
X	=	% crossfall (sign relative to curvature)
H	=	absolute curvature (rad/km) = 1000/R
R	=	horizontal radius of curvature (m).

If R <0 then the sign of X was switched. The range of X is limited to 0–30%.

The resulting value of AS was capped at 110km/h for rural sections of state highways and 70 km/h for urban sections of state highways.

Figure C.1 shows a plot of advisory speed versus curvature with crossfall set to zero. The curve is horizontal for radii of curvature less than 10m (in absolute value) because for this statistical modelling exercise, all radii of curvature less than 10m have been replaced with a value of 10m.

Figure C.1 Advisory speed versus curvature



C.2 The out-of-context-curve (OOCC) effect

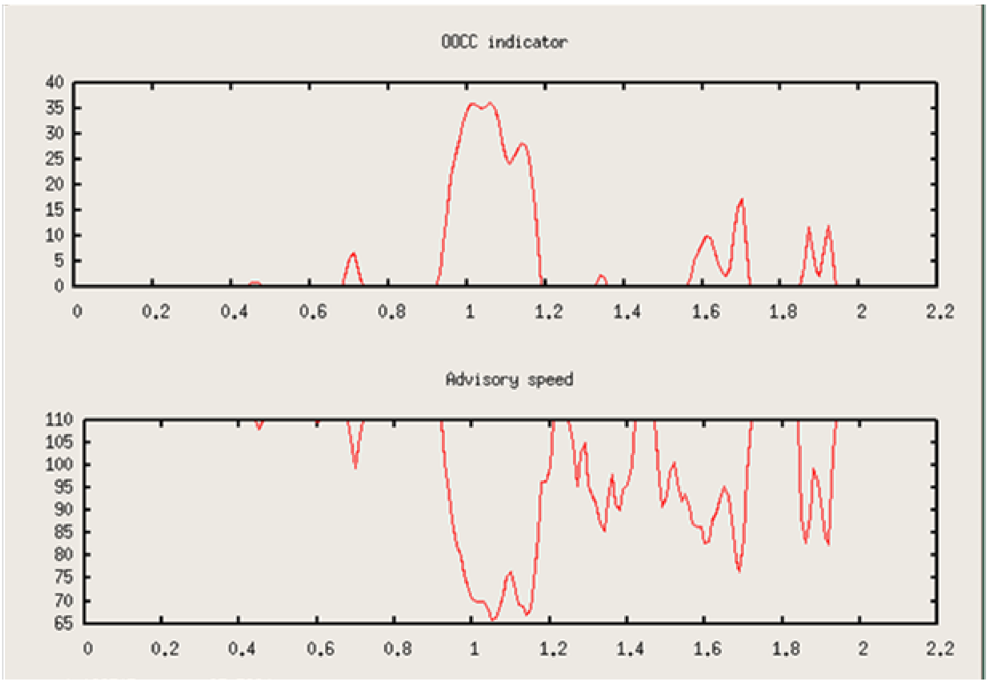
The *OOCC* indicator for a particular 10m road section on the increasing lane is calculated as follows:

1. The local speed, AS1, is derived by averaging the advisory speed calculated using equation C.1 for the current 10m section (ie 10m section of interest) and the two preceeding 10m sections.
2. The approach speed, AS2, is derived by averaging the advisory speed calculated using equation C.1 for the 50 10m sections preceding the three used to calculate the *AS1*.
3. The *OOCC* indicator is zero if $AS2 \le AS1$, and $AS2 - AS1$ if $AS2 > AS1$.

For the *decreasing* lane, the same calculation steps are performed, but in the opposite direction.

With reference to figure C.2, the ideal situation would be for the OOCC indicator to die away rather more quickly after the beginning of the curve. However, in terms of predicting crash rates, the present formulation works better than anything else tried.

Figure C.2 OOCC indicator graph



Appendix D IRI adjustment calculation

D.1 Calculation steps

The adjustment of the IRI roughness value to account for the effect of curvature and gradient involves the following steps:

1. The horizontal curvature is transformed by taking the log_{10} of the absolute value and setting the result to the lower bound value of 1 if less than 1 or the upper bound value of 5 if greater than 5.
2. The transformed value of horizontal curvature and gradient are input into the adjustment calculation given by rows 1–8 in table C1, which involves a constant (row 1), a fifth-order polynomial of log_{10}*(absolute curvature)* (rows 2–6) and a second-degree polynomial of gradient (rows 7–8).
3. The correction is calculated by subtracting the sum of the adjustment calculation components from the base level of 0.3484115, which will provide 0 correction when horizontal curvature equals 100000m and gradient equals 1%.
4. The correction is subtracted from the log_{10}*(IRI)* to give the *adjusted* log_{10}*(IRI)* value
5. The corrected IRI value can be calculated simply by taking the antilog of the *adjusted* log_{10}*(IRI)* or dividing the IRI value by the antilog of the correction value.

Table D1 Adjustment polynomial

Row no.	Predictor variable	Model coefficient
1	constant	-0.51774158
2	bound_log10_abs_curvature**1	2.736878766
3	bound_log10_abs_curvature**2	-2.27852495
4	bound_log10_abs_curvature**3	0.82384106
5	bound_log10_abs_curvature**4	-0.13815523
6	bound_log10_abs_curvature**5	0.008803766
7	gradient**1	0.000184087
8	gradient**2	0.000890999

D.2 Illustrative calculation

The calculation procedure described in section D.1 is applied to the example situation of:

- a horizontal curvature of 5000m
- a gradient of 0%
- a lane IRI roughness of 2 =mm/m.

The transformed horizontal curvature value is 3.69897 (ie log_{10}*(5000))*. Since the value lies between 1 and 5, it does not need to be bounded.

Table D.2 summarises the calculation steps as performed in the Microsoft Excel™ spreadsheets *fitted_all.xls*, *fitted_wet.xls*, *fitted_sel_xls* and *fitted_wet_sel.xls*, which can be downloaded from the NZTA website (www.nzta.govt.nz).

