An Exploratory Study Using Big Data for Improved Safety and Operational Efficiency: A New Zealand Case Study
An Exploratory Study Using Big Data for Improved Safety and Operational Efficiency: A New Zealand Case Study

Gareth Robins¹ and Dr. Salvador Hernandez²
Director of Analytics, EROAD, New Zealand, gareth.robins@eroad.com¹
Oregon State University, Corvallis, Oregon, USA, sal.hernandez@oregonstate.edu²

MOTIVATION
Concerns related to crashes has increased globally over the years because of the potential level of injury severity sustained by those involved and their resulting economic impacts. “This is especially true for New Zealand with over 4.9 million inhabitants and up to 3.68 million visitors each year (Stats New Zealand, 2017). The countries limited road network of 11,000 kilometers of state highways and 83,000 kilometers of local roads provides a challenge to overseas drivers that must navigate mountain passes to shared heavy light motorways (at different speeds) (Walker, 2014). Recently New Zealand has joined a long list of jurisdictions that are establishing Vision Zero initiatives and directing policy makers towards the implementation of these initiatives (Office of the Associate Minister of Transport, 2018). To achieve this goal, New Zealand is focusing on mechanisms to effect change that include changing government policy, setting sensible speed limits, reducing the tolerance for speeding during periods of high travel such as public holidays, media campaigns, and better road engineering. This is important because in 2017 there were 379 deaths on New Zealand roads (NZ Ministry of Transport, Research and Statistics, 2017a). Along with the all too high human cost, New Zealand’s Ministry of Transport estimates the social cost of road crashes and associated injuries in 2016 to be NZ$4.17 billion (NZ Ministry of Transport, Research and Statistics, 2017a). This is estimated by the amount of money that the New Zealand population would be willing to pay for safety improvements that result in the expected avoidance of one premature death. These costs will only continue to grow if nothing is done to help mitigate and reduce roadway crashes.

In New Zealand the major causes of road death on open roads are losing control, travelling too fast for the conditions, driving impaired, and failing to keep left. In regard to Urban causes, failure to give way and failure to see the other party lead the way (NZ Ministry of Transport, Research and Statistics, 2018). However, in recent years there has been a reduction in the average vehicle speed across the New Zealand road network, and the incidence of vehicles travelling over the speed limit has also seen a reduction. An analysis by Sam Warburton (2017) looked at crashes per kilometer driven in New Zealand and found that the chance of a person in a car dying on the road was 41 percent higher than it was in 2013, and 12 percent higher than 2016. Moreover, New Zealand is a popular tourist destination, with 5.6% of GDP being generated from tourism (New Zealand. Statistics New Zealand., 2016). According to the New Zealand Ministry of Transport, 4.3 % of fatal and injury crashes occurred when an overseas driver was at fault. Most of these crashes were the fault of drivers that originated from countries that also drive on the left side of the road, indicating that unfamiliarity with left hand driving was not to blame (NZ Ministry of Transport, Research and Statistics, 2017b).
With these statistics in mind, the goal of this paper is to better understand contributing factors that increase an individual's road risk. The road is a dynamic system, evolving throughout the day as the population uses it. As such the risk of an individual road can therefore be defined in this study by how fatigued, frustrated, and/or familiar the users of that road are at the time they are driving. To accomplish this, this paper utilizes telematics data collected in New Zealand that provides a unique opportunity to study human factors related to driving at an aggregated level and provides insights into how the collective population affects road risk across the day. This will be done through an exploratory statistical analysis of that telematics data and the development of risk measures for fatigue, frustration and familiarity of the road network. This data is collected in collaboration with EROAD, that have a strong presence in New Zealand, and have the critical mass of telematics devices on the roads to study this behavior and provide an additional level of risk that can be used to enhance the existing models to include population behavior. Because of the regulatory motivations behind the installation of these devices, the accuracy and polling frequency are higher than comparable sources, enabling the analyst to better understand the network corner by corner. The use of telematic data to study driving behavior is not new (Brackstone et al., 1999; Hickman and Hanowski, 2011; Wahlstrom et al., 2015; Wouters and Bos, one et al., 1999; Hickman and Hanowski, 2011; Wahlstrom et al., 2015; Wouters and Bos, 2000), however past studies have been limited by the number and quality of the data collection process. In addition, these same studies have either been cell phone based and/or instrumented vehicles (Botzer et al., 2017; Hickman and Hanowski, 2011; Itoh, 2008). The current data set used for this study is collected directly from various vehicle classes in the New Zealand and amount to a sample size of over 1 billion records (See Data Description). To the best of the authors knowledge these are the first attempts at utilizing such data to better understand fatigue, frustration and familiarity of the road network.

