Combining accelerometer data and contextual variables to evaluate the risk of driver behaviour[☆]

Johan W. Joubert[a,1,*], Dirk de Beer[b,2], Nico de Koker[a,3]

[a]*Centre of Transport Development, Department of Industrial and Systems Engineering, University of Pretoria, Private Bag X20, Hatfield, 0023.*
[b]*Digicore Technologies, Regency Office Park, 9 Regency Drive, Route 21 Corporate Park, Irene X30, Centurion, South Africa.*

Abstract

Telemetry devices are generating and transferring increasingly more data, with notable potential for decision makers. In this paper we consider the accelerometer and speed data produced by in-vehicle data recorders as a proxy for driver behaviour. Instead of extracting harsh events to cope with the large volumes of data, we discretise the data into a tractable and finite *risk space*. This novel methodology allows us to track both acceptable and non-acceptable driving behaviour, and calculate a more comprehensive risk model using the envelope of the data, and not a priori thresholds. We show how thresholds suggested in literature can characterise some driving behaviour as good, even though our empirical evidence has not even registered such extreme driving behaviour.

We demonstrate the model using accelerometer data from 124 vehicles over a one month period. Three rules, each a combination of accelerometer and/or speed data, are applied to the risk space to derive person-specific scores that are comparable among the individuals. The results show that the scoring is useful to identify specific risk groups. The proposed model is also dynamic in that it dynamically adjusts to the observed records, instead of data having to abide by a limited model specification.

Keywords: Risk profiling; Driver behaviour; Accelerometer; Speeding

1. Introduction

Road accidents have a large social and financial cost to any economy. Not only the direct cost in terms of injury and loss of economic productivity, but also the associated cost of the insured occupants and vehicles. Unfortunately accidents are primarily attributed to driver behaviour. To study *safety*, Musicant et al. [12] argue that we can use extreme driving events such as harsh braking and acceleration, improper turning and swerving as surrogates for safety: a high frequency of undesirable events is a direct indication of risky and unsafe driving.

As mobile sensing platforms become more readily available and affordable, so the volume of detailed data from the physical world also increases. Such data is useful for decision making. One particular application is the insurance sector where telematic data is used to assess and ultimately mitigate the insured risk [16]. If insurers want to base decision-making and risk profiling on driver behaviour models, such models must withstand scientific scrutiny.

Rodríguez González et al. [21] consider *how* driver behaviour data can be obtained, and note three approaches: self-reporting by drivers; driving simulators and from measuring more naturalistic driving conditions. The latter is advantageous as it allows behaviour to be observed and studied while the driver behaves 'normally'. That is, not as if for the sake of a specific and controlled experiment. Following from Wu et al. [27] there are two ways to identify risky driver behaviour in naturalistic settings. Firstly there are physiological signals of the driver vigilance that includes fatigue, distraction and driving under toxic influence. To identify these signals intrusive sensors are often required, or nonintrusive cameras from which driver characteristics can be inferred. In this project we consider the second way by studying the behaviour of the vehicles. For this approach we consider vehicles fitted with in-vehicle data recorders (IVDRs).

The benefit of using IVDRs is that it is a reliable and objective source of driving behaviour and vehicle usage data [21, 25]. As a technology, it is inexpensive and provides continuous measurement of on-road driving behaviour that is otherwise difficult to observe. Compared to handheld devices, as Paefgen et al. [16] note, it provides a more accurate detection of extreme driving events as it is pre- or auto-calibrated based on the device pose (orientation). Drivers appear to display safer behaviour when they are (knowingly) monitored and when feedback is provided [9], although such considered behaviour is not always sustained without reinforcement [24].

A simplistic approach to measure driver risk is to calculate the total number of extreme, or harsh events over a specified period, with or without a weighting for each event type. The driver with the highest number of harsh events is the most risky. Consequently, a driver's risk is only determined by *bad* behaviour. Desyllas and Sako [4] review *pay-as-you-drive* business models where the level of risk is weighted by usage (for example distance driven, speed and location of use) and traditional underwriting considerations (for example driver age, gender and marital status). Insurers cluster the drivers using their respective risk profiles, and can then provide incentives, through for example risk-based premiums, in an effort to curb and discourage risky driving. In this paper we propose a more comprehensive approach to measure driver risk. Instead of only looking at harsh events, we propose to aggregate the continuous IVDR data into a multi-dimensional *risk space*. The contribution is methodologically novel as it is, to our knowledge, the first attempt to provide a richer description of (technical) driver risk from the perspective of the insurer.

The paper is structured as follows: the next section reviews the recent literature on how harsh driving events are used to infer driver risk. In Section 3 the proposed risk space is introduced, and implemented in section 4 via a numerical example of 124 drivers for whom IVDR data was recorded during August 2014. The subsequent discussion explores how different risk-evaluation rules can influence the driver scoring.

2. Literature

Musicant et al. [12] acknowledge the vast amounts of data that IVDRs generate, and use this argument to motivate why much of the research is directed towards identifying only extreme events from the raw data. The result is much less information to process in evaluating a particular driver. The extreme driver behaviour events are mainly derived from the US Patent by Raz et al.

[20], which is the technology in the *Green-Box* by Greenroad Driving Technologies Ltd. In these approaches preset thresholds are considered, and whenever the accelerometer reading exceeds the thresholds, a *harsh event* is recorded.

In this review we distinguish between studies that are mainly from an academic research perspective, and those that are industry-driven.

2.1. Research

There are studies that mainly focus on studying driving behaviour, and do so mainly from a research perspective that will most likely not be state-of-practice. Consider for example the massively sensorised *Argos* system that can measure a driver's point of gaze, distance to lateral road markings and can generate and record different in-car light and audio stimuli. Such systems hold much opportunity for research as an experiment supervisor can interact directly, or schedule specific actions to take place during an orchestrated experiment. However, such levels of vehicular telematics is unlikely to have a large footprint in the coming years.

The field of driver behaviour modelling is new, and still evolving. Much of the literature relates to driving event recognition. The study by Johnson and Trivedi [9] combines the data from multiple sensors like accelerometers, geospatial positions system (GPS) and gyroscopes, often referred to as *sensor fusion*, and are shown to identify specific manoeuvres very accurately.

Regarding devices of choice, there seems to be two schools of thought: (fixed) in-vehicle data recorders, and smart phone-based technologies. Díaz Álvarez et al. [5] claim that IVDR is not that freely available and hence argue for the use of mobile devices, while others like Paefgen et al. [17] argue that a variety of affordable commercial offerings are indeed available, and also, that smartphones seem to overestimate the number and severity of harsh events.

Driver behaviour and errors are known to be a major cause of vehicle crashes, especially for fleet vehicles [25]. But evidence regarding whether event data recorders (EDRs) influence crash occurrences is mixed.

According to Paefgen et al. [16] consumers do not favour so-called *pay-as-you-drive* and *pay-per-risk* policies. Added to consumer resistance, the telemetry incurs additional costs to insurers and, ultimately, customers. They argue that smartphones are a viable alternative as they are operated at the users' discretion. However, following a controlled field study they note that the smartphone devices tend to overestimate critical driving events when compared to fixed devices that were calibrated based on the initial device pose.

The most common approach for profiling driver behaviour is to extract critical driving events, for example harsh braking, swerving, etc. Such events are then considered violations of thresholds imposed on vehicle acceleration. In studying the use of smartphones, Paefgen et al. [16], only concern themselves with event counts, that is, the number of occurrences that the preset thresholds were violated. Different contributions in literature uses different thresholds. In this paper we will express acceleration using the non-standard International System of Units (SI) measure of *standard gravity*, $g_n = 9.81m/s^2$, that is the mean acceleration due to to gravity at the Earth's surface. In the remainder of this paper we will denote standard gravity with g, and '*milli-g*' with mg. Paefgen et al. [16] use longitudinal thresholds $a_x < -0.1g$ for harsh braking, $a_x > 0.1g$ for harsh acceleration, and lateral acceleration of $|a_y| > 0.2$ for swerving, while Baldwin et al. [2] use $0.15g$ in all directions. Johnson and Trivedi [9] use both longitudinal and lateral acceleration, but only report the lateral thresholds. For aggressive driving, they distinguished between left and right turns ($0.73g$), swerving ($0.74g$), and U-turns ($0.91g$). Their results show that 97% of events can be accurately identified. Bergasa et al. [3] use accelerometer thresholds of $0.4g$ in all directions

3

to indicate aggressive driver behaviour, and calculate a driver score using the count and intensity (three levels) of these events.

A number of contributions have studied what underlying factors and contextual variables are good predictors of whether a person will register harsh driving events. Age is a determining factor. More precisely, novice (young) drivers are over-represented in vehicle crashes and injuries, especially during their first year of unsupervised driving [18]. They also note that males are more likely to take driving risks.

Musicant et al. [12] note that temporal factors influence the number of harsh driving events. They identified both the time of day, and day of week as playing a significant role in predicting the number of harsh driving events. They indicate that time of day could be reduced to *daytime*, 06:00–24:00, and *nighttime*, 00:00–06:00. According to their research, gender too plays a significant role with men being more prone to perform harsh driving events.

In the United States, the 100-car naturalistic driving study is an ongoing instrumented-vehicle study undertaken with the primary purpose of collecting large-scale (100 vehicles), naturalistic driving data. Neale et al. [14] introduced the proposed study design early in the project while Dingus et al. [6] report in detail on the subsequent field study. From the telematics data they created an event database that includes *crashes*, *near-crashes*, and *other incidents*. The study did not (to our knowledge) consider the occurrence of *non-events*. That is, the effect of drivers behaving well, and the proportion of driver time that is considered non-risky.

One approach is to identify k number of driving classes, and then use the classical k-means clustering approach to group the event data into k classes, and score the individuals accordingly [28]. Wu et al. [27] use a *hidden Markov model* (HMM), a statistical Markov model in which the states are not directly known, but the output variables (driving events) that are dependent on the states (driver's risk attributes) are visible. They used seven behavioural events, and also applied low, medium and high intensities for those events. Their experimental results show that they have a detection ratio of harsh events of between 98.2% and 100%.

Much of the recent work, like Rodríguez González et al. [21], aim to provide real time feedback to drivers to warn them, and hopefully influence their behaviour positively. Farah et al. [7] demonstrate how family feedback can improve driving behaviour beyond the target group of young novice drivers. Although some literature [13, 24, 25] state that initial positive impact of feedback unfortunately diminishes over time if not reinforced, we are not aware of longitudinal studies of driver feedback. Such feedback systems often do create even *more* distractions for the driver to deal with. The research effort, though, still moves us closer to future intelligent transport systems in which advanced driver assistance systems will play an increasing role.