Table D.2 Example IRI adjustment calculation

Predictor variable	Model coefficient	Value of variable	Product (value × coefficient)
constant	-0.51774158	1	-0.517742
bound_log10_abs_curvature**1	2.736878766	3.69897	10.123632
bound_log10_abs_curvature**2	-2.27852495	13.68238	-31.17564
bound_log10_abs_curvature**3	0.82384106	50.61071	41.695181
bound_log10_abs_curvature**4	-0.13815523	187.2075	-25.86369
bound_log10_abs_curvature**5	0.008803766	692.4749	6.0963872
gradient**1	0.000184087	0	0
gradient**2	0.000890999	0	0
		Σ:	0.358122
		Base level:	0.3484115
		Correction:	0.0097105
		$10^{(Correction)}$	1.0226111

Therefore for a lane IRI value of 2:

- the adjusted $\log_{10}$(IRI) = $\log_{10}$(2) – 0.0097105 = 0.2913195
- the *adjusted IRI = 2÷1.022611= 1.956 mm/m*

Appendix E Model coefficients

E.1 Tables of model coefficients

The following tables summarise the regression coefficients for the 'all', 'wet', 'selected' and 'wet selected' crash rate prediction models corresponding to the coding of the Microsoft Excel™ spreadsheets *fitted_all.xls*, *fitted_wet.xls*, *fitted_sel_xls* and *fitted_wet_sel.xls,* which can be downloaded from the NZTA website (www.nzta.govt.nz).

Table E.1 Model coefficients for 'all' casualty crash rates

Predictor variable	Model statistics		
	Coefficient	S.E.	Ratio
constant	-8.91855	1.5417	-5.785
year:2000	0		
year:2001	0.109205	0.032422	3.3683
year:2002	0.247343	0.031694	7.804
year:2003	0.238247	0.031266	7.62
year:2004	0.232857	0.031305	7.4384
year:2005	0.235531	0.031055	7.5842
year:2006	0.295369	0.030688	9.625
year:2007	0.365291	0.030352	12.035
year:2008	0.202345	0.031429	6.4382
year:2009	-0.25118	0.034691	-7.2405
region:R01	0		
region:R02	-0.3796	0.043852	-8.6564
region:R03	-0.14205	0.027619	-5.1432
region:R04	-0.14638	0.034058	-4.298
region:R05	-0.1046	0.054999	-1.9019
region:R06	0.047882	0.037183	1.2877
region:R07	0.053738	0.037214	1.444
region:R08	-0.06228	0.031594	-1.9713
region:R09	-0.01674	0.040607	-0.41217
region:R10	-0.0313	0.036024	-0.86887
region:R11	-0.24174	0.032098	-7.5312
region:R12	-0.28411	0.046289	-6.1377
region:R13	0.039511	0.03157	1.2515
region:R14	0.096712	0.039059	2.4761
urban_rural:U	0		
urban_rural:R	0.119504	0.022661	5.2736
adj_skid_site:4	0		
adj_skid_site:3	1.610236	0.025046	64.291
adj_skid_site:1	1.871158	0.050226	37.255
bound_OOCC**1	-0.01228	0.011831	-1.0376
bound_OOCC**2	0.00319	0.000927	3.443
bound_OOCC**3	-5.5E-05	1.81E-05	-3.0522
bound_log10_abs_curvature**1	-3.48945	0.67689	-5.1551
bound_log10_abs_curvature**2	0.491136	0.11127	4.414
log10_ADT**1	0.36854	0.24587	1.4989
log10_ADT**2	-0.12283	0.034023	-3.6103
scrim-0.5000**1	-1.77861	0.12049	-14.762
scrim-0.5000**2	1.168532	1.0013	1.167
bound_abs_gradient**1	0.164931	0.53955	0.30569
bound_abs_gradient**2	-0.01713	0.084138	-0.20356
bound_abs_gradient**3	0.000751	0.004139	0.18145
bound_adj_log10_iri**1	0.118761	6.3295	0.018763
bound_adj_log10_iri**2	-27.8012	9.1483	-3.0389
bound_adj_log10_iri**3	-1.57226	0.93719	-1.6776
bound_log10_abs_curvature**1.bound_adj_log10_iri**1	-0.26655	4.3747	-0.06093
bound_log10_abs_curvature**1.bound_adj_log10_iri**2	18.8887	6.3699	2.9653
bound_log10_abs_curvature**2.bound_adj_log10_iri**1	-0.03185	0.71379	-0.04462
bound_log10_abs_curvature**2.bound_adj_log10_iri**2	-2.79786	1.0265	-2.7257

Table E.2 Model coefficients for 'wet' casualty crash rates

Predictor variable	Model statistics		
	Coefficient	S.E.	Ratio
constant	-13.7068	2.6587	-5.1556
year:2000	0		
year:2001	0.216156	0.058892	3.6704
year:2002	0.289379	0.058896	4.9134
year:2003	0.161567	0.058971	2.7398
year:2004	0.296033	0.057195	5.1759
year:2005	0.196402	0.058016	3.3853
year:2006	0.238524	0.057388	4.1563
year:2007	0.330196	0.056824	5.8108
year:2008	-0.05255	0.060038	-0.87527
year:2009	-0.33419	0.065415	-5.1088
region:R01	0		
region:R02	-0.19626	0.074638	-2.6295
region:R03	-0.08758	0.048895	-1.7912
region:R04	-0.08954	0.061937	-1.4457
region:R05	-0.21315	0.10574	-2.0158
region:R06	-0.00386	0.067623	-0.05709
region:R07	0.264025	0.065537	4.0286
region:R08	-0.08725	0.058465	-1.4923
region:R09	0.040161	0.075486	0.53203
region:R10	-0.21106	0.070041	-3.0134
region:R11	-0.49337	0.064213	-7.6833
region:R12	0.264128	0.075002	3.5216
region:R13	-0.21238	0.062577	-3.3939
region:R14	0.274234	0.0708	3.8733
urban_rural:U	0		
urban_rural:R	0.28952	0.046404	6.2391
adj_skid_site:4	0		
adj_skid_site:3	1.323964	0.052527	25.205
adj_skid_site:1	1.291555	0.11779	10.965
bound_OOCC**1	-0.03688	0.019991	-1.845
bound_OOCC**2	0.005748	0.001504	3.8215
bound_OOCC**3	-0.00011	2.87E-05	-3.697
bound_log10_abs_curvature**1	-4.95618	1.2537	-3.9533
bound_log10_abs_curvature**2	0.685837	0.20951	3.2736
log10_ADT**1	2.158552	0.48061	4.4913
log10_ADT**2	-0.36243	0.066587	-5.443
scrim-0.5000**1	-4.00498	0.22821	-17.549
scrim-0.5000**2	4.3763	1.6805	2.6042
bound_abs_gradient**1	1.3885	0.86513	1.605
bound_abs_gradient**2	-0.19777	0.13435	-1.4721
bound_abs_gradient**3	0.009417	0.00658	1.4312
bound_adj_log10_iri**1	2.949255	11.298	0.26103
bound_adj_log10_iri**2	-32.6665	15.84	-2.0623
bound_adj_log10_iri**3	-0.24495	1.7287	-0.1417
bound_log10_abs_curvature**1.bound_adj_log10_iri**1	-1.82795	7.9899	-0.22878
bound_log10_abs_curvature**1.bound_adj_log10_iri**2	21.43343	11.353	1.888
bound_log10_abs_curvature**2.bound_adj_log10_iri**1	0.236115	1.3252	0.17817
bound_log10_abs_curvature**2.bound_adj_log10_iri**2	-3.25395	1.8619	-1.7477