The following sections outline the data, provide an exploratory statistical analysis of the data to gain insights into the contributing factors related to fatigue, frustration and familiarity.
DATA DESCRIPTION

The data used for this study was generated from a GPS-GNSS enabled device installed in vehicles owned by customers of EROAD, a global regulatory telematics company. The devices are installed in a wide variety of vehicle makes (over 2800) and across 39 industry sectors. Specifically, the data was collected from 20,866 LIGHT vehicles (GVM<3,500kg) and 32,740 HEAVY vehicles (GVM>3,500kg).

The device central collects GPS and telematics data continuously, which is transmitted to a central platform approximately every 250 meters when the vehicle is in motion, and for significant events such as harsh braking, ignition on/off, and idling. The events are enriched with information about the road, obtained from HERE Maps (a map generating company). For example, the posted speed limit was abstracted, which facilitated the calculation of vehicle speed (e.g., if they were speeding), and the road functional class as described below.

Speeding events were defined when a vehicle exceeded the speed limit by 5 km/h or more. Heavy vehicles have a reduced speed limit of 90 km/h in an otherwise 100 km/h limit, therefore speeding events were counted when a HEAVY vehicle exceeded 95 km/h. To satisfy privacy concerns, none of the speeding events had information that could be used to identify the EROAD customer.

Harsh braking events were generated on the device when the change in velocity exceeds 10 km/h in 1 second. This equates to a deceleration of 2.77ms² or 0.28g.

The functional class of the road was used to summarize the results throughout the paper and ranges from class 1 roads designed for high speed, high volume traffic to class 5 roads designed as local access. As seen in Figure 1 the density of these classes increases to serve smaller regions. The road network reference being used is divided into road segments, primarily at intersections, but also at jurisdictional boundaries.

Finally, a trip is defined as the duration between a vehicle’s ignition on event and ignition off event. Outside of the scope of this study was trip chaining, whereby we would expect longer travel times when combining trips with short intervals between them, for example courier drivers delivering packages and turning the vehicle off for just a few minutes, or vehicles waiting at rail crossings and not idling.

<table>
<thead>
<tr>
<th>Road Class</th>
<th>Description</th>
<th>Count of Road Segments</th>
<th>AADT</th>
<th>Speed Limit Range (km/h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>National High Volume</td>
<td>19,354</td>
<td>&gt; 35,000</td>
<td>70 - 110</td>
</tr>
<tr>
<td>2</td>
<td>Regional, linking remote regions</td>
<td>11,811</td>
<td>&gt; 15,000</td>
<td>60 - 100</td>
</tr>
<tr>
<td>3</td>
<td>Arterial, critical connectivity</td>
<td>32,462</td>
<td>&gt; 5,000</td>
<td>50 - 100</td>
</tr>
<tr>
<td>4</td>
<td>Collector, linking between 200 and 2000 populations</td>
<td>42,420</td>
<td>&gt; 3,000</td>
<td>50 - 100</td>
</tr>
<tr>
<td>5</td>
<td>Access, all other roads</td>
<td>263,083</td>
<td>&gt; 200</td>
<td>50 - 100</td>
</tr>
</tbody>
</table>
Figure 1. Map of Road Functional Classes. Class 1 (state highways) at the left to Class 5 (local roads) at the right.

As part of the investigation into how frustration is affected by the curvature of the road, a curvature ratio is defined. This curvature ratio is defined as the ratio of the straight-line distance between either end to the contoured length (see Figure 2 and 3). Accounting for compound curves, reverse horizontal curves, or changes in vertical alignment was out of the scope of this study.