Musicant et al. [13] recognise the heterogeneity among individuals, but note that it is often dealt with by considering subgroups. The downside is that each subgroup is still characterised by its *average driver*. In their contribution, they do have an interesting and novel approach to fit a distribution to the counts of undesirable events at the individual level. They studied the relationship between event counts and trip duration, the magnitude of changes in driver behaviour over time and how rapidly drivers change their behaviour. They confirmed earlier observations reported in Musicant et al. [12] that there is an inflation of event counts at the beginning and at the end of the trips, regardless of the trip lengths. Possible explanations stated in their paper include that trip ends are often *near home* or other familiar places where non-defensive driving occurs more frequently. Also, trip ends are more likely to occur in urban areas where there is more potential for harsh driving events like sharp turning and harsh braking at traffic signals.

Quantifying the risk associated with driver behaviour is not merely an academic exercise to study and better understand behaviour. Much of the research finds its way to, and is applied by insurers who aim to mitigate their risk of insuring drivers and their vehicles. In this section we review a number of these commercial applications.

Paefgen et al. [17] provide apt justification for the use of IVDR in industry. Referring to the naturalistic driving study of Neale et al. [14] and Dingus et al. [6], they note that the per-vehicle cost of data acquisition can easily exceed USD 20,000. As a viable (transitional) alternative to dedicated research solutions they recommend collaborating with vehicle fleets equipped with IVDR for commercial purposes. The market penetration of commercial IVDRs, across various sectors, are high enough to reward researchers with rich data and large sample sizes at a low cost. Malm and Fagerberg [10] estimate that the original equipment manufacturers alone have shipped 11-million units by 2014 with an estimated 20.5-million subscribers using embedded telematics.

In 2004 Progressive Corporation piloted their *TripSense*® usage-based insurance program in Minnesota to research driving habits [19]. In 2008 they introduced *MyRate*^{SM}, a pricing scheme providing a driver-based insurance rate based on driving performance (the programme is optional). Using Progressive as a case study, Desyllas and Sako [4] bring the business model and intellectual property protection perspective, arguing that pay-as-you drive placed Progressive in front of their competitors and the market as a whole.

Although literature mentions the *SAGA* system developed in Iceland, a reproducible citation could not be found. The SAGA system uses event data recorders to monitor and report location and usage of vehicles, speed (compared to the speed limit) and driving behaviour according to predefined criteria. The system was mainly aimed at vehicle fleets and was used on 70 company fleets. In the Global Road Safety Partnership [8] *speed manual*, the Iceland case study is said to have achieved a 56% reduction in crash cost; 43% reduction in the total number of crashes, and 51% reduction in the number of crashes in which the employees were responsible.

Toledo et al. [25] calculate risk indices as indicators for the likelihood of crash involvement, and these correlate well with past crash observations. Also, they observe a significant reduction in crash rates after IVDRs were installed. They measured acceleration (40 Hz), speed and position and using these to detect a range of 20 vehicle manoeuvres that are considered risky events. The manoeuvres were further classified by direction (left or right), as well as three levels of severity. An individual driver risk index is calculated as a linear function of the numbers and severity of the various types of manoeuvres that the driver has performed. The risk index is then normalised using the drive time of the individual. Their experiment is based on 191 fleet vehicles: compact pickup trucks/light delivery vehicles from a single company.

The *Aviva Drive* app measures a (voluntary) user's acceleration, braking, and cornering characteristics over 300 kilometres (200 miles) to evaluate the driving behaviour. This is done using smart phone technology and allows the user to choose which journey(s) to evaluate. The inputs are then used to provide a personalised score for the driver on a 0–10 scale, allowing for as much as 20% discount on premiums [1].

In the United Kingdom, the Co-operative Insurance Group has a *Young Driver Insurance* product that uses a *Smartbox* to collect *safer driving parameters*: average speeds on different types of road; harsh acceleration and braking; time of day; and harsh cornering. Each parameter is scored on a 1–5 scale to compute an overall score. Although they indicate how the overall score will influence your premium, they do not divulge what thresholds they use for the different

parameters [23].

The Italian company Octo Telematics uses their *Octo Clearbox*, an onboard device that varies from smart phones, self install solutions, to semi and full professional onboard devices [15]. The company indicates that they monitor journey types, distance travelled, and driving style, but do not provide additional information.

Other commercial applications include the recent *DriveSafe* app [3] and the *In-Drive* program, introduced by StateFarm [22], that offer savings based on driver behaviour. They specifically mention harsh braking and acceleration, turning, time of day and level of speeding (exceeding 80 miles per hour).

Should we invest so much in detecting extreme events? We infer from literature that a disproportionate amount of attention is given to the study and use of harsh driving events. McGwin and Brown [11] report that in nearly 50% of the studied cases an accident was preceded by some particular driving event, yet in only 6% of the cases was it the primary cause of the crash. Harsh braking will quite plausibly precede a crash as a driver attempts an avoidance manoeuvre. The last thing the driver should hear is some driver assistance system saying in a soothing voice: *"please don't break so hard, Joe, that's risky, and may influence your premium..."* Drivers with frequent harsh braking events may in fact be more inattentive, which in turn may make them more accident prone if given untimely feedback. It would appear that as researchers we forgot that correlation do not necessarily imply causation.

3. Methodology

The majority of research contributions identify behaviour at the individual level. In this project we want to use the full data set to find an overall view of a pool of drivers, and rate each individual relative to all the others. As a result, we avoid imposing *a priori* thresholds of what constitutes *good* and *bad* behaviour, but rather use the envelope of all the available data to determine relative performance. This is in line with the field of *Data Envelopment Analysis*, a non-parametric method in operations research and decision science where one can empirically measure the productive efficiency (driver risk) of decision-making units (drivers).

The data used in this paper was obtained from three-axes accelerometers installed in the vehicles of 124 randomly drawn individuals who subscribe to Digicore's existing *Ctrack* service offering. Like Rodríguez González et al. [21] we choose sensors that are already available in vehicles. Our objective is to make our model results transferable to the insurance market and real applications. Raw accelerometer data is sampled at $4mg$ resolution at frequencies of at least $50\,\text{Hz}$. The data stream is then passed through an orientation matrix, transforming the raw data stream to the vehicle orientation before being pushed through Kalman filters. The result is a $5\,\text{Hz}$ sampled data set with an expected data accuracy and repeatability between vehicles of $25mg$.

We postulate that we can represent complex driving behaviour with a simple model based on the statistics of external driving signals. In effect, what we want to do is provide a measure of *exposure* to high risk resulting from driver behaviour. Paefgen et al. [17] noted that, in the context of transportation research, exposure is often interpreted as the accumulated mileage of a vehicle or the vehicle's duration of driving. Effectively we want to see how many accelerometer observation points do we observe for different drivers in different areas of the risk space. In this section we introduce an aggregation scheme that addresses both the extent and the degree of exposure as this has indeed been identified as a dominant research area to pursue.

3.1. Discretising the acceleration space for risk

Our approach in this paper is different from that followed in previous work. Instead of extracting only harsh events, we aggregate all the accelerometer records by discretising the three-dimensional accelerometer space. The three dimensions represent longitudinal (x), lateral (y) and vertical (z) acceleration. To do this we bin our data to a space filling, face-centred close packed (FCC) three-dimensional arrangement providing a more efficient space filling geometry than a simple cubic arrangement. Grid points in this arrangement are delimited by rhombic dodecahedra (the primitive cell of an FCC lattice). The scale of each cell, w, is a user input into the tessellation and denotes the distance from the centroid of the cell to the furthest corner point (of which there are six).

From each accelerometer record we use the unique anonymised identification number for the customer vehicle; the time stamp of the record; the three-axis acceleration namely longitudinal (x), lateral (y) and vertical (z); and the derived speed. The speed estimation is done on-board using the embedded Geographical Positioning System (GPS) unit. The three-dimensional tessellation is populated by associating each record with the polyhedral cell that encloses the record's (x, y, z) coordinate. After processing all the records there is a set of C cells, denoted by $\boldsymbol{C} = \{1, \ldots, C\}$, that has one or more records associated with it. The number of records associated with cell $i \in \boldsymbol{C}$ is denoted by r_i. This results in a significant computational saving as we only have to keep record of r_i, hence $|\boldsymbol{C}|$ values. So, like extracting harsh events are trying to reduce the computational burden, discretising the risk space achieves a similar outcome. However, our discretisation provides a richness of the records over a time period longer than just observing harsh events. This is because we account for good driving as well as bad driving.

Consider the month of August 2014 for which there was a total of 51.43-million records. Table 1 shows summary statistics for the r_i counts when using different cell scales. In the three-dimensional risk space the result of the discretisation is determining how many records *fall within* each cell. Consider the first row, for example. If the accelerometer space is discretised with cells that measure

Table 1: Summary statistics for aggregating accelerometer records as a function of the cell scale.

Cell scale	Number of	Percentile of record counts (r_i)						
(*mg*)	cells ($|\boldsymbol{C}|$)	0.25	0.50	0.75	0.90	0.95	0.99	$\max\limits_{i \in C} r_i$
10	44 117	2	7	59	463	1 541	19 216	1 114 196
20	8 833	2	11	136	1 602	6 146	83 915	4 928 796
30	3 420	2	15	219	3 002	13 985	227 919	5 740 776
40	1 730	3	19	344	6 419	27 248	342 454	9 107 871
50	1 035	3	21	535	9 316	41 014	515 031	21 212 745
75	409	4	42	1001	29 294	191 977	3 252 093	15 344 942
100	224	6	44	1583	50 899	152 715	2 018 825	38 040 692

$10mg$ from its centroid to the sharp point edge, the risk space contains 44,117 cells with one or more records associated with it. Of all the cells, 50% will have 7 or fewer records associated with it, and 90% will have 463 or fewer records associated with it. The cell with the highest r_i value has more than 1.1-million records associated with it.

3.2. Profiling the risk space

To classify the risk space, an a priori set of quantiles, $\boldsymbol{Q}$, is required that denotes the upper limits of the various risk classes. For the purpose of illustration in this paper, we used an arbitrary four-stage risk classification with $|\boldsymbol{Q}| = 4$ such that $\boldsymbol{Q} = \left\{ q^{\text{none}}, q^{\text{low}}, q^{\text{med}}, q^{\text{high}} \right\}$ with $q^{\text{high}} = 1.0$.

