Appending GPS traces with road grade data to estimate driver behaviour risk

Jeanne-Marie Hugo

A project report in partial fulfilment of the requirements for the degree

Magister (Industrial Engineering)

in the

Faculty of Engineering, Built Environment, and Information Technology

University of Pretoria

April 2021

Abstract

Title: Appending GPS traces with road grade data to estimate driver behaviour risk
Student name: Jeanne-Marie Hugo
Student number: u15038832
Supervisor: Prof. Johan W. Joubert

Road gradients have various impacts when it comes to the performance of a vehicle and its driver. Literature shows how steep grades can cause an increase in emission rates and accident rates, but the actual behaviour and risk of a driver on different road grades are neglected. To find reliable road grade values to use with a behavioural model, this dissertation proposes a smoothing method to be used on the elevation profile of a freely available Digital Elevation Model, namely the 1 Arc second Shuttle Radar Topography Mission dataset.

A behavioural model from literature is used to determine the risk of the drivers within an accelerometer dataset from 45 trucks for a day's travel, using all the records and not only those that fall within certain thresholds of harsh events. The model also allows for road grade to be added as a contextual variable, where records are filtered on five different grade categories ranging from steep downhill to steep uphill roads. These categories each have their own risk space and ranking of driving performance. The majority of drivers have relative constant rankings, but the results show that some drivers behave differently with changes in gradient. Therefore, road grade can influence the behaviour of drivers and can be a useful addition to defining and understanding the risk of a driver.

Contents

List of Figures

List of Tables

Acronyms

GPS	Global Positioning System
NREL	National Renewable Energy Laboratory
SRTM	Shuttle Radar Topography Mission
DEM	Digital Elevation Model
InSAR	Interferometric Synthetic Aperture Radar
MSD	Mean Standard Deviation
MAD	Mean Average Deviation
BWF	Butterworth filter
SG	Savitzky-Golay
DDR	Dynamic Data Reconciliation
US	United States
USGS	United States Geological Survey
PAYD	Pay As You Drive
PHYD	Pay How You Drive
CRS	Coordinate Reference System
CEP	Circular Error Probable
IMU	Inertial Measurement Unit
IDW	Inverse Distance Weighting
RSS	Residual Sum of Squares

Chapter 1

Introduction and background

The behaviour of a driver can be profiled and used in industries such as insurance or fleet management, or in the areas of autonomous technologies or fuel economy analysis. This section discusses how one can construct these driver behaviour profiles when a vehicle's Global Positioning System (GPS) traces and telematics data are obtained.

1.1 Introduction

Vehicle insurance premiums were historically largely based on static factors such as a driver's socio-demographic information, the type of vehicle, annual kilometres, and driving record that includes accidents and violations (Hu et al., 2018). These static factors, however, do not reflect dynamic factors such as distance travelled by a driver per trip, speeding, aggressive driving, and road geometry. Static factors also have the implication of increasing social inequality since lower income drivers tend to drive less (Litman, 2002).

Insurers started using driving patterns that are much more individualised for their pricing schemes more recently, using an individual's speed and erratic manoeuvring as an indicator of behaviour. This agrees with Hu et al. (2018) who argue that speed and pedal operations are the most important measures when characterising driving behaviour. Bener et al. (2020) added that while factors such as speeding and aggressiveness are the biggest predictors for road accidents, mobile phone usage is just as dangerous.

These accidents do not only carry the direct cost of injury or vehicle effectiveness, it also carries insurance-related costs. Insurers always aim to better manage and price their risk and the on-board sensors of a vehicle can provide them with the necessary information to decide whether a driver is a high, medium, or low risk client. Not only can they better quantify their insurance risk, but the drivers can have more information on their own driving to modify their behaviour, ultimately making the roads safer and more efficient.

Two approaches are set out by Shinar and Oppenheim (2011) that can be used to model driver behaviour: descriptive and functional. A descriptive model focusses on what the driver is doing, as opposed to functional models that focus on why the driver is behaving in a certain way. Acceleration and speed are used in a model to profile a driver's risk in a descriptive manner and provides some idea as to how fast and/or aggressive a person normally drives. An accelerometer is the on-board sensor that measures acceleration on three axes, whereby the GPS determines the position of the vehicle and derives the speed from its change in position. The three-axis accelerometers used in this dissertation sense dynamic forces of movements and vibrations within a vehicle, the vertical-axis data can indicate the upward or downward movement a driver makes on the road, specifically when encountering a pothole or speed bump. Acceleration's measurement can be made in units

of meters per second square (m/s^2) or in G-forces (g). Gravitation impacts the magnitude of a single G-force, but it is roughly equivalent to $9.8\,m/s^2$ and can be accounted for when processing vertical acceleration. As with most sensors, the data generated is noisy and needs to be smoothed. This dissertation proposes a useful smoothing method, one which still finds importance in significant spikes in the signal. These spikes can represent, for example, a car driving over a speed bump, or through a pothole.