Figure 2. Calculation of curvature ratio, dividing the shortest distance by the contoured distance.
RESULTS
Below we describe the analysis in terms of three risk measures for fatigue, frustration, and familiarity. For fatigue we looked how distant each road segment is from the origin of the vehicles that visit it. Frustration was measured by analyzing how driver’s propensity to speed changes based on the curvature of the road preceding their current location. Finally, for familiarity we looked at the proximity of speed and harsh braking events to the start and stop location of a trip.

FATIGUE

Time to Arrive
According to the New Zealand Ministry of Transport, in 2016 fatigue was identified as a contributing factor in 28 fatal crashes, 119 serious injury crashes and 438 minor injury crashes. These crashes resulted in 36 deaths, 160 serious injuries and 574 minor injuries. The total social cost of crashes involving driver fatigue was about $291 million; this is about 7 percent of the social cost associated with all injury crashes (Ministry of Transport, Research and Statistics, 2017c).

For this study access to individual driver’s work plans or rosters were not obtained, so instead an estimate of fatigue based on trip time was derived. Over 9.5 million trips were studied where vehicles visited a total of 49.2 million road segments. The trips were then filtered to remove those where the amount of rest time before the trip was less than one hour. The median time it took each vehicle to reach each road segment was then calculated and this was split by the vehicle weight type.

This allowed us to show a distribution of travel times from the last sufficient rest period to the current road segment. This distribution is shown in Figure 4 below, which plots the distribution of median travel times for each road segment. We can see that HEAVY vehicles visit road segments after driving for significantly longer than LIGHT vehicles, which fits with the assumption of HEAVY vehicles undertaking long haul duties, and LIGHT vehicles being used for localized travel.
In order to create an index that represents the remoteness of the road from the drivers rest location, the median travel time between the top and bottom 5% of all travel times was normalized, and the visit counts to the segment were normalized between the top and bottom 5% of all segments. These two indexes were in the range 0 - 1, and were combined into a single index using equation 1,

$$\sqrt{\left(\frac{tti^2}{\sqrt{2}} + \frac{vci^2}{\sqrt{2}}\right)}$$

Where, $tti$ is the travel time index and $vci$ is the visit count index. Because the trips were split into those from HEAVY and LIGHT vehicles, the final step was to calculate the combined index using a weighted average between the HEAVY vehicle index and LIGHT vehicle index, the visit count was used as the weighting variable.

The road segments were split by this final index using kmeans into 5 clusters (Lloyd, 1982). Figure 5 below shows the extremes of these clusters, with the lowest risk roads on the left, located mainly in city centers, and the highest risk roads on the right. The intermediate clusters (not shown) are located on a mixture of arterial and collector roads. We can see that the riskiest roads are visually close to functional class 1 roads, which makes sense as the travel time to reach these roads is a lot higher. However, we also see that roads in urban areas score high on this index.

Figure 4. Average drive time from all vehicles to reach all segments after 60mins rest period.
New Zealand has some of the most challenging roads in the world, particularly when it comes to travel in rural areas. The topology of the land, and the layout of our major centers means that many of the functional class 1 and 2 roads, designed for high volume, high speed travel must be designed with a significant number of curves.

We describe each road segment in terms of its curvature and the effect that has on the propensity to speed. To understand this behavior we looked at each location event that was preceded by either a cornering event or another location event. Cornering events are generated when a vehicle has completed a bearing change of at least 35 degrees. With these two scenarios, corner-straight (cs) or straight-straight (ss) we plotted (Figure 6) the distribution of speeds that fell within +/- 10% of the road speed limit. We also ensured that the preceding event was on a road with the same speed limit as the subject event. The distance between the subject event and the preceding event was 250m +/- 13m. We eliminated events where the driver was seen to change their speed by more than 10km/h per second, which would have generated a harsh acceleration or harsh braking event.

Figure 5 Maps of low fatigue risk located in city centers and high fatigue risk located along state highways.