In populating the risk space we aim to calculate r_i for each cell i and denote the total number of records with $R = \sum_{i=1}^{C} r_i$. The cells are sorted in decreasing order based on r_i so that $r_1 \geq r_2 \geq \ldots \geq r_C$, where ties are broken arbitrarily. Next find the value a so that the cumulative number of records $\left[\sum_{i=1}^{a} r_i \right]/R \leq q^{\text{none}}$, and $\left[\sum_{i=1}^{a+1} r_i \right]/R > q^{\text{none}}$. The interpretation of this step is as follows: cells $i \in \boldsymbol{C}^{\text{none}} = \{1 \ldots a\}, \boldsymbol{C}^{\text{none}} \subseteq \boldsymbol{C}$ accounts for a fraction of just less or equal to q^{none} of all records and are considered *no risk* cells. All record within these cells are considered risk free. Are they truly without risk in an absolute sense? That we do not know, but they do represent the cells with the *most common* accelerometer characteristics.

For the next risk class we find b such that $\left[\sum_{i=1}^{b} r_i \right]/R \leq q^{\text{low}}$, and $\left[\sum_{i=1}^{b+1} r_i \right]/R > q^{\text{low}}$ and identify cells $i \in \boldsymbol{C}^{\text{low}} = \{a+1, a+2, \ldots, b\}, \boldsymbol{C}^{\text{low}} \subseteq \boldsymbol{C}$ as having *low risk*.

Similarly, we find c such that $\left[\sum_{i=1}^{c} r_i \right]/R \leq q^{\text{med}}$, and $\left[\sum_{i=1}^{c+1} r_i \right]/R > q^{\text{med}}$ and identify cells $i \in \boldsymbol{C}^{\text{med}} = b+1, b+2, \ldots, c\}, \boldsymbol{C}^{\text{med}} \subseteq \boldsymbol{C}$ as having *medium risk*. The remaining cells $i \in \boldsymbol{C}^{\text{high}} = \{c+1, c+2, \ldots, C\}, \boldsymbol{C}^{\text{high}} \subseteq \boldsymbol{C}$ are considered *high risk* cells.

It is therefore up to the decision-maker, for example the insurer, to determine the a priori quantile set $\boldsymbol{Q}$. In the next section we apply the risk profile and evaluate its sensitivity to different risk assumptions, and demonstrate how different quantile sets influence the risk profiles.

4. Results and discussion

The proposed discretisation of the risk space is much more flexible in that it can be tailored to different risk appetites, and accommodate a variety of contextual variables. We start the result section by showing the effect that different user parameters have on the risk space. We then report the driver risk profiles for three sets of rules.

4.1. Varying risk quantiles

The effect that different risk quantiles, $\boldsymbol{Q}$, have on the profiling of the risk space is illustrated in Figure 1. All records were considered in the risk space based on their accelerometer characteristics as described in the previous section.

The overwhelming majority of accelerometer records fall within the no risk area (coloured green), close to the centroid of the risk space. The model is quite sensitive to the chosen quantiles, as is seen in the varying thicknesses of the four shading bands as one moves out from the core to the periphery.

In this paper we opt for the risk profile illustrated in Figure 1b where $\boldsymbol{Q} = \{0.960, 0.994, 0.999, 1.000\}$. Our choice is empirically based simply on a visual balance between the four risk categories. In the rest of the paper we consistently use these risk quantiles unless otherwise specified.

The choice of quantiles is left to the insurer to tailor to its risk profile requirements, and is beyond the scope of this manuscript. Later in the paper we describe how the risk categories are translated into person-specific scores, and it is only at that point where the insurer needs to carefully consider the score thresholds that may influence pricing dynamics.
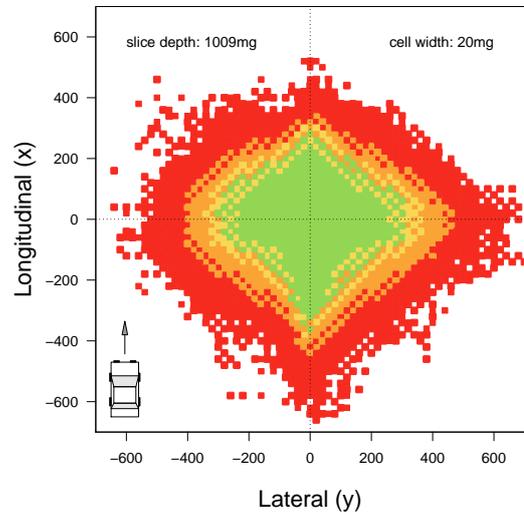
8
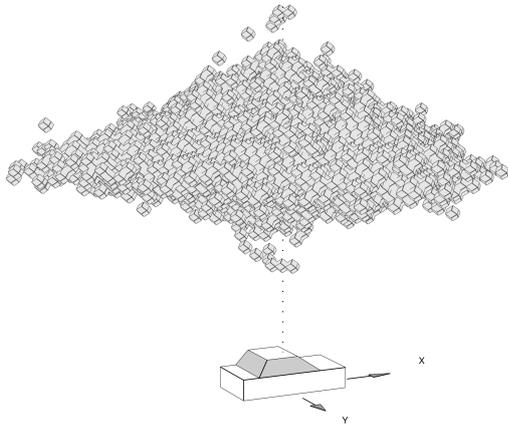
(a) Three-dimensional risk space showing all polyhedral cells with one or more records associated with it.
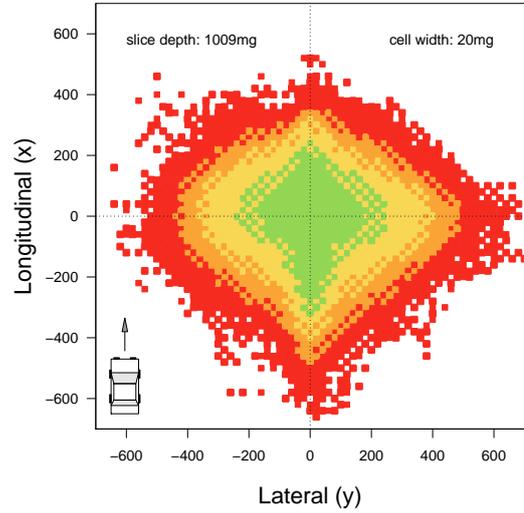
(b) $Q = \{0.960, 0.994, 0.999, 1.000\}$
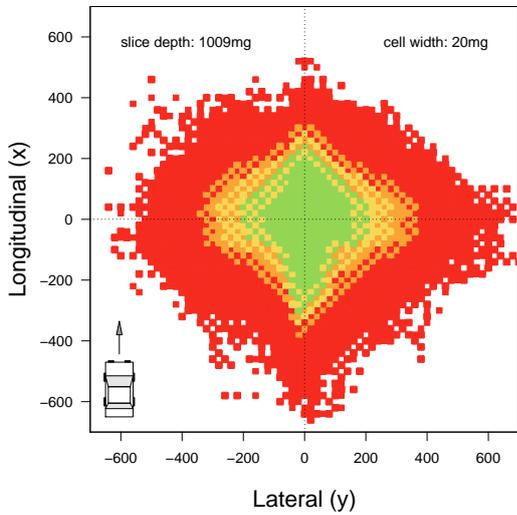
(c) $Q = \{0.950, 0.975, 0.990, 1.000\}$

(d) $Q = \{0.980, 0.990, 0.998, 1.000\}$

Figure 1: Discretised risk space showing the effect of using different risk quantiles, Q. The uncategorised, three-dimensional risk space is shown in (a), while (b)–(d) shows a horizontal slice through the polyhedral body at height $z = 1009$mg. The green indicates cells with no risk, while the yellow, orange and red represent low, medium and high risk cells, respectively.
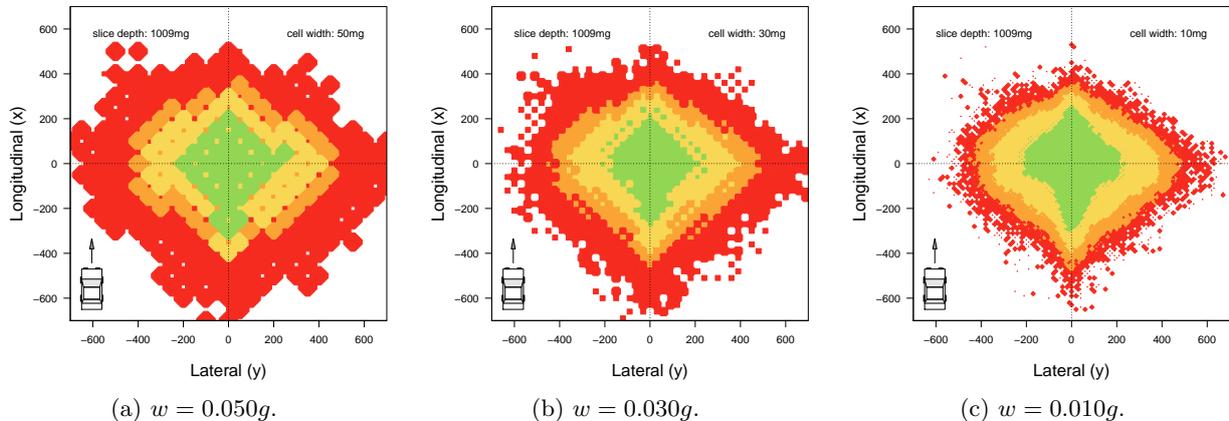
Figure 2: The influence of cell width, w, on the granularity of the risk space. All figures were populated with the same records and shows a horizontal slice through the polyhedral body at height $z = 1009$mg. The white indicates cells with no risk, while the increasing darker shades represent low, medium and high risk cells.

4.2. Granularity

The results reported in Table 1 are clarified visually in Figure 2 to illustrate how the cell scale influence the discretisation of the risk space. With fined granularity, a clover-shape for risk categories emerges. The elongation along the axes is quite plausible: hard breaking ($x \ll 0$) and acceleration ($x \gg 0$) typically happens when the vehicle is moving in a fairly straight line, that is, with little lateral acceleration ($y \approx 0$). Similarly, lateral acceleration is more common, and arguably also *safer* at constant speed ($x \approx 0$), and usually occur *after* any harsh braking has been completed.

Compare this with the fixed, rectangular accelerometer thresholds that are reported in literature. Figure 3 shows the suggested thresholds proposed by a number of authors.

Recall from the earlier review that Baldwin et al. [2] suggested $0.15g$ thresholds in all directions. This is indicated in Figure 3 by the dotted square. Compared to this study, their suggested thresholds mainly cover the no-risk area. Paefgen et al. [16] also take a conservative approach in suggesting their thresholds: $0.10g$ and $0.2g$ in the longitudinal and lateral directions respectively. This is denoted by the inner solid rectangle in Figure 3. We agree with their higher thresholds in the lateral direction as we have observed similar phenomena in this study.