Table E.3 Model coefficients for 'selected all' casualty crash rates

Predictor variable	Model statistics		
	Coefficient	S.E.	Ratio
constant	-12.6718	1.6748	-7.566
year:2000	0		
year:2001	0.085456	0.038321	2.23
year:2002	0.228284	0.037399	6.1039
year:2003	0.238775	0.036741	6.4988
year:2004	0.218525	0.036876	5.9259
year:2005	0.253614	0.036365	6.9741
year:2006	0.313933	0.035937	8.7357
year:2007	0.407871	0.035457	11.503
year:2008	0.151282	0.037105	4.0771
year:2009	-0.25663	0.040835	-6.2845
region:R01	0		
region:R02	-0.2643	0.0516	-5.1221
region:R03	-0.09066	0.032174	-2.8177
region:R04	-0.09987	0.039972	-2.4985
region:R05	-0.08047	0.063133	-1.2746
region:R06	0.027534	0.043507	0.63287
region:R07	0.045147	0.045748	0.98687
region:R08	-0.03222	0.037267	-0.86469
region:R09	0.099612	0.049175	2.0256
region:R10	-0.05864	0.042708	-1.373
region:R11	-0.18855	0.037969	-4.9658
region:R12	-0.2261	0.052246	-4.3277
region:R13	0.117788	0.03686	3.1955
region:R14	0.201889	0.044644	4.5222
urban_rural:U	0		
urban_rural:R	0.310655	0.031277	9.9324
adj_skid_site:4	0		
adj_skid_site:3	0.784518	0.044496	17.631
adj_skid_site:1	1.169093	0.080368	14.547
bound_OOCC**1	-0.01378	0.01311	-1.0512
bound_OOCC**2	0.003379	0.001007	3.3564
bound_OOCC**3	-5.9E-05	1.95E-05	-3.0363
bound_log10_abs_curvature**1	-2.63723	0.73981	-3.5647
bound_log10_abs_curvature**2	0.312073	0.12272	2.5429
log10_ADT**1	1.324669	0.2893	4.5789
log10_ADT**2	-0.27911	0.04049	-6.8933
scrim-0.5000**1	-2.28265	0.13921	-16.398
scrim-0.5000**2	2.711952	1.1216	2.4179
bound_abs_gradient**1	0.732892	0.57867	1.2665
bound_abs_gradient**2	-0.09748	0.090149	-1.0813
bound_abs_gradient**3	0.004273	0.004431	0.96448
bound_adj_log10_iri**1	7.691234	7.0121	1.0968
bound_adj_log10_iri**2	-30.0854	10.272	-2.9289
bound_adj_log10_iri**3	-0.19299	1.0581	-0.18239
bound_log10_abs_curvature**1.bound_adj_log10_iri**1	-6.07777	4.884	-1.2444
bound_log10_abs_curvature**1.bound_adj_log10_iri**2	20.57531	7.2091	2.8541
bound_log10_abs_curvature**2.bound_adj_log10_iri**1	1.001927	0.80565	1.2436
bound_log10_abs_curvature**2.bound_adj_log10_iri**2	-3.20082	1.1762	-2.7213

Table E.4 Model coefficients for 'selected wet' casualty crash rates

Predictor variable	Model statistics		
	Coefficient	S.E.	Ratio
constant	-17.2725	2.769	-6.2379
year:2000	0		
year:2001	0.20353	0.066016	3.083
year:2002	0.255531	0.06629	3.8548
year:2003	0.172717	0.065745	2.6271
year:2004	0.298435	0.063825	4.6758
year:2005	0.224584	0.064414	3.4866
year:2006	0.244509	0.06402	3.8193
year:2007	0.365524	0.063177	5.7857
year:2008	-0.09517	0.067219	-1.4158
year:2009	-0.3164	0.0727	-4.3522
region:R01	0		
region:R02	-0.11131	0.082467	-1.3498
region:R03	-0.0714	0.053841	-1.3262
region:R04	-0.07784	0.068983	-1.1284
region:R05	-0.24264	0.1169	-2.0756
region:R06	0.01294	0.073661	0.17566
region:R07	0.198854	0.076236	2.6084
region:R08	-0.07059	0.065185	-1.0829
region:R09	0.148088	0.084976	1.7427
region:R10	-0.20001	0.076945	-2.5993
region:R11	-0.47437	0.072369	-6.5549
region:R12	0.294735	0.081223	3.6287
region:R13	-0.15302	0.069277	-2.2088
region:R14	0.33728	0.078119	4.3175
urban_rural:U	0		
urban_rural:R	0.524459	0.060305	8.6967
adj_skid_site:4	0		
adj_skid_site:3	0.682127	0.083215	8.1972
adj_skid_site:1	0.763025	0.17115	4.4582
bound_OOCC**1	-0.02929	0.021181	-1.3828
bound_OOCC**2	0.005114	0.001576	3.2452
bound_OOCC**3	-9.6E-05	2.99E-05	-3.194
bound_log10_abs_curvature**1	-4.20988	1.2875	-3.2699
bound_log10_abs_curvature**2	0.529936	0.21714	2.4406
log10_ADT**1	3.243258	0.54674	5.932
log10_ADT**2	-0.53266	0.076438	-6.9685
scrim-0.5000**1	-4.45343	0.25287	-17.612
scrim-0.5000**2	6.062047	1.7858	3.3946
bound_abs_gradient**1	1.787674	0.89791	1.9909
bound_abs_gradient**2	-0.25464	0.13933	-1.8276
bound_abs_gradient**3	0.011912	0.006819	1.7469
bound_adj_log10_iri**1	8.614876	11.782	0.73121
bound_adj_log10_iri**2	-34.1862	16.94	-2.0181
bound_adj_log10_iri**3	-0.70335	1.9226	-0.36582
bound_log10_abs_curvature**1.bound_adj_log10_iri**1	-6.01232	8.352	-0.71986
bound_log10_abs_curvature**1.bound_adj_log10_iri**2	22.75693	12.111	1.8791
bound_log10_abs_curvature**2.bound_adj_log10_iri**1	0.895003	1.3962	0.64103
bound_log10_abs_curvature**2.bound_adj_log10_iri**2	-3.40385	1.9976	-1.7039