FRUSTRATION

Curvature

We describe each road segment in terms of its curvature and the effect that has on the propensity to speed. To understand this behavior we looked at each location event that was preceded by either a cornering event or another location event. Cornering events are generated when a vehicle has completed a bearing change of at least 35 degrees. With these two scenarios, corner-straight (cs) or straight-straight (ss) we plotted (Figure 6) the distribution of speeds that fell within +/- 10% of the road speed limit. We also ensured that the preceding event was on a road with the same speed limit as the subject event. The distance between the subject event and the preceding event was 250m +/- 13m. We eliminated events where the driver was seen to change their speed by more than 10km/h per second, which would have generated a harsh acceleration or harsh braking event.
We see in the graph above the two distributions for each event tuple (corner-straight and straight-straight). In all speed limits we see a greater proportion of speeds under the speed limit in corner-straight tuples. This indicates that drivers are accelerating out of corners but are more likely to remain under the speed limit. In many roads this could be because of the lead up to another corner, where exceeding the speed limit will likely result in a sharp deceleration to meeting the next corner’s advisory speed. In the straight-straight event tuple we see a larger proportion of events that exceed the speed limit than in the corner-straight event tuple. Even though the inter-event distance is 250m +/- 13m, which is sufficient to accelerate to above the speed limit from a corner, drivers are more likely to exceed the limit when they are already travelling in a straight road. This speed violation is less likely however on roads with lower speed limits, perhaps due to increased congestion and less opportunity to speed.
For HEAVY vehicles we see the largest change in speeding behavior on functional class 1 roads, where vehicles exhibit 30.5% more over-speeding in straight-straight tuples compared to corner-straight tuples. Table 2.

Table 2. Proportion of HEAVY vehicle over-speeding when following a curved or straight road

<table>
<thead>
<tr>
<th>Road Class</th>
<th>Proportion of Overspeed in Corner to Straight Event Tuple</th>
<th>Proportion of Overspeed in Straight to Straight Event Tuple</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. National</td>
<td>0.323</td>
<td>0.422</td>
</tr>
<tr>
<td>2. Regional</td>
<td>0.282</td>
<td>0.364</td>
</tr>
<tr>
<td>3. Arterial</td>
<td>0.273</td>
<td>0.333</td>
</tr>
<tr>
<td>4. Collector</td>
<td>0.270</td>
<td>0.346</td>
</tr>
<tr>
<td>5. Access</td>
<td>0.256</td>
<td>0.326</td>
</tr>
</tbody>
</table>

Figure 7. Proportion of overspeed events from HEAVY vehicles that follow a straight road or curve, split by road class. Proportions are of all events between -10% and +10% of speed limit.
For LIGHT vehicles we see a marginally higher incidence of speeding when exiting a corner on functional class 5 (4.1%), class 4 (1.1%) and class 3 roads (2.1%). This could be due to the faster acceleration profiles of LIGHT vehicles on local roads. As with the HEAVY vehicle data, we also see an increase in the propensity to speed in functional class 1 roads (10.8%) roads when vehicles are already travelling on a straight road. This could be due to overtaking lanes. The results are presented in Figure 8.

Table 3. Proportion of LIGHT vehicle over-speeding when following a curved or straight road

<table>
<thead>
<tr>
<th>Road Class</th>
<th>Proportion of Overspeed in Corner to Straight Event Tuple</th>
<th>Proportion of Overspeed in Straight to Straight Event Tuple</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. National</td>
<td>0.295</td>
<td>0.327</td>
</tr>
<tr>
<td>2. Regional</td>
<td>0.285</td>
<td>0.299</td>
</tr>
<tr>
<td>3. Arterial</td>
<td>0.309</td>
<td>0.302</td>
</tr>
<tr>
<td>4. Collector</td>
<td>0.354</td>
<td>0.350</td>
</tr>
<tr>
<td>5. Access</td>
<td>0.323</td>
<td>0.310</td>
</tr>
</tbody>
</table>

Figure 8. Proportion of overspeed events from LIGHT vehicles that follow a straight road or curve, split by road class. Proportions are of all events between -10% and + 10% of speed limit.
To translate this back to an individual road's likelihood of facilitating speeding events, we use equation 2 on each road segment.

$$\sqrt{(ct_i^2 + vs_i^2)}/\sqrt{(2)} \quad (2)$$

c is the curvature of the road as defined above, s is the reverse of c indicating the amount of straightness, PLSS & PHSS are the proportions of overspeed for LIGHT & HEAVY vehicles respectively in a straight-straight tuple, PLCS & PHCS are the proportions of overspeed for LIGHT & HEAVY vehicles in a corner-corner tuple.