The thresholds of $0.4g$ in all directions suggested by Bergasa et al. [3] is indicated by the outer solid square in Figure 3. It is especially the corner points that we challenge. Using fixed thresholds implies that a manoeuvre in which, for example, a driver simultaneously breaks hard ($y = -350mg$) and swerves to the left ($x = -350mg$) is considered *acceptable* and within limits. Yet, such a manoeuvre is so extreme that we do not observe it in our study, and is arguably *very* dangerous.

Changing the cell scale to allow for fined granularity therefore provide more realistic risk assessment. The four corners at the extremities of the traditional fixed-thresholds risk squares are consequently categorised more realistically as high risk.

4.3. Contextual variables

The accelerometer records used in this paper include contextual variables. Of interest in this manuscript is the speed of the vehicle for each record, a unique (anonymised) person identifier, and
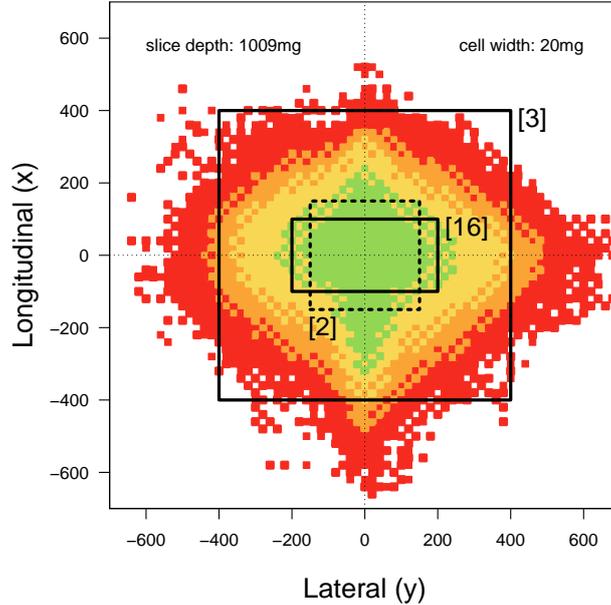
Figure 3: Comparing the multi-level, clover-shaped risk profile with fixed thresholds suggested in literature. The inner solid rectangle was proposed by Paefgen et al. [16]; the dotted rectangle by Baldwin et al. [2]; and the outer rectangle by Bergasa et al. [3].

the road type. One can generate different risk spaces for each available contextual variable, or for different levels of a variable. To illustrate this, consider the examples given in Figure 4. As a base for comparison, Figure 4a shows the risk profile when considering all records.

If one builds the risk space from records associated with freeway driving, one expects higher speeds, and less erratic manoeuvring. This expectation is indeed in line with the observations captured in Figure 4b. In a similar manner, as shown in Figure 4c, one can build the risk space for all road types, but focusing on only those records with speeds exceeding 90 km/h (56mph). Why would this be of interest to, for example, an insurer? It allows them to have a richer description and quantification of risk for different road types.

Since each record is associated with a specific individual, it is also possible to generate a unique risk space for each individual. In the examples shown in Figures 4d–4f each risk profile only contains the records of a particular individual, and the risk quantiles are applied to each individual, independent of their performance relative to other drivers. This allows the insured individual to measure him/herself over a period of time.

Much of the current research provides person-specific or vehicle-specific scores based only on an individual's own driving behaviour. These scores may have been influenced by the parameterisation following the calibration based on a cohort of drivers, yet is still based on risk thresholds set *a priori*. We argue that from an insurer's point of view there is much benefit to compare an entire pool of drivers relative to one another, and as a result base the thresholds on the envelope of empirical driver data the insurer have on their books.
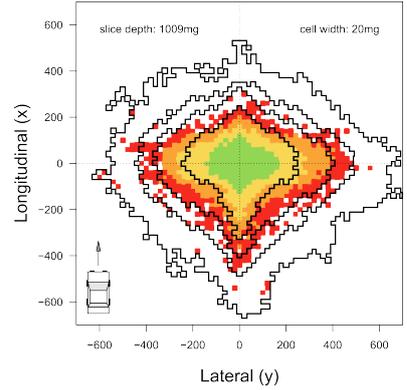
4.4. Driver scoring

To calculate the individual score of person i based on the performance of the population of drivers, $\boldsymbol{P} = \{1, \ldots, P\}$ where $P = |\boldsymbol{P}| = 124$ in this paper, we explore three rules. The objective

11
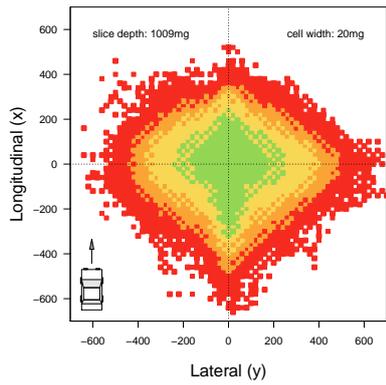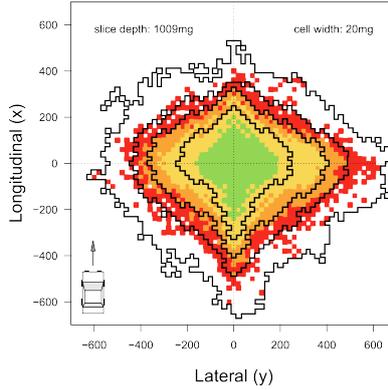
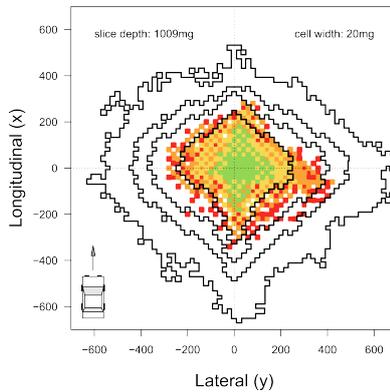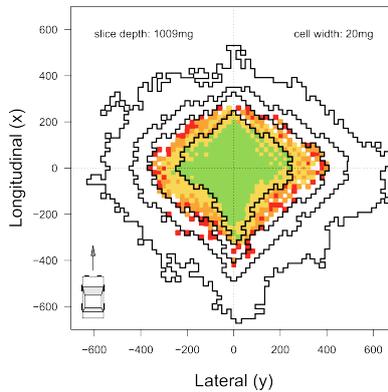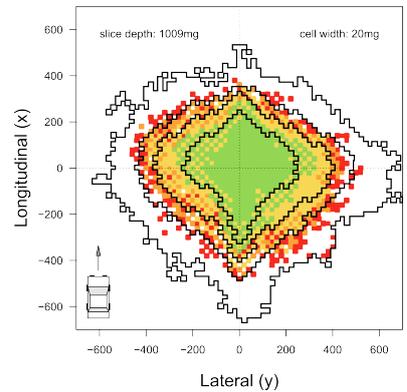(a) All records.　(b) Freeways.　(c) Speed exceeding 90km/h≈56mph.

(d) Person A.　(e) Person B.　(f) Person C.

Figure 4: Different risk spaces conditional to contextual variables provided. Each figure's risk space was populated with the subset of records that contained the specific contextual variable and shows a horizontal slice through the polyhedral body at height $z = 1009mg$. The green indicates cells with no risk, while the yellow, orange and red represent low, medium and high risk cells, respectively. As a reference, the overall risk boundaries from (a) are provided.
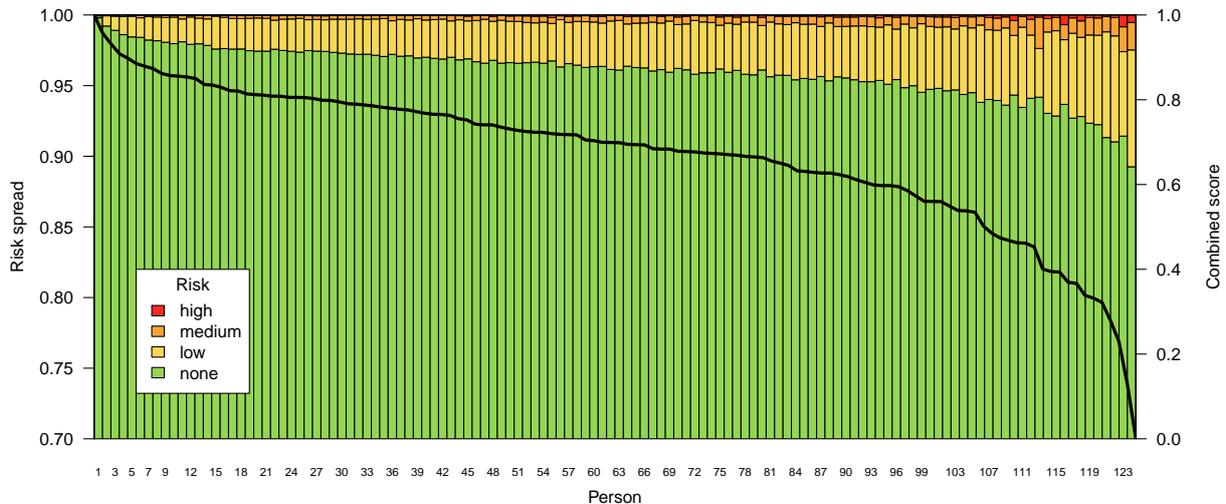
12

Figure 5: Individual scores when combining all risk categories and considering only accelerometer measurements.

is to show that the discretisation of the risk space is again flexible enough so that an insurer can calibrate it to a set of rules that best describe its risk appetite.

4.4.1. Rule 1: Acceleration only

The first rule serves as the basis and is described in Section 3. Under this rule, all records are evaluated based *only* on the measured vehicle acceleration. Once the entire risk space is discretised and all records for all individuals are accounted for in one of the cells, we cumulatively apply the risk quantiles until each cell is assigned a risk category.

Each individual's records are subsequently evaluated against the populated risk space and we calculate the proportion of records each individual has in each of the four risk categories. We denote these proportions by p^{none}, p^{low}, p^{med} and p^{high} respectively.

To combine the risk categories into a single score per person, each risk category is weighted. For the purpose of illustration we choose a simplistic weighting scheme where $w^{none} = 0$ denotes the weight for the no risk category, and similarly $w^{low} = 1$, $w^{med} = 2$ and $w^{high} = 3$. The score for each person i, denoted by s_i, is then calculated using

$$s_i = 3 - \left[p^{none}w^{none} + p^{low}w^{low} + p^{med}w^{med} + p^{high}w^{high} \right] \qquad \forall i \in \boldsymbol{P} \qquad (1)$$

We rank the individuals based on the combined score, and normalise the scores to the range [0,1]. That is, the worst performing individual's score is rescaled to 0, while the highest performing individual's score is rescaled to 1. Figure 5 shows the resulting graph.