E.2 Illustrative calculation

To see how the model is applied, the baseline parameter values used in generating the crash rate trend plots presented in section 3.5.1 have been inputted to the 'all' casualty crash rate model summarised in table E.1. The baseline parameter values are as tabulated in table 3.7 and this table has been reproduced as table E.5 below for ready reference.

Table E.5 Baseline parameter values used in generating crash rate trend plots

Parameter	Baseline value
year	2008
region	R03
urban_rural	R
adj_skid_site	4
OOCC	0
curvature	5000
ADT	1000
gradient	0
scrim	0.5
adj_log10_iri	0.3

Table E.6 shows the results of applying the the 'all' casualty crash rate model. As can be seen the estimated personal risk (ie crash rate) is 12.63 injury crashes per 100 million kilometres travelled and the estimated collective risk (ie crash density in terms of all reported injury crashes per year per 10m of lane) is 0.00046.

Table E.6 Calculation of personal and collective risks using baseline parameter values as input

Predictor variable	Model coefficient	Value of variable	Product (value × coefficient)
constant	-8.91855	1	-8.91855
year:2008	0.202345	1	0.202345
region:R03	-0.14205	1	-0.14205
urban_rural:R	0.119504	1	0.119504
adj_skid_site:4	0	1	0
bound_OOCC**1	-0.01228	0	0
bound_OOCC**2	0.00319	0	0
bound_OOCC**3	-5.5E-05	0	0
bound_log10_abs_curvature**1	-3.48945	3.69897	-12.9074
bound_log10_abs_curvature**2	0.491136	13.68238	6.719914
log10_ADT**1	0.36854	3	1.105619
log10_ADT**2	-0.12283	9	-1.1055
scrim-0.5000**1	-1.77861	0	0
scrim-0.5000**2	1.168532	0	0
bound_abs_gradient**1	0.164931	4	0.659725
bound_abs_gradient**2	-0.01713	16	-0.27404
bound_abs_gradient**3	0.000751	64	0.048069
bound_adj_log10_iri**1	0.118761	0.290289	0.034475
bound_adj_log10_iri**2	-27.8012	0.084268	-2.34275
bound_adj_log10_iri**3	-1.57226	0.024462	-0.03846
bound_log10_abs_curvature**1.bound_adj_log10_iri**1	-0.26655	1.073772	-0.28622
bound_log10_abs_curvature**1.bound_adj_log10_iri**2	18.8887	0.311705	5.887697
bound_log10_abs_curvature**2.bound_adj_log10_iri**1	-0.03185	3.971851	-0.12649
bound_log10_abs_curvature**2.bound_adj_log10_iri**2	-2.79786	1.152987	-3.22589
			Σ=-14.59
	Personal risk (injury crashes per 10^8 vkt)		12.63
	Collective risk (injury crashes per 10m)		0.00046

Appendix F KiwiRAP model variant

F.1 Model fit

The analysis of variance, parameter bounds and model coefficients are sumamrised in tables F.1–F.3. below and pertain to 'all' casualty crashes.

Table F.1 Table of variance

Predictor variable	Degrees of freedom (df)	Chi-square	
		Type III (added last)	Type 1 (sequential)
year	5	128.56	88.397
region	6	89.968	120.86
urban_rural	1	12.532	175.3
adj_skid_site	2	2352.2	3145.5
poly3_bound_OOCC	3	268.81	2678.7
poly2_bound_log10_abs_curvature	2	39.81	53.279
poly2_log10_ADT	2	388.07	343.29
poly2_scrim-0.5000	2	111.01	127.09
poly3_bound_adj_log10_iri	3	46.175	45.391
poly3_bound_abs_gradient	3	14.096	14.096

Table F.2 Parameter ranges

Limitations in the range of data that was available for the model fitting and the analysis method means that the model is limited in its application to the following parameter ranges:	
year:	1997 to 2002 (beyond these years requires estimation of the yearly coefficient)
region:	R1 to R7 (= NZTA administration regions, where R1=Auckland, R2=Hamilton, R3=Napier, R4=Whanganui, R5=Wellington, R6=Christchurch and R7=Dunedin)
urban_rural:	U (urban) or R (rural)
skid_site:	T10 site category 1, 3 or 4 (category 2 has been combined into category 4)
curvature:	100m to 10000m radius (absolute value used; ie does not differentiate left- from right-hand curves). For radii outside this range, use 100m for values less than 100m, and 10,000m for values greater than 10,000m
ADT:	average daily traffic, unlimited range of values
gradient:	4 to 10 (absolute value is used, and values less than 4 are set equal to 4)
SCRIM:	0.3 to 0.7 SCRIM coefficient
IRI:	2.0 to 10.0 IRI (m/km) lane roughness