Functional models are more focussed on the road geometry and traffic. These outside factors can also have a major impact on the behaviour of a driver. Joubert et al. (2016) split up driver behaviour categories in different road type scenarios to analyse the behaviour. One can use categories such as highways, urban roads, and rural roads as different contextual variable categories.

The other aspect of road geometry is the slope of a road, also referred to as *road grade*, and how driver behaviour changes when the geometry changes. The current practice is to leave out road grade or assume it to be zero when studying driver behaviour. An example of this is found with Sentoff et al. (2015) who purposefully left it out when the grade data was not collected or was not available at the exact time of the rest of the data recording. The complexity in determining the road grade is another reason why behavioural models rather assume it to not have an impact.

Authors such as Glennon (1987) and Yu and Abdel-Aty (2014), on the other hand, argue that road geometry does indeed impact driver behaviour. Their works show that level sections of roads have less accidents than that of steeper roads. Hamdar et al. (2016) also argue that the changes in slope or the angle of a road affects a driver's behaviour more than another outside factor like the weather would.

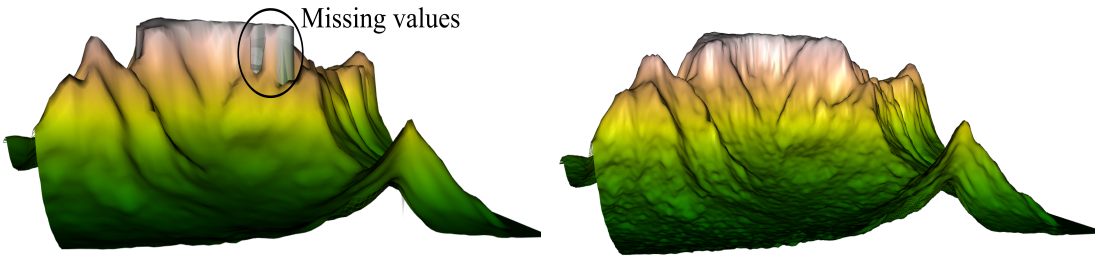
The collection of accurate elevation values for longitudinal and lateral coordinates when calculating the slope causes the before-mentioned complexity, therefore road grade is usually neglected. Capturing or calculating road grade accurately is generally expensive as it requires sophisticated GPS devices or expensive measurement systems (Wood et al., 2014a). The modern way to get the road grade of a roadway is by using a more automated process. This process is a combination of consumer-grade GPS devices to determine position and an automated model to determine road grade. Such models are, typically, unsupervised machine-learning algorithms. The vertical accuracy of these GPS devices is between 10 and 20 meters and therefore not enough for the automotive industry. This forces one to use other types of elevation estimates to aid in the automated road grade calculations.

The work of the National Renewable Energy Laboratory (NREL) extracted elevations from a Digital Elevation Model (DEM) to determine road grades (Wood et al., 2015). The authors produced large-scale road profiles of major United States (US) highways to mainly assess energy usages on certain grades. Finding these road grades from elevations is done by first extracting height values from each pixel in a DEM's grid-like structure of the surface terrain. Each pixel value represents the height/elevation of a squared area. DEMs are assembled using remote sensing techniques like satellite sensors and used for a multitude of scientific applications (Hawker et al., 2018). It seems to provide a reliable representation of topography, often calming the user into a false sense of belief in the accuracy thereof (Wechsler, 2003). As a consequence, one has to be aware of the inaccuracies of DEMs and how to combat them. Davis and Keller (1997) sum it up well saying that "landscapes are not uncertain, but knowledge about them is".

The DEM that is used in this dissertation is that of the Shuttle Radar Topography Mission (SRTM) that is freely available. The global SRTM dataset is measured by synthetic-aperture radar which produces high-resolution surface images from phase differences in

radar waves that are emitted from an antennae attached to a satellite.

There are a few versions of SRTM datasets. To visualise how they can differ in accuracy, two datasets are illustrated in Figure 1.1 using the iconic Table Mountain in Cape Town, South Africa, as a reference. Here the two different granularity scales are visible. Figure 1.1a is of a 3 Arc second representation with $90m \times 90m$ cells and the figure on the right, Figure 1.1b, is the 1 Arc second with $30m \times 30m$ cells. The latter is the most recent version found in Africa and produces a rougher, more realistic surface with a more refined granularity.



(a) 3 Arc second (b) 1 Arc second

Figure 1.1: Table Mountain topography when using different SRTM granularities.

The images not only show how the granularity differs, but they also show how some elevation values are missing in Figure 1.1a – the older version. SRTM derives elevations from synthetic aperture radar and can cause 'voids' or cells where no elevation value could be found. This happens especially with water surfaces and patches of very rough terrain, as seen in the steep inclines of Table Mountain's top right corner. The newest version (with a smaller cell size) of SRTM data in Figure 1.1b does not have these empty cells, producing a more reliable representation of the surface that does not need to be interpolated by multivariate interpolation if one wants a full surface area.

If one desires to have an accurate estimation of the road's surface to use as a variable in a driving behaviour model, the better option would be to use the newest version of a DEM. Road grade can then be derived with much greater reliability.