Each bar represents an individual and shows the *risk spread* on the left y-axis, namely the proportion of the total number of observed records in each of the risk classes. The thick black line represents the combined score, and its units are shown on the right-hand side y-axis. For later reference, individuals were identified by numbers 1 through 124.

For example, individual 1 has close to 100% of all its observed accelerometer records in cells in the risk space that were considered *no risk* (green). On the other extreme we see individual 124 who had only 89% of its records in no risk cells, 9% in low risk cells (yellow), 1.5% in medium risk

13

cells (orange), 0.5% in high risk cells (red). The score for individual 124 is calculated as

$$s_{124} = 3 - [0.89(0) + 0.9(1) + 0.015(2) + 0.005(3)] = 2.055 \tag{2}$$

Since this is the lowest combined score achieved in the population, it is normalised to 0.0.

4.4.2. Rule 2: Speeding preempted

So far we have not considered speed, or more specifically *speeding*, in any of the results. One of the contextual variables available to us was the vehicle's speed at the time of the record, as well as the speed limit for the road segment associated with the record.

In this second rule we preempt populating the risk space by first considering the level of speeding. If the driver's speed for the record was within the speed limit, then we assumed *no* speeding risk. An arbitrary threshold of 10% above the ruling speed limit is set to represent a *low* speeding risk, and speeds up to 20% above the ruling speed limit are considered *medium* risk. All records exceeding the speed limit by 20% or more are considered *high* risk. The choice of 10% is based on the tolerance used by law enforcement in South Africa. Of the observed records, 88.1% reflected no speeding risk; 6.6% low risk; 2.6% medium risk; and the balance of 2.7% reflected high risk. Consequently, approximately 95% of records were within the *tolerated* speed limit. We do note that some road segments did not have speed limits set, and visual inspection of a random sample of these instances revealed that such road segment are often not classified and not part of the official road network. Examples include large open parking areas and on-site roads at industrial facilities. In such cases we conservatively consider all these records as having *no* speeding risk.

After the records are filtered based on speed, only *no*-risk records are used to populate the risk space. The populated risk space is subsequently categorised in the same cumulative way using the risk quantiles.

Similar to the acceleration-only rule, we next evaluate each of the individuals' records. If a record's speed was above the ruling speed limit, the record is categorised according to the speeding thresholds. Alternatively, the record's risk is associated with the risk category of the cell in the populated risk space. The result is again an indication of the proportion of records that each individual had in each of the four risk classes.

Using (1) we calculated a combined score for each individual; scaled the scores to the range [0,1]; and ranked the individuals from best to worst. The results are shown in Figure 6.

There is a noticeably higher proportion of records, across the board, that are considered higher risk than when only looking at the accelerometer data. The ranking of individuals have also changed. Person '2' has dropped significantly, approximately three quarters of the way down.

4.4.3. Rule 3: Combined speed and acceleration

To find a balance between the first two rules, we combine the accelerometer and speed risk dimensions in this third rule, and we do this by associating each record with both an accelerometer and a speed risk category. We use the same weighting scheme, $w^{none} = 0$, $w^{low} = 1$, $w^{med} = 2$ and $w^{high} = 3$ to assign a numeric value to each record for both accelerometer *and* speed risk dimensions.

A single risk value is then calculated for each record by weighing the two risk dimensions. Figure 7 illustrates how different weights for the two risk dimensions impact the final score for a record.

We see that when speed weighs 10%, actually anything strictly *less than* 25%, it does not affect the scoring and acceleration alone determines the final score. The result is therefore the same
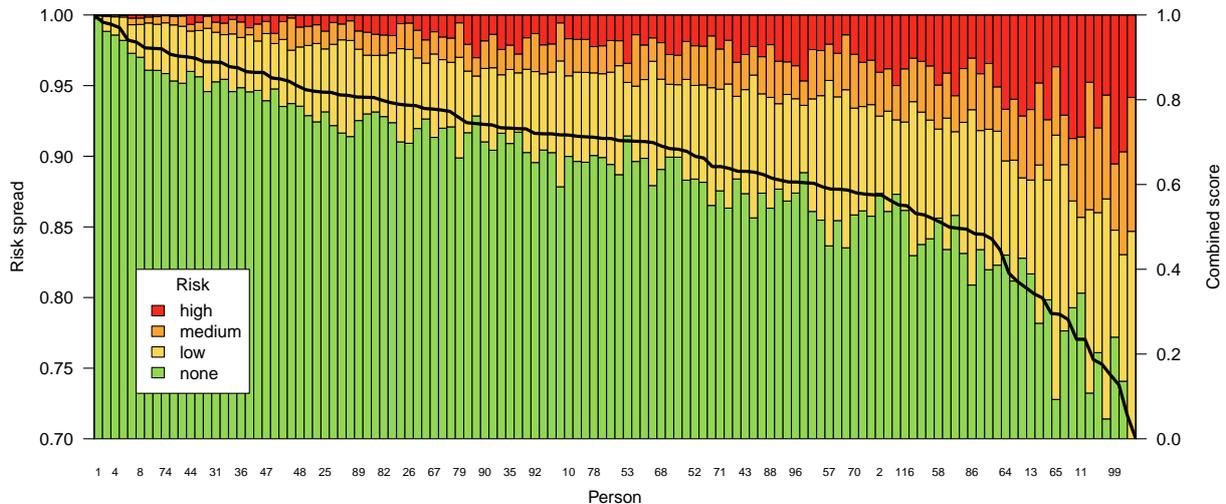
Figure 6: Individual scores when combining all risk categories and first considering the level of speeding before considering the accelerometer measurements.

as applying rule 1. Conversely, if speed weighs more than 75%, as is the case in Figure 7f, then speed alone determines the final risk category of the record. This is different from rule 2 in which preempting speed implies that no-risk speed records can still be assigned a low, medium or high risk category based on the acceleration characteristics.

Of interest in this third rule is for an insurer to consider a customisable weighting scheme that reflects their preferred trade-off between speed and acceleration. Suggesting a reasonable weighting scheme is beyond the scope of this paper.

For illustration purposes we conveniently choose the 25%:75% weighting for acceleration and speed (Figure 7e). Using (1) we calculated a combined score for each individual; normalised the scores to the range [0,1], and ranked the individuals from best to worst. The results are shown in Figure 8 and demonstrates a trade-off between the first two rules in having more exposure to low and medium risk than when only considering acceleration (rule 1), yet less high risk exposure than when preempting speed (rule 2).

4.5. Discussion

So how do the different scoring rules impact the 124 individuals considered in our study?

4.5.1. Differences between rules

First we compare rules 1 and 3. Under each of these two schemes each individual is ranked based on their combined score, which was calculated using (1). In Figure 9 we show how individual rank positions have changed when using three different weighting schemes for rule 3.

In the left diagram each point represents an individual, with the x-value denoting the person's ranking under rule 1, and the y-value denoting the person's ranking under rule 3. Points directly on the diagonal imply individuals whose ranking has stayed constant. Points above the diagonal represent those individuals who have *gained* in ranking as a result of rule 1, while points below the line are those who have gained as a result of rule 3. The histogram on the right shows the distribution of changes in ranking: negative values mean people have dropped in rankings as a result of rule 3. That is, we've deducted the ranking under rule 3 from the ranking under rule 1.

15

Risk category

	0	1	2	3

0	0.75	1.50	2.25	3

Combined risk score
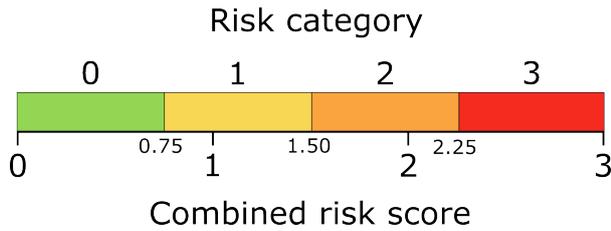
(a) Converting the combined, weighted risk score to a discrete risk category.

Acceleration 90%

Speed 10%		0	1	2	3
	0	0.00	0.90	1.80	2.70
	1	0.10	1.00	1.90	2.80
	2	0.20	1.10	2.00	2.90
	3	0.30	1.20	2.10	3.00

(b)

Acceleration 75%

Speed 25%		0	1	2	3
	0	0.00	0.75	1.50	2.25
	1	0.25	1.00	1.75	2.50
	2	0.50	1.25	2.00	2.75
	3	0.75	1.50	2.25	3.00

(c)

Acceleration 50%

Speed 50%		0	1	2	3
	0	0.00	0.50	1.00	1.50
	1	0.50	1.00	1.50	2.00
	2	1.00	1.50	2.00	2.50
	3	1.50	2.00	2.50	3.00

(d)

Acceleration 25%

Speed 75%		0	1	2	3
	0	0.00	0.25	0.50	0.75
	1	0.75	1.00	1.25	1.50
	2	1.50	1.75	2.00	2.25
	3	2.25	2.50	2.75	3.00

(e)

Acceleration 10%

Speed 90%		0	1	2	3
	0	0.00	0.10	0.20	0.30
	1	0.90	1.00	1.10	1.20
	2	1.80	1.90	2.00	2.10
	3	2.70	2.80	2.90	3.00

(f)

Figure 7: Each record is associated with two risk categories: one with speed, and another with acceleration. The two categories are traded off one another to find a single risk category to associate with the record. Weighing the two dimensions differently results in different trade-offs. In tables (b)–(f) each row represents the risk category of the record when considering speed, and each column represents the risk category of the record when considering acceleration. The cell value indicates the combined risk score using the specific weighting, and the cell colour represents the final risk category associated with the record. The conversion from the combined risk score to a discrete risk category is shown in (a)
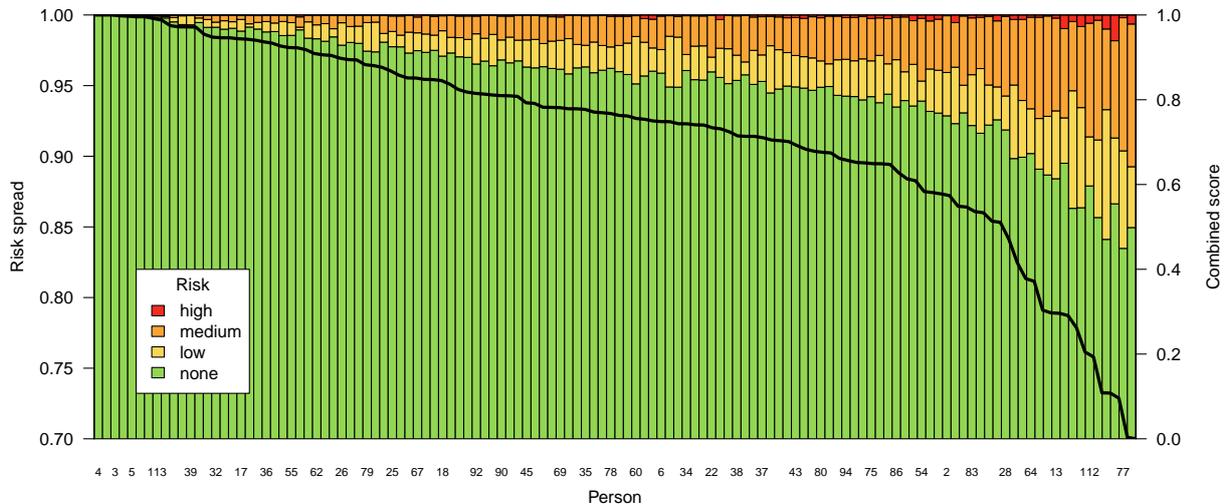
Figure 8: Individual scores when combining all risk categories and considering the level of speeding (weighing 75%) and acceleration (weighing 25%).