Table F.3 Model coefficients

Predictor variable	Model statistics		
	Effect	Standard error	t-value
constant	-13.916	1.9212	-7.2434
constant (SCRIM values replaced by IL values)	-14.043	–	–
year:1997	0		
year:1998	-0.06314	0.031775	-1.9872
year:1999	-0.05173	0.032188	-1.6071
year:2000	-0.10808	0.032434	-3.3325
year:2001	-0.00217	0.031895	-0.06794
year:2002	0.19928	0.030735	6.4836
region:R1	0		
region:R2	0.12921	0.033399	3.8686
region:R3	0.19913	0.045251	4.4006
region:R4	0.29469	0.037131	7.9365
region:R5	0.23685	0.040852	5.7978
region:R6	0.080057	0.040135	1.9947
region:R7	0.12308	0.040482	3.0404
urban_rural:R	0		
urban_rural:U	-0.11288	0.031887	-3.5401
adj_skid_site:4	0		
adj_skid_site:3	1.6191	0.034804	46.519
adj_skid_site:1	1.8544	0.080171	23.13
bound_OOCC**1	0.018871	0.020104	0.93865
bound_OOCC**2	0.001442	0.001554	0.92789
bound_OOCC**3	-1.69E-05	3.02E-05	-0.56063
bound_log10_abs_curvature**1	1.0318	0.43304	2.3827
bound_log10_abs_curvature**2	-0.1952	0.066934	-2.9164
log10_ADT**1	0.50289	0.31224	1.6106
log10_ADT**2	-0.14548	0.043513	-3.3433
scrim-0.5000**1	-1.6266	0.15451	-10.527
scrim-0.5000**2	0.28664	1.2857	0.22293
bound_adj_log10_iri**1	-12.503	4.6567	-2.6849
bound_adj_log10_iri**2	23.159	8.8839	2.6068
bound_adj_log10_iri**3	-12.319	5.2829	-2.3318
bound_abs_gradient**1	-0.01497	0.75572	-0.0198
bound_abs_gradient**2	0.008727	0.1179	0.074022
bound_abs_gradient**3	-0.00049	0.005799	-0.08447

F.2 Illustrative application

Parameter values as tabulated in table F.4 were input in the model to provide example calculations of personal and collective risk calculated from:

$$\text{Collective risk} = \alpha \exp(L) \qquad \text{Equation F.1}$$

$$\text{Personal risk} = \frac{10^{10}}{365} \exp(L) \qquad \text{Equation F.2}$$

where L is the sum of the components of the log-linear model.

Table F.4 Parameter values for example application of KiwiRAP model

Parameter	Example value
year	2002
region	R2
urban_rural	rural
adj_skid_site	4
OOCC (km/h)	15
curvature (m)	300
ADT (v/l/d)	10000
gradient (%)	4
SCRIM skid resistance	0.4 (=IL for skid site 4)
Iri lane roughness (mm/m)	0.3

The required steps in calculating L are summarised in table F.5 below.

In applying the model to calculate the horizontal alignment road protection score, the SCRIM skid resistance values are replaced by the T10:2002 skid resistance investigatory-level value pertaining to the 10m segment of interest. In this case, the model constant changes from -13.916 to -14.043.

Also, when calculating the horizontal alignment road protection score, we remove the effect of traffic as our only interest is geometric features. This is achieved by either setting the ADT model coefficients to zero or setting ADT to 1 when calculating the terms making up 'L'.

Figure F.1 shows the difference in fit resulting from using investigatory-level skid resistance rather than measured skid resistance for a section of SH29.

Figure F.1 Level of agreement between predicted and observed crashes on a section of SH29

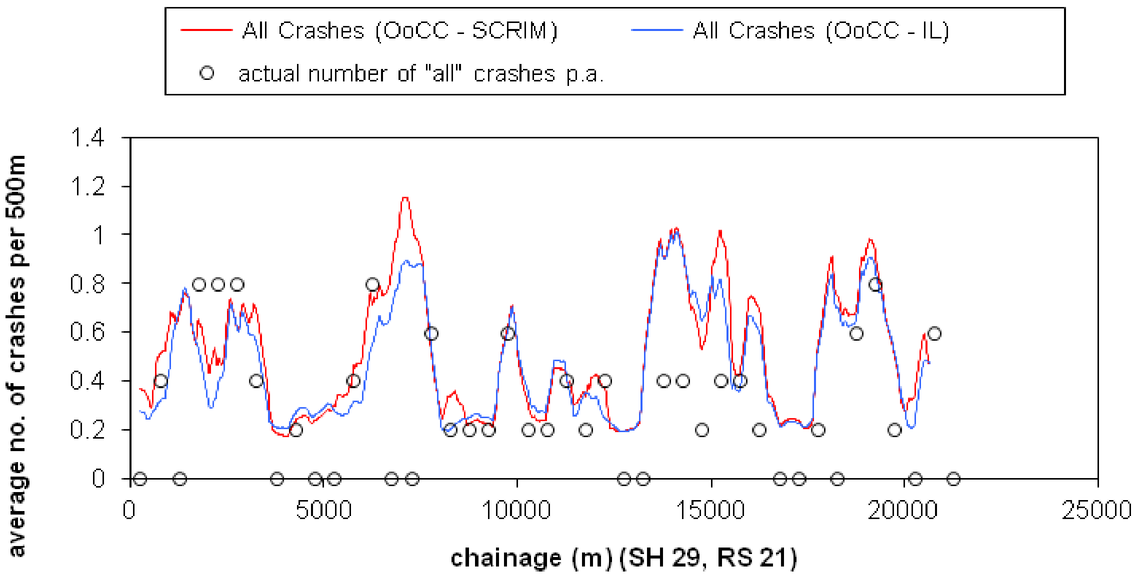


Table F.5 Example application of KiwiRAP model

Predictor variable	Value of variable	Model coefficient	Product (value x coefficient)
constant	1	-14.043	-14.043
year	1	0.19928	0.199
region	1	0.12921	0.129
urban_rural	1	0	0.000
skid_site	1	0	0.000
$\log_{10}(\lvert curvature \rvert)$	2.477121255	1.0318	2.556
$[\log_{10}(\lvert curvature \rvert)]^2$	6.136129711	-0.1952	-1.198
$\log_{10}(ADT)$	4	0.50289	2.012
$[\log_{10}(ADT)]^2$	16	-0.14548	-2.328
$\lvert gradient \rvert$	4	-0.014965	-0.060
$\lvert gradient \rvert^2$	16	0.008727	0.140
$\lvert gradient \rvert^3$	64	-0.00048983	-0.031
(SCRIM-0.5)	-0.1	-1.6266	0.163
$(SCRIM-0.5)^2$	0.01	0.28664	0.003
$\log_{10}(iri)$	0.477121255	-12.503	-5.965
$[\log_{10}(iri)]^2$	0.227644692	23.159	5.272
$[\log_{10}(iri)]^3$	0.108614121	-12.319	-1.338
OOCC	15	0.018871	0.283
$OOCC^2$	225	0.0014419	0.324
$OOCC^3$	3375	-0.00001693	-0.057
			Σ=-13.940=*L*
	Personal risk (expected casualty crashes per year per 10^8 vkt)		24.20
	Collective risk (expected casualty crashes per year per10m)		0.008833