This produced a range of scores in the domain -1.369 to 10.439. We truncated this distribution to the upper and lower 1%, and coerced values that lay outside to these limits. Finally, this domain of -1.369 to 8.429 was normalized to the range 0 to 1 by equation 3.

$$\log(1+(LV \cdot c \cdot PLSS + HV \cdot c \cdot PHSS + LV \cdot c \cdot PLCS + HV \cdot c \cdot PHCS + LV))$$

where x was the score and p01 was the 1% percentile of the score, and p99 was the 99% percentile of the score.

![Distribution of Road Segment Frustration Indexes](image)

**Figure 9.** Distribution of frustration indexes from equation 1, normalised with equation 2.
FAMILIARITY

Proximity of harsh driving to destination

It is widely accepted that crashes are more likely to occur when a vehicle is close to their home location. This was studied by Burdett et al., (2017). The study used the New Zealand Household Travel Survey and found that roads within 11 km (6.8 miles) of home accounted for half of all travel and 62% of all crashes. The reason behind many of these close-to-home crashes involve a lapse in attention. The study also suggests that local roads are over represented in the dataset.

For this study, we also looked at the proximity of harsh braking events to the final destination of their respective trip. The difference in approach to the Burdett et al., (2017) paper is that we are able to focus on harsh braking and speeding as precursors to a crash. The authors measured the straight line distance between the crash and the center of the driver’s home suburb, we are measuring the exact distance from the location of the poor driving event, and the final destination of the driver. Given the vast difference in the shape of suburbs, this distance is much more accurate. Lastly, we are able to compare the results of harsh driving to destination distances where that final destination is a common location for the driver vs any location.

We studied a sample of 1.9 million trips that showed 610,000 harsh braking events. Of these trips, 71.32% terminated on local roads (functional class 5), and 48.26% of harsh braking occurred on these roads. On the other functional classes, harsh braking is more represented than trip terminations, indicating this behavior happens during a trip.

We calculated the distance between the location of the harsh braking event and the final destination and normalized this over the trip distance. We found that 50% of harsh braking events occurred within 8.6km of the trip’s final destination. Furthermore 90.4% of harsh braking events occurred in the last 50% of their trip. When looking at the distribution of event to stopped location distances, normalized to the trip distance, we find that fewer harsh braking events occur at the start of the trip, and more occur at the end of the trip.

We saw no relationship between the intensity of the harsh braking event and the proximity to the origin or destination.

Figure 10. Distribution of harsh braking events along trips
Speeding

The same 1.9 million trips were also analyzed for speed events. When normalized across the trip, the distribution of speeds to the start and stop position is different to that of harsh braking. We see that most speeding occurs in the middle of the trip, after a vehicle is approximately 10% of the way through its trip. Speeding again drops off as the vehicle approaches its destination. This is understandable given the large number of trips that terminate on functional class 5 roads with lower speed limits, however we have seen in the curvature analysis that there are pockets of speeding occurring after corners on these road classes.

Figure 11. Distribution of speeding events along trips

To create the index that represents the propensity to speed when near the final destination we first describe each road segment by the average location along the length of all trips that encounter it. The density curves above were used to obtain the probability of speeding or harsh braking for a given location along the trip. This was repeated for HEAVY and LIGHT vehicles for both harsh braking and speeding curves. The result is that road segments that are consistently close to the destination of the trips that intersect them will show a high-risk of harsh braking, and a lower risk of speeding. Conversely, segments that are close to the origin of the trips that intersect them will show a lower likelihood of both harsh braking and speeding. The four resulting indexes where then simply normalized between the lower and upper 1% of values, and the mean was taken to reduce into a single familiarity index.
COMBINING INDEXES TOGETHER

The last part of the analysis was to combine the three indexes for frustration, familiarity, and fatigue into a single index that can be used to make routing decisions on. We chose to simply take the mean in this case. The resulting distribution shows that functional class 5 roads show a consistently lower risk index than functional class 1. Road classes 1, 2, and 3 show a multi-modal distribution which is driven by varying speed limits, whereas class 4 and 5 speed limits are more consistent.