As expected, the spread around the diagonal reduces as the acceleration component's weight increase under rule 3.

In Figure 10 we compared rules 2 and 3, and used the same three weighting schemes. What was expected in this case is that the spread increased as the acceleration component's weighting was increased from 25% (Figure 10a) to 75% (Figure 10c). However, what was not expected is that the spread actually *decreases* when the acceleration component's weight increased to 50% (Figure 10b).

Answering the question *"are the distribution of winners and losers centred around zero?"* allows us to identify if some of our rules and rule configurations are, over the entire population of individuals, balancing the winners and losers, or if it is skewed towards producing either more winners or losers. To answer the question we did a number of t-tests for which we report the results in Table 2.

In each case the null hypothesis was that the mean of the distribution of winners and losers was centred around zero, i.e. $\mu = 0$. The alternative hypothesis is simply that it is *not*, i.e. $\mu \neq 0$. In all the cases we accept the null hypothesis. So even if there are many individuals who experience large changes in their ranking positions, seen over the entire population of individuals, the distribution of rankings is still centred around zero. This is expected. The earlier observation that the 50:50 weight results in the smallest spread is confirmed with the 95% confidence interval being the smallest at ± 3.59.

Table 3 shows the result for similar t-tests, but here we used changes in the combined score as opposed to the changes in ranking. Contrary to the ranking results, the majority of null hypotheses are rejected: the 95% confidence interval only includes the zero mean in one instance. With the 95% confidence interval mainly having negative values, the rejection of the null hypotheses indicate that the scores under rule 3 is consistently lower than under rules 1 & 2.
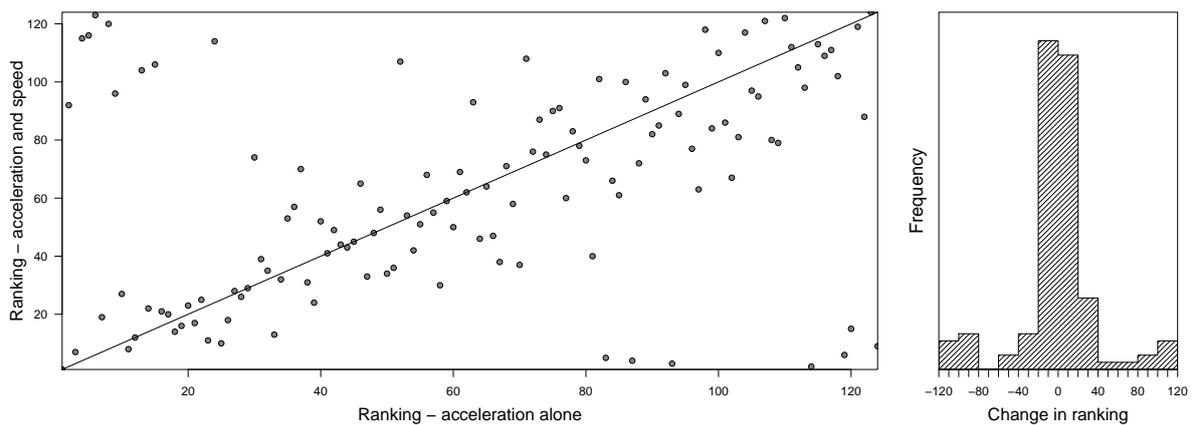
4.5.2. Comparing fixed thresholds

This section reports on our attempt to compare the proposed model with the three fixed-threshold studies in literature that were highlighted in Figure 3. We start by summarising, in
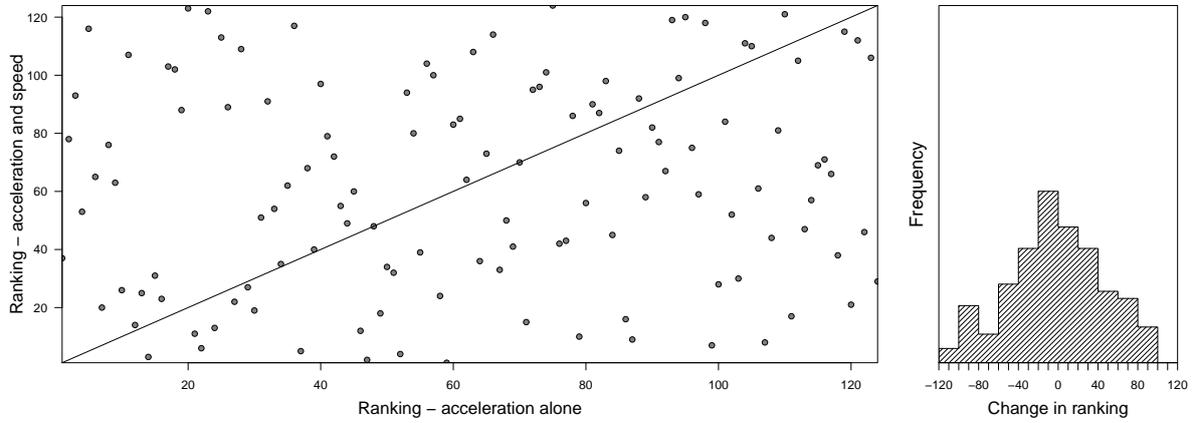
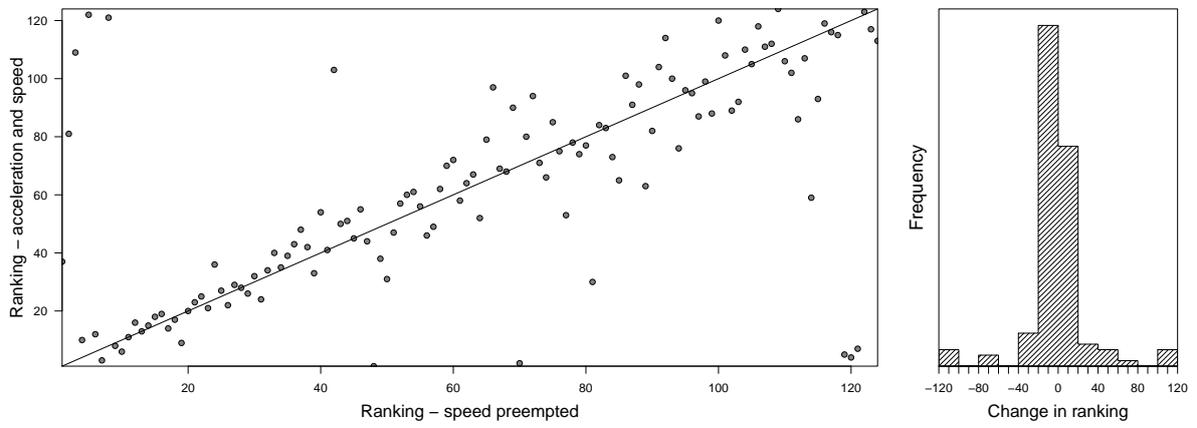(a) Acceleration: 25%; Speed: 75%.
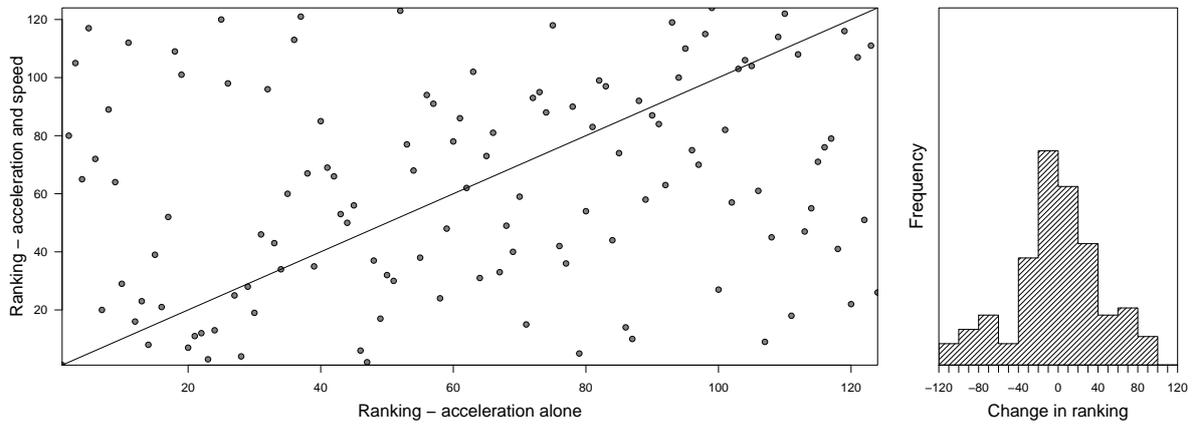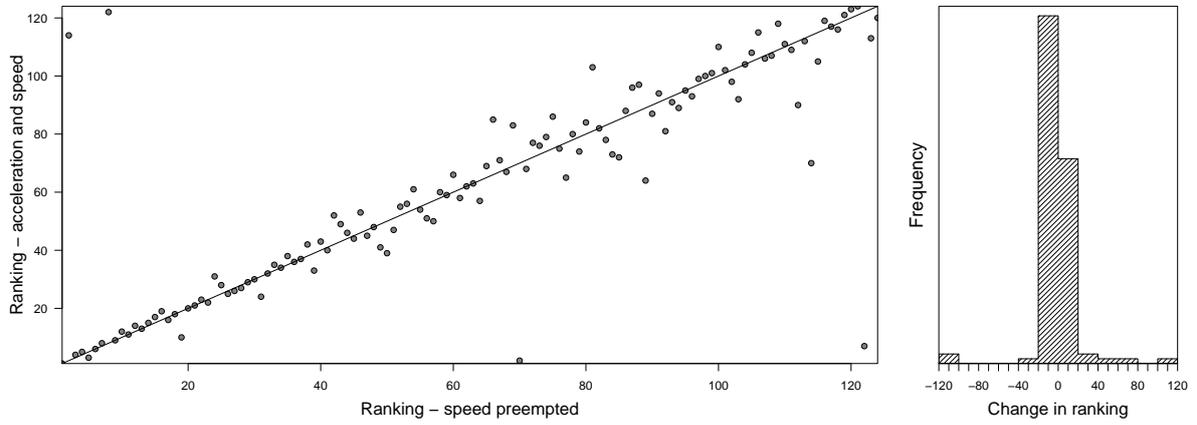