Appendix G Curve context model variant

G.1 Introduction

Loss of control on curves remains the largest cause of crashes on New Zealand's rural state highways, comprising 1309 reported injury crashes in 2009 (MOT 2010). This represents 49% of reported injury crashes on rural state highways and 36% of all reported injury crashes. Of these 1309 crashes, 1210 (92%) occurred on curves classified by New Zealand Police as either moderate or easy, and 471 (36%) occurred in wet conditions. Since 1997/98, with the issuing of the T10 specification for state highway skid resistance management, curves with a horizontal radius of curvature less than 250m have been effectively managed to a skid resistance level that is 25% greater than for all other curves on rural state highways. The T10 specification aimed to equalise the risk across the state highway network of a skidding crash in the wet, by assigning investigatory skid resistance levels (in terms of equilibrium skid resistance – ESC) for different site categories that are related to different friction demands. A description of these site categories and associated investigatory levels (IL) are summarised in table G.1 below. As can be seen, curves below 250m horizontal radius of curvature are assigned a higher IL than curves with a horizontal curvature of radius 250m or greater.

By incorporating the concept of a 'threshold level' (TL)for skid resistance, the policy effectively sets a minimum level of service. The TL is the trigger level at which urgent remedial work should be undertaken. The TL is currently set at 0.1ESC below the IL. In practice, the policy results in curves of less than 250m horizontal radius of curvature being immediately investigated and treated when the skid resistance falls below the TL of 0.4ESC. Curves of 250m horizontal radius of curvature or more are treated only when the skid resistance falls below the TL of 0.3ESC.

Table G.1 T10:2002 skid site categories

Site category	Description	Notes	Investigatory level (ESC)
5	Divided carriageways		0.35
4	Normal roads	Undivided carriageways only	0.40
3	Approaches to road junctions – down gradients 5–10%	Includes motorway on/off ramps	0.45
2	Curve <250m radius, down gradients >10%		0.50
1	Highest priority	Railway level crossings, approaches to roundabouts, traffic lights, pedestrian crossings and similar hazards	0.55

As not all small-radius curves constitute a safety hazard, and not all moderate-to-large radius curves have a low crash risk, statistical modelling was undertaken to allow estimation of crash risk for any curve of less than 500m horizontal radius on New Zealand's rural state highways. The resulting model is presented below following the section outling the methodology used to define a curve.

G.2 Curve identification

Curves on rural state highways were identified using 10m horizontal curvature data in the 'high speed' (HS) geometry table found in the NZTA's RAMM database. After a number of iterations, the following set of rules was settled on.

1 *What constitutes a curve?*
Curves are defined as consisting of at least three sequential 10m segments in the same lane that have a 30m rolling average radius less than the threshold of 500m and the sign of the radius is the same for all three segments. For simplicity, this is referred to as the curve apex (see figure G.1).

2 *Start and end points*
For a lane, the start and end of a curve is when the average radius value over three consecutive 10m readings (recorded at the middle reading; ie the 30m rolling average comprises the 10m section before, the 10m section under consideration, and the next 10m section) is greater than 800m. This takes account of the curve transition/spiral and the braking zone leading into a curve.

For the carriageway, the start point of a curve is the lane curve start location with the lower chainage, and the end point is the lane curve end location with the higher chainage.

3 *Compound and reverse curves*
If there is more than one instance within the length of the curve where condition (1) is met (ie three sequential 10m segments where the 30m rolling average radius is less than 500m and all are of the same sign), these are to be treated as part of one large curve, provided the following condition is met:

> For one of the lanes, there are no instances throughout the length of the curve where there are more than two sequential 10m segments with a 30m moving average radius greater than 800m. For simplicity, a gap is defined as being a 10m segment whose 30m moving average radius is greater than 800m. Therefore it is possible to have one or several 10m or 20m gaps in one lane of a compound or reverse curve for this condition to be met. In other words, gaps of 10m or 20m can be ignored.

Provided this condition is met, a gap of any size can be tolerated in the other lane.

A compound curve is when the sign of curvature at the apexes doesn't change throughout the curve length. A reverse curve is when there is a change in the sign of curvature between successive apexes.

4 *Start and end points of reverse curves*
The point where one curve ends and the next curve begins is defined by splitting the difference between:

i) the latest point in the first curve where the curvature is in the same direction and the 30m moving average radius is less than or equal to 800m in both lanes

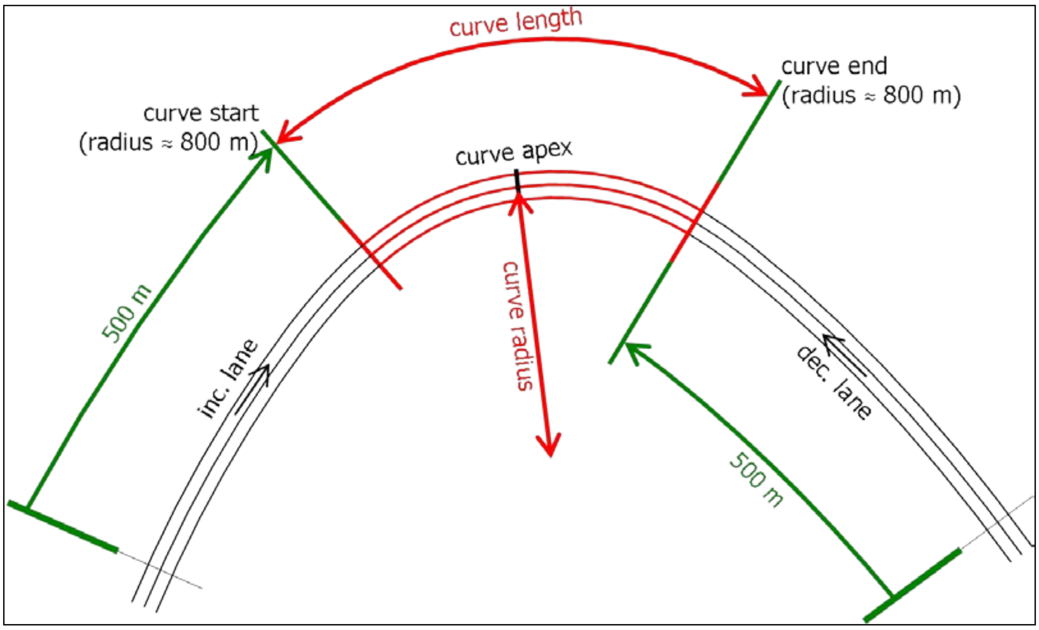
ii) the earliest point in the second curve where the curvature is in the same direction and the 30m moving average radius is less than or equal to 800m in both lanes.