Figure 12. Distribution of final risk scores by road class
MAPPING ROUTES

The resulting index was split into 5 clusters using kmeans and plotted on a map. Figure 13 shows an example of a route choice between Taupo and Rotorua. The left-hand route is 107km, whereas the right-hand route is 81.4km according to Google. We calculated the cumulative risk index for each route by taking the weighted average of the risk indexes for each road segment, weighted by the road segment distance. This is because this current model cannot assess where in the road segment the vehicle is subjected to the calculated risk, and the lengths of the road segments vary substantially.

Of the two route choices in figure 13, the left-hand route shows a weighted mean index of 0.408 whereas the right hand route shows a weighted mean index of 0.595. Thus, the right hand route has 23.9% less distance, but 45.8% higher risk.

a) Varying Drive Times  

b) Varying Risk

Figure 13. Route choice between Taupo and Rotorua showing varying drive times, right image shows the varying risk on the two routes.
SUMMARY AND CONCLUSIONS

This paper set out to describe the road network in terms of the risk imposed on individual drivers by others using the network. We explored ideas such as changing shift schedules as a measure of fatigue, the number of times a driver visited a particular road segment as a measure of familiarity, and the driver’s adherence to the speed limit on the various roads they encounter across their trips. We settled on three metrics that demonstrated a strong correlation with harsh braking and speeding, and attempted to convert these measures into a single metric that could be used to make improved routing decisions on the network. This example clearly shows that the shortest route, that would normally be suggested by most consumer routing applications, exposes the vehicle to a higher risk than the alternative route.

Optimized routing offers a competitive advantage for many transport operators, as margins are waning, they are looking for more ways to increase efficiency, and reduce wear and tear on the vehicle. Consumer grade routing applications work to minimize a distance-based penalty between an origin and destination. Some more advanced algorithms will also include other factors such as topology, intersections, congestion, and special use roads. We believe this is the first implementation that offers a population risk penalty to also optimize for. As a next step we will be implementing this algorithm over a directed graph of road segments and exploring how the risk model changes across the time of day or day of week. We will also automate the calculation of the various risk indexes to increase the sample size from one billion records to the full EROAD dataset.

The potential for EROAD customers is that we can dynamically route their vehicles as we see a change in the network ahead of their trip and offer drivers a safer alternative to reach their destination on time. Having an optimized route model also offers EROAD customer’s a competitive advantage in terms of their obligations under the Health and Safety Act 2015. Moreover, we believe we are in a position to better inform the design of static factors such as the design of infrastructure and dynamic factors such as choice of route, and signage to soften the negative consequences of situational factors.

EROAD’s mission is to solve complex transportation problems and the goal of this research project was to showcase options that assist in achieving Vision Zero. It will further give government agencies a new layer of information for evidence-based spending on infrastructure funding.
REFERENCES
Office of the Associate Minister of Transport, 2018. IMPROVING ROAD SAFETY IN NEW ZEALAND.

Gareth Robins is the Director of Analytics at EROAD, a global provider of regulatory telematics. His work is focussed on using GPS data to identify areas of potential risk on the network, visualise traffic and commodity flows to assist planners, and produce network usage and reliability metrics to assist transport authorities. Gareth has over 20 years of experience in the analytics industry, having worked across three continents.

Dr. Salvador Hernandez research derives knowledge from social sciences and computational sciences to create meaningful solutions for problems in Transportation Safety, Freight Logistics, Transportation Systems Modeling, and Disaster Management. He blends the development of new science-based approaches with practical applications and implementation.

Dr. Hernandez methodological research interests and expertise are in the areas of statistical & econometric and mathematical modeling, including discrete choice analysis, and network flow analysis & optimization. Dr. Hernandez utilizes interdisciplinary modeling techniques to address issues at the intersection of multiple disciplines.

ABOUT EROAD
ERoad modernizes road charging and compliance for road transport by replacing paper-based systems with easy-to-use electronic systems. The company is headquartered in Auckland, New Zealand, and listed on the New Zealand Exchange (NZX). Its US business is based in Portland, Oregon. EROAD is also a leading provider of health and safety compliance services, including vehicle management and driver behaviour and performance measures.