(b) Acceleration: 50%; Speed: 50%.
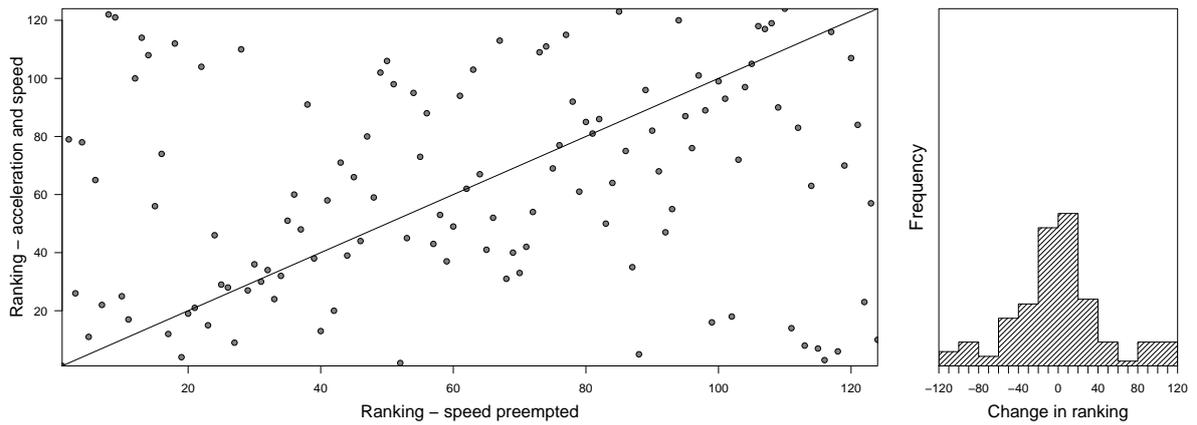

(c) Acceleration: 75%; Speed: 25%.

Figure 9: Comparing the change in rankings when comparing rule 1 to rule 3 with three different weight-combinations.

(a) Acceleration: 25%; Speed: 75%.

(b) Acceleration: 50%; Speed: 50%.

(c) Acceleration: 75%; Speed: 25%.

Figure 10: Comparing the change in rankings when comparing rule 2 to rule 3 with three different weight-combinations.

Table 2: Results of t-tests checking if the distribution of winners and losers (using the ranking) are centred around mean zero.

Rules compared	Weights (Rule 3) acceleration	speed	Statistics μ	σ	95% Confidence interval	Reject H_0
1 & 3	25%	75%	0.0000	47.9526	[-8.5234; 8.5235]	✕
	50%	50%	0.0000	44.4968	[-7.9097; 7.9097]	✕
	75%	25%	0.0000	39.4121	[-7.0059; 7.0059]	✕
2 & 3	25%	75%	0.0000	30.1379	[-5.3573; 5.3573]	✕
	50%	50%	0.0000	20.1938	[-3.5896; 3.5896]	✕
	75%	25%	0.0000	45.9836	[-8.1740; 8.1740]	✕

Table 3: Results of t-tests checking if the distribution of winners and losers (using the combined score) are centred around mean zero.

Rules compared	Weights (Rule 3) acceleration	speed	Statistics μ	σ	95% Confidence interval	Reject H_0
1 & 3	25%	75%	-0.0498	0.2308	[-0.0909; -0.0088]	✓
	50%	50%	-0.0282	0.2103	[-0.0656; 0.0091]	✕
	75%	25%	-0.1776	0.0806	[-0.1920; -0.1633]	✓
2 & 3	25%	75%	-0.0601	0.0556	[-0.0700; -0.0503]	✓
	50%	50%	-0.0385	0.0384	[-0.0454; -0.0317]	✓
	75%	25%	-0.1879	0.1611	[-0.2166; -0.1879]	✓

Table 4, the number of harsh events per kilometre travelled for the cohort of 124 drivers. For each

Table 4: Summary statistics of the number of harsh events per kilometre travelled when using different thresholds taken from literature.

Source	μ	σ	Percentile 0.25	0.50	0.75	0.90	0.95	0.99
Baldwin et al. [2]	14.6	92.1	3.5	5.0	7.6	9.2	10.1	104.4
Bergasa et al. [3]	0.04	0.14	0.0	0.0	0.0	0.1	0.1	0.8
Paefgen et al. [16]	28.2	184.8	6.2	8.7	13.3	15.8	18.1	190.9

driver we counted the number of records that fall outside the given threshold, considering each instance as a harsh event, and divided the total by the estimated kilometres travelled.

As expected, the high threshold values suggested by Bergasa et al. [3] result in very few harsh events being detected, at least in our observed data set. The South African data seems quite high when compared with the Swiss data reported by Paefgen et al. [16]. The large deviation can be attributed to two factors. Firstly, the Swiss data was based on a controlled study with a single, pre-defined route. Secondly, South Africa has a *much* higher road traffic death rate of 25.1 per 100,000 inhabitants compared to Switzerland's 3.3 deaths per 100,000 inhabitants [26].

More importantly, we are interested in seeing how the rank position of the individuals change under the different thresholds, and specifically when compared to our proposed model. We compare, in Figure 11, the threshold ranks with our *Rule 3* model using an acceleration weight of 75% and a speeding weight of 25%. In all three cases the change in ranking histograms follow a unimodal distribution. The deviation, however, is quite large. This can partially be attributed to the inclusion of speeding in our proposed model. Also, with thresholds there are only two options, risky or non-risky, while our quantile definition allows for a more progressive scaling of the risk, albeit customisable.

4.5.3. Impact of quantile set

Earlier in the paper we noted that the model is sensitive to the choice of the quantile set. What is the *preferred* quantile set? That will be application specific and will indeed be up to the insurer to determine based on their client base. Looking at the combined score lines in Figures 5–8 one can see a sharp downturn at around the 80th percentile of the individuals, towards the right-end of the graphs. An insurer could consider such a turning point to identify where dynamic pricing bands should be established. Similarly, on the upper end of the score scale (left-end of the graph), after the first 10 customers there is a flattening out, a more linear and constantly decreasing shape to the graph. An insurer could use this analysis to identify what proportion of high-scoring customers should be subsidised by the lower-scoring tail.

We argue in this paper that comparing and scoring your customer base against the envelope of performance is more useful than comparing them against some fixed threshold set a priori. With *useful* we imply that it can be used rigorously to influence insurance pricing. In suggesting that, we do acknowledge that pricing an insurance product for low-performing clients higher may indeed push such clients away, making the envelope against which one evaluate a moving target. A driver's insurance premium may therefore not only be dependent on his or her own driving performance, but also on other drivers, each of whom may be aiming to improve their own driving scores.

Yet it is in these dynamics that our proposed discretisation and an individual's evaluation hold much potential. A premium dependent on the dynamics and evolution of other drivers may just be carrot (or stick) that drivers need to maintain on a trajectory of improved driving behaviour.

Vehicle dynamics and performance are constantly changing. Combined with updating your risk space on a rolling horizon, it may indeed be quite plausible to assume that the risk space will adjust dynamically to newly introduced vehicular technologies.