If the split point is found to be in the centre of a 10m section, the displacement is rounded down; ie the split 10m section is added to the second curve.

5 *Minimum separation distance between curves*

A curve is regarded as being isolated if the length of break (ie radius greater than 800m) between curves is 20m or greater.

Figure G1 Schematic of curve



The above curve identification process (figure 7.1) was validated by comparing the derived curve extents with those manually determined from true tangent points (ie where there is the first indication of deviation from the straight approach) for curves located at SH5/RS29, SH30/RS158 and SH30/RS170. A total of 55 curves, including a number of compound and reverse curves, were used in this validation exercise. In the majority of cases, the start and end locations agreed to 20m or closer.

In this study a program was written in Matlab® (version R2007b) to automatically generate a report of rural curves with a radius of less than 500m. The program can process curves on both divided and undivided carriageways. It also flags curves whenever the location of the increasing lane apex is 40m or greater than the location of the decreasing lane apex, to indicate a possible concern with the geometry data. There is an expectation that the difference in curve location between increasing and decreasing directions should not be greater than 10–20m in the majority of cases, as surveys of road geometry made since 2009 have employed a GPS-based location-referencing system, so the start point is common for both increasing and decreasing directions.

Application of the curve identification process to New Zealand's rural state highways yields the following results:

- There are 18,771 curves with a horizontal radius of curvature less than 500m, of which about 65% (11,800) are of <250m radius.
- The combined length of <250m radius curves amounts to 1699.02km, and the curves between 250m and 500m radius, amounts to 1138.04km. This gives a total of 2837.06km, which equates to about 26% of the entire state highway network.
- The average length of a <250m radius curve is 144m, compared with 163.3m for curves between 250m and 500m radius.

G.3 Crash risk model for curves

Curves at intersections were excluded from the dataset. The statistical modelling attempted to fit the number of crashes in each curve (and in each year of the analysis period) to geometric elements of the curve, exposure (ie ADT), and the difference between the approach and the curve speeds (ie OOCC).

Previous New Zealand research (Koorey and Tate 1997) had identified that the risk and severity of crashes on curves was not only a function of absolute curve radius but also the difference between the approach speed and the curve speed. Furthermore, the crash rate was shown to increase significantly when the difference between the approach speed and curve speed exceeded 15km/h. The determination of approach and curve speeds was seen as a critical input to the statistical modelling.

The approach and curve speeds can be reasonably determined by inputting 10m radius and crossfall data from the geometry table in the NZTA's RAMM database into the 'advisory speed' formula as detailed in appendix C.

Tate and Turner (2007) tested a range of variables and among other things, identified a strong correlation between curve crashes and the difference between the approach speed over a 500m length and the minimum curve speed over a 30m length.

For the statistical modelling, the approach speed in the increasing lane was defined as the average of the advisory speeds from 500m prior to the start of the curve, to the start of the curve. The approach speed in the decreasing lane was defined as the average of the advisory speeds 500m prior to the end of the curve to the end of the curve. For the 500m lead-in of an analysis, for which there was no data in either lane, the advisory speeds were assumed to be equal to 110km/h, which was 10% above the open-road speed limit. Setting the advisory speed to the maximum expected speed of 110km/h ensures that calculated differences between approach speed and curve speed will err on the high side for situations when geometry data is not available over the entire 500m lead-in. The schematic in figure G1 shows the position of the lead-in relative to the start of the curve.

The curve speed was defined as the minimum 30m averaged advisory speed over the length of the curve. The 30m average was derived from the advisory speed calculated for the current and preceding two 10m sections.

For the rural environment, the advisory speed was capped at 110km/h, and for the urban environment it was capped at 70km/h.

An issue with the statistical modelling was the difficulty in allowing for errors in the location of the crashes. A partial solution was to assign each crash that had occurred within 50m of a curve to that curve. Where this would have resulted in a crash being assigned to two curves, it was assigned to the one nearest to the crash.

A modification of a Poisson linear/log-linear model was fitted to the data.

The modelling assumed that each side of each curve can generate crashes at the rate (per year) according to the following relationship:

$$a \times L_1 \times \exp(L_2) \qquad \text{Equation G1}$$

where a is the ADT per lane and L_1 and L_2 are linear combinations of transforms of the road characteristics as follows:

For L_1 : a constant

square root of curve length (sqrt_lengthR)

For L_2 : OOCC (ie difference between the approach and curve speeds)

curve speed (AS)

skid resistance (SCRIM)

approach gradient (gradient_app)

$\log_{10}$(ADT)

year

NZTA administration region.

Therefore, the fitted model was a combination of the linear model (the L_1 part) and the log-linear model (the L_2 part). The coefficients in the L_1 and L_2 linear combinations were the unknown parameter that had to be estimated.

The following equation was used to calculate the overall number of crashes per 100 million vehicles passing through the curve for the side of the road of interest:

$$\frac{10^8}{365} \times L_1 \times \exp(L_2) \qquad \text{Equation G2}$$

The overall personal risk associated with a particular curve was obtained by averaging the crash rate calculated from the above equation for each side of the road.

Table G.2 summarises the results of the analysis of variance of the model fit. To assist the fitting process, the 50 percentile value was subtracted from the non-categorical variable, apart from the approach gradient. Third-degree polynomial transforms were used for OOCC, curve speed and $\log_{10}$(ADT), whereas a second-degree polynomial transform was used for SCRIM skid resistance, approach gradient and square root of curve length.

Table G.2 Table of variance – 'all' casualty crashes

Term	Degrees of freedom (df)	Chi-square	
		SS(3) (term added last)	SS(1) (term added sequentially)
L_1			
poly2_((sqrt_lengthR)-15)	2	81.97	81.97
L_2			
year	5	35.89	28.45
NZTA administration region	6	61.35	89.67
poly3_(OOCC-30)	3	189.66	457.41
poly3_(AS-50)	3	28.03	14.16
poly2_(SCRIM-0.5)	2	63.43	47.63
poly3_(log10(ADT)-3)	3	35.48	35.99
poly2_(gradient_app)	2	13.81	13.81

The column SS(3) in table G.2 gives the chi-square value when the corresponding term is the last one added to the analysis, and SS(1) gives the chi-square value when the terms are included sequentially. When calculating the SS(1) values for the L_1 and L_2 terms, it was assumed that the other linear set of terms had already been fitted.