5. Conclusion

One consequence of higher frequency telemetry data is the large amount of raw data generated by vehicles. Instead of dealing with the data by merely looking at the quantity of harsh driving events, we propose an alternative approach that uses all accelerometer records, and we do this for two reasons. Firstly because many accidents are shown to *not* be preceded by some harsh driving manoeuvre. Secondly, and more importantly, the proposed discretisation approach allows us to deal with not only the *unwanted* driving behaviour and use it as a proverbial stick, but also with *wanted*, acceptable behaviour that can act as a motivational carrot. The second reason is achieving the same as when one weighs the absolute number of unwanted events by the total distance travelled for a particular driver.
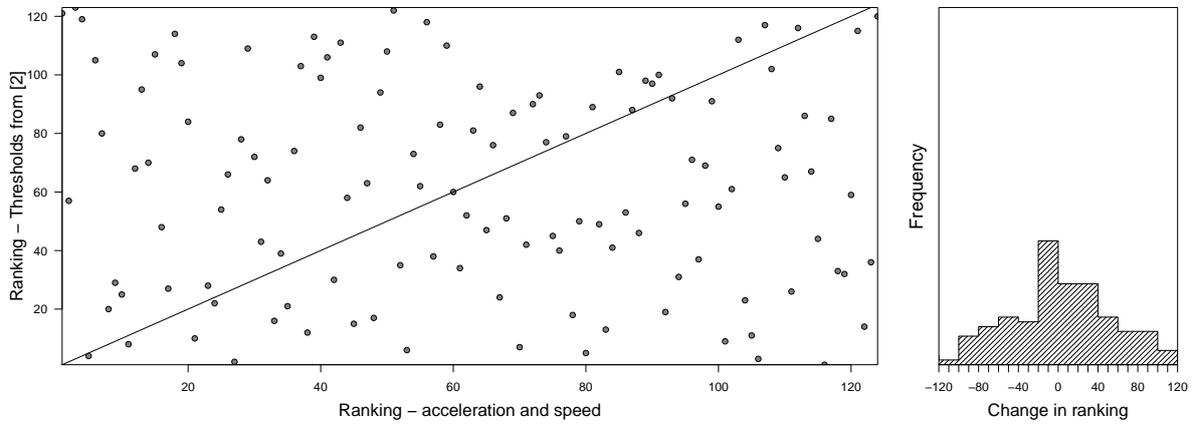
(a) Using thresholds from Baldwin et al. [2].
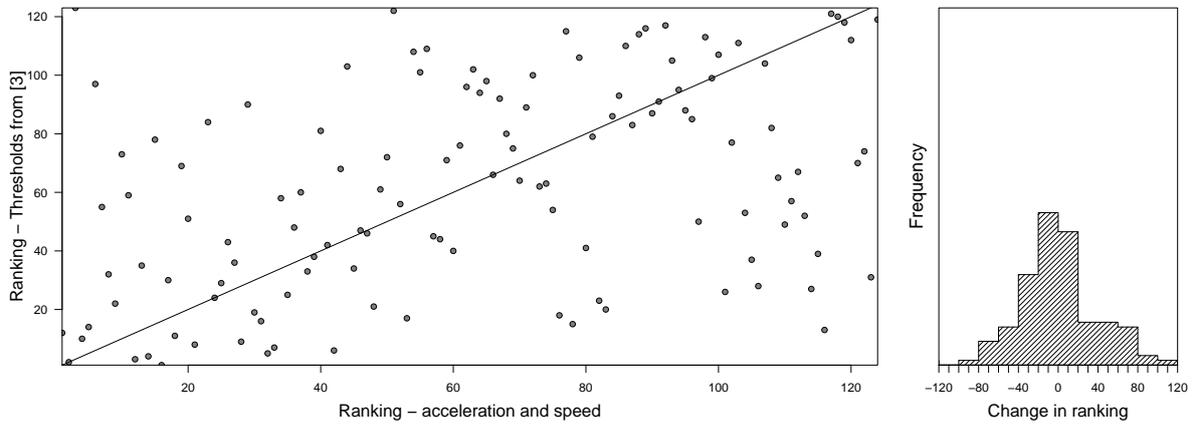
(b) Using thresholds from Bergasa et al. [3].
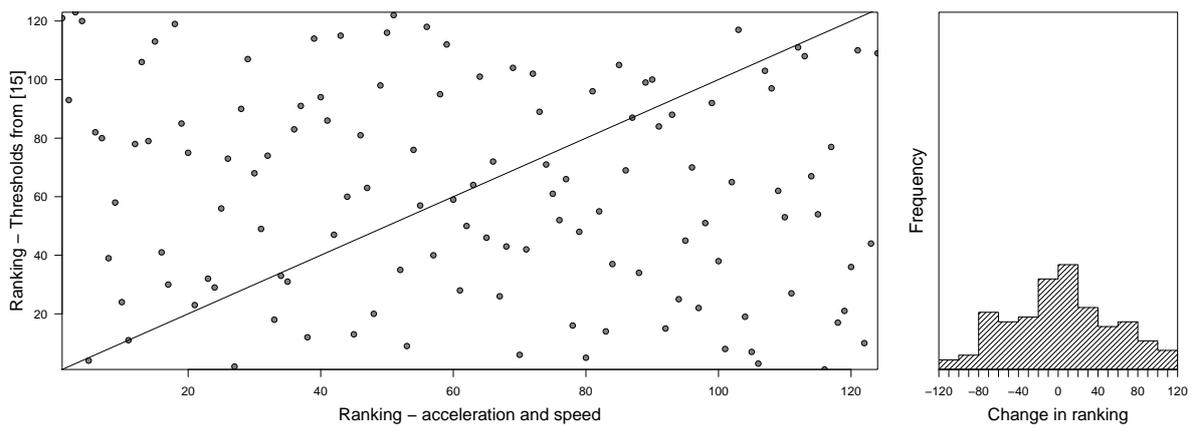
(c) Using thresholds from Paefgen et al. [16].

Figure 11: Comparing the change in rankings when using threshold-based risk from literature.

We specifically choose to not use harsh events in this paper. We acknowledge that such events may indeed be communicated back to the driver, either in real time or in after-the-fact reporting. Our proposed approach would consequently not be able to provide real-time feedback to the driver.

Measurement devices, like the accelerometer and speed sensors used in this paper, has inherent noise associated with them, making extreme event measurements uncertain. Our approach to aggregate the data and use the risk space statistics effectively works around this problem.

In the paper we describe the polyhedral discretisation, its risk classification, and demonstrate three rules for defining risk that are based on combinations of accelerometer and speeding characteristics of the telemetry records. The risk scores are then aggregated to an individual level, and the ranking of customers then show quite clear distinctions between the good, average, and bad performers within the envelope of the customer base.

Our model is able to account for changes in vehicle dynamics and technology in that the model adapts to the pool of customer vehicles, i.e. the actual observed envelope instead of predefined *extreme events* that differs per vehicle type and may prove in future to be not-so-extreme. The objective of this paper is not to provide a calibrated model to assess the risk of driver behaviour. Instead, we provide a discretisation that is more flexible and provide a richer description of risk than the traditional approach using only harsh events and fixed thresholds.

To comment on the use of the results to provide driver feedback, we suspect that the initial novelty will wear off if not reinforced with (recurring) incentives over time. A later study will explore the time variability of the risk envelope.

References

[1] Aviva (2014). Aviva Drive. Available online from http://www.aviva.co.uk/drive/ (accessed: 9 September).

[2] Baldwin, K. C., Duncan, D. D., and West, S. K. (2004). The driver monitor system: A means of assessing driver performance. *John Hopkins APL Technical Digest*, 25(3):269–277.

[3] Bergasa, L. M., Almería, D., Almazán, J., Yebes, J. J., and Arroyo, R. (2014). DriveSafe: an app for alerting inattentive drivers and scoring driving behaviors. In *IEEE Intelligent Vehicles Symposium (IV)*, pages 240–245.

[4] Desyllas, P. and Sako, M. (2013). Profiting from business model innovation: Evidence from pay-as-you-drive auto insurance. *Research Policy*, 42(1):101–116.

[5] Díaz Álvarez, A., Serradilla García, F., Naranjo, J. E., Anaya, J. J., and Jiménez, F. (2014). Modeling the driving behavior of electric vehicles using smartphones and neural networks. *IEEE Intelligent Transportation Systems Magazine*, 6(3):44–53.

[6] Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., Perez, M. A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z. R., Jermeland, J., and Knipling, R. (2006). The 100-car naturalistic driving study, Phase II — results of the 100-car field experiment. Technical report DOT HS 810 593, U.S. Department of Transport: National Highway Traffic Safety Administration.

[7] Farah, H., Musicant, O., Shimshoni, Y., Toledo, T., Grimberg, E., Omer, H., and Lotan, T. (2014). Can providing feedback on driving behavior and training on parental vigilant care affect male teen drivers and their parents? *Accident Analysis & Prevention*, 69:62–70.

[8] Global Road Safety Partnership (2008). *Speed management: a road safety manual for decision-makers and practitioners*. Global Road Safety Partnership: A hosted programme of the International Federation of Red Cross and Red Crescent Societies., Geneva.

[9] Johnson, D. A. and Trivedi, M. M. (2011). Driving style recognition using a smartphone as a sensor platform. In *14th International IEEE Conference on Intelligent Transportation Systems*, pages 1609–1614.

[10] Malm, A. and Fagerberg, J. (2015). The global automotive OEM telematics market. Available online from .

[11] McGwin, G. and Brown, D. B. (1999). Characteristics of traffic crashes among young, middle-aged, and older drivers. *Accident Analysis & Prevention*, 31(3):181–198.

[12] Musicant, O., Bar-Gera, H., and Schechtman, E. (2010). Electronic records of undesirable driving events. *Transportation Research Part F: Traffic Psychology and Behaviour*, 13(2):71–79.

[13] Musicant, O., Bar-Gera, H., and Schechtman, E. (2014). Temporal perspective on individual driver behavior using electronic records of undesirable events. *Accident Analysis and Prevention*, 70:55–64.

[14] Neale, V. L., Klauer, S. G., Knipling, R. R., Dingus, T. A., Holbrook, G. T., and Petersen, A. (2002). The 100 car naturalistic driving study, Phase 1 — experimental design. Interim report DOT HS 808 536, U.S. Department of Transport: National Highway Traffic Safety Administration.

[15] Octo Telematics (2014). Clearbox and mobile data collections. Available online from http://www.octotelematics.com/en/site/pages/driving-behaviour(1) (accessed 9 September).

[16] Paefgen, J., Kehr, F., Zhai, Y., and Michahelles, F. (2012). Driving behavior analysis with smartphones: Insights from a controlled field study. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, MUM '12, pages 36:1–36:8, New York, NY, USA. ACM.

[17] Paefgen, J., Staake, T., and Fleisch, E. (2014). Multivariate exposure modeling of accident risk: Insights from pay-as-you-drive insurance data. *Transportation Research Part A: Policy and Practice*, 61:27–40.

[18] Prato, C. G., Toledo, T., Lotan, T., and Taubman Ben-Ari, O. (2010). Modeling the behavior of novice young drivers during the first year after licensure. *Accident Analysis & Prevention*, 42(2):480–486.

[19] Progressive (2014). Progressive firsts. Available online from http://www.progressive.com/progressive-insurance/first/ (accessed: 8 September).

[20] Raz, O., Fleishman, H., and Mulchadsky, I. (2008). System and method for vehicle driver behavior analysis and evaluation. US Patent US7389178 B2.

[21] Rodríguez González, A. B., Mark Richard Wilby andVinagre Díaz, J. J., and Sánchez Ávila, C. (2014). Modeling and detecting aggressiveness from driving signals. *IEEE Transactions on Intelligent Transportation Systems*, 15(4):1419–1428.

[22] StateFarm (2014). Drive safe and save — in-drive. Available online from https://www.statefarm.com/insurance/auto/discounts/drive-safe-save/indrive (accessed 11 September).

[23] The Co-operative Insurance (2014). Yound drivers insurance. Available online from http://www.co-operativeinsurance.co.uk/youngdriverinsurance (accessed 9 September).

[24] Toledo, T. and Lotan, T. (2006). In-vehicle data recorder for evaluation of driving behavior and safety. *Transportation Research Record*, 1953:112–119.

[25] Toledo, T., Musicant, O., and Lotan, T. (2008). In-vehicle data recorders for monitoring and feedback on drivers' behavior. *Transportation Research, Part C*, 16(3):320–331.

[26] WHO (2015). Global status report on road safety 2015. Technical Report Table A2, World Health Organisation, Geneva, Switzerland.

[27] Wu, B.-F., Chen, Y.-H., and Yeh, C.-H. (2014). Driving behaviour-based event data recorder. In *IET Intelligent Transport Systems*, volume 8, pages 361–367.

[28] Yeo, A. C., Smith, K. A., Willis, R. J., and Brooks, M. (2001). Clustering technique for risk classification and prediction of claim costs in the automobile insurance industry. *Intelligent Systems in Accounting, Finance and Management*, 10(1):39–50.