The 1% and 5% levels of significance are tabulated in table G.3. A comparison with the SS(3) values shows all the fitted variables are statistically significant at the 1% level if the Poisson model is valid. OOCC is shown to be the most significant predictor variable, followed by curve length. However, the SS terms for the curve length are underestimating its importance, probably because of the use of the L_1 and L_2 terms.

Table G.3 Levels of significance

Levels of significance	Degrees of freedom					
	1	2	3	4	5	6
5%	3.84	5.99	7.81	9.49	11.07	12.59
1%	6.63	9.21	11.34	13.28	15.09	16.81

It is likely that there is more variability in the data than the Poisson model implies, because the model does not fit perfectly. As a consequence, the SS values in table G.2 should be substantially above the critical levels in table G.3 before declaring a term to be statistically significant.

Table G.4 is the resulting table of effects. Values greater than 2 under the column headed 'Ratio' are statistically significant if the Poisson model is believed to be correct. This test should be applied only to the highest-degree term in each polynomial.

Table G.5 provides an illustrative application of the model using the following inputs:

year:	2002
region:	R2
OOCC:	30(km/h)
AS:	80(km/h)

SCRIM: 0.5(ESC)

ADT: 1000(v/d)

gradient: 0 (%)

length: 100(m).

In table G.5 the quantity being modelled is personal risk in units of 'all' casualty crashes per 100 million vehicles entering the curve, and collective risk in units of annual number of 'all' casualty crashes per curve.

In applying the model to assign skid resistance investigatory levels to curves with horizontal radius of curvature of 400m or less, a fixed SCRIM skid resistance value of 0.4, corresponding to the T10 skid site category 4 investigatory level, was used.

Table G.4 Model coefficients

Variable	Model statistics		
	Coefficient	Standard error	Ratio
L_1:			
constant	1.77E-05	1.80E-06	9.9
(sqrt(lengthR)-15.0)**1	1.61E-06	1.92E-07	8.4
(sqrt(lengthR)-15.0)**2	6.84E-09	1.21E-08	0.6
L_2:			
year:1997	0		
year:1998	-0.02352	0.062	-0.4
year:1999	0.04360	0.063	0.7
year:2000	0.02011	0.063	0.3
year:2001	0.19874	0.061	3.3
year:2002	0.25136	0.061	4.1
region:R1	0		
region:R2	0.13161	0.064	2.1
region:R3	0.38803	0.080	4.8
region:R4	0.40065	0.074	5.4
region:R5	0.28962	0.079	3.7
region:R6	0.33949	0.079	4.3
region:R7	0.43579	0.076	5.7
(OOCC-30.0)**1	0.04387	0.004	10.6
(OOCC-30.0)**2	0.00039	0.000	3.7
(OOCC-30.0)**3	-1.24E-05	4.95E-06	-2.5
(AS-50.0)**1	0.01570	0.003	4.6
(AS-50.0)**2	-9.43E-05	1.71E-04	-0.6
(AS-50.0)**3	-9.87E-07	2.65E-06	-0.4
(SCRIM-0.5)**1	-2.17050	0.273	-8.0
(SCRIM-0.5)**2	-1.14390	2.159	-0.5

Variable	Model statistics		
	Coefficient	Standard error	Ratio
(log10_ADT-3.0)**1	-0.05904	0.094	-0.6
(log10_ADT-3.0)**2	-0.17294	0.206	-0.8
(log10_ADT-3.0)**3	-0.08039	0.155	-0.5
(gradient_app)**1	-0.02628	0.008	-3.4
(gradient_app)**2	0.00035	0.001	0.4

Note:

R1 to R7 are the following NZTA administration regions: R1=Auckland; R2=Hamilton; R3=Napier; R4=Whanganui; R5=Wellington;R6=Christchurch; R7=Dunedin.

Table G.5 Example application of curve crash risk model

Variable	Value of input variable	Processed value of input variable	Model coefficient	Product (value × coefficient)
L_1:				
constant		1	1.77E-05	1.77E-05
(sqrt(lengthR)-15)**1	100	-5	1.61E-06	-8.04E-06
(sqrt(lengthR)-15)**2	100	25	6.84E-09	1.71E-07
				Σ = 9.84E-06
L_2:				
year:2002		1	0.25136	2.51E-01
region:R2		1	0.13161	1.32E-01
(OOCC-30.0)**1	30	0	0.04387	0.00E+00
(OOCC-30.0)**2	30	0	0.00039	0.00E+00
(OOCC-30.0)**3	30	0	-1.24E-05	0.00E+00
(AS-50.0)**1	80	30	0.01570	4.71E-01
(AS-50.0)**2	80	900	-9.43E-05	-8.48E-02
(AS-50.0)**3	80	27,000	-9.87E-07	-2.66E-02
(SCRIM-0.5)**1	0.5	0	-2.17050	0.00E+00
(SCRIM-0.5)**2	0.5	0	-1.14390	0.00E+00
(log10_ADT-3.0)**1	1000	0	-0.05904	0.00E+00
(log10_ADT-3.0)**2	1000	0	-0.17294	0.00E+00
(log10_ADT-3.0)**3	1000	0	-0.08039	0.00E+00
(gradient_app)**1	0	0	-0.02628	0.00E+00
(gradient_app)**2	0	0	0.00035	0.00E+00
				Σ = 7.42E-01
			Personal risk:[a]	**5.66**
			Collective risk:[b]	**0.02**

a) Personal risk is in terms of ‘all’ casualty crashes per 10^8 vehicles entering the curve.

b) Collective risk is in terms of annual number of casualty crashes per curve.

G.4 Model fit

It was difficult to carry out a goodness-of-fit test on the model. The usual chi-squared goodness-of-fit test on the actual and fitted values did not work in the present situation because of the small expected number of crashes for most curves.

One approach was to divide the data into categories based on one or two of the predictor variables and then compare the observed and modelled numbers of crashes for each of the categories. Figure G.2 shows the comparison when one categorises by length and curve speed.

Mostly, the agreement is good, but there appears to be an interaction between length and curve advisory speed for the shortest curves. The tightest curves are less dangerous than predicted by the model and the straighter ones are more dangerous.

Tests on length and OOCC, and on curve speed and OOCC, also gave good agreement.

Figure G.2 Predicted and actual crash numbers