1.2 Problem statement

Modern methods of determining driver behaviour do not have road grade as a variable. The variables that are used and have a significant impact on behaviour are speed, lateral acceleration, and longitudinal acceleration. Road grade does impact the behaviour of a driver, but it is not clear by how much. The literature has mostly been revolving around higher emission and accident rates on steep(er) road grades, but behavioural changes on these same grades have not yet been quantified.

When one wants to add road grade to a behavioural model it is important to have an accurate estimation or representation of the road profile. Historically, road grades are found with the usage of consumer-grade GPS devices and software. This, however, does not produce reliable elevation estimations and does not create a smooth road profile. Open source DEM data is much more reliable, but introduces a new problem: granularity.

A step-wise elevation profile is the result of a DEM's granularity. When a derivative (road grade) is taken from this step-wise function or curve, the value can be unrealistic because of steep drops and jumps in elevations. Using an unrealistic road grade estimate within the behavioural model will result in an erroneous inference about a driver's risk.

1.3 Research design

In this dissertation, a risk model is built for driver behaviour with different types of road grade on which the driver has driven: very steep uphill and downhill, medium steepness uphill and downhill, and flat roads. This is not only done to profile a driver's behaviour on different road conditions, but is also an analysis as to how much impact slope has on a driver's behaviour.

Behaviour is measured and determined by using the vehicles on-board accelerometer data: three-axis acceleration and speed. This on-board telemetry unit contains information with each recording (more or less each second) that the vehicle is moving and/or is stationary. The GPS that accompanies the accelerometer contains geographical information on the vehicle such as lateral and longitudinal location.

When open source SRTM data is available, one only needs the on-board accelerometer data (with its GPS traces) to profile a driver's risk on different grades of roads. The model provides the road grade values of each coordinate after smoothing the SRTM data. The profiling is then done after a driver's trip information is recorded – it is not instantaneous. At the end, a driver is labelled as a low, medium, or high risk driver.

1.4 Research methodology

The first development to get to this dissertation's behavioural model is finding a sample road. This sample road's purpose is to provide one with a reference of elevation values to find the best model to use for the calculation of road grade. A $28km$ route in Pretoria acts as the reference road profile to test which road grade calculation method is the closest to the *true* road. The sample is made up of highways and urban areas having roughly a 60/40 composition. The *true* elevation values are measured by driving the road multiple times in more than one vehicle and with more than one sensor.

The extracted sets of 1 Arc second ($30m \times 30m$) SRTM elevation data from the United States Geological Survey (USGS) database is compared to the *true* road profile of the $28km$. The sample road now has both the actual and the DEM estimated elevations. Different road grade calculation methods are tested against the reference elevations to find the closest representation of the road.

After slope estimation model is finalised, the next dataset is introduced while the sample road dataset is not necessary any more. Accelerometer data from 45 different long-haul interlink trucks is used to formulate a base behavioural model that calculates a driver's risk based on their planar (longitudinal and lateral) acceleration. These trucks travelled across South Africa and the dataset is only from one month's travelling. The on-board recordings are available for every second the vehicles were running. The base model follows the steps in Figure 1.2.

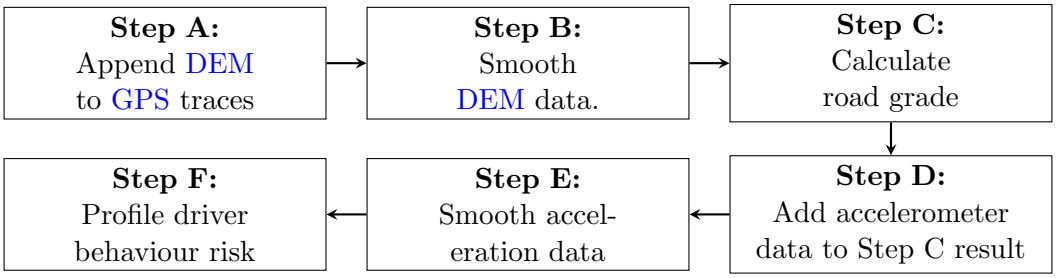


Figure 1.2: Step-by-step generation process of the behavioural model.

The GPS traces in Step A are extracted from the truck dataset. Using the smoothing method, the stepwise SRTM elevations are smoothed on equidistant points in Step B. After smaller equidistant points have been interpolated from the smoothed road profile the road grades are calculated as a function of the distance travelled by the vehicles for each point as set out in Step C.

From the equidistant points, the model interpolates road grade values for each GPS trace of the original dataset. These road grade values are added to the rest of the accelerometer's parameters. The three-axis accelerations are still to be filtered as they are sets of data from a sensor and have some digital noise, therefore, all three accelerations are smoothed in Step E.

For each GPS trace one now has the following valuable parameters: a time stamp, lateral and longitudinal coordinates, road grade, vehicle speed, speed limit, lateral acceleration, longitudinal acceleration, and vertical acceleration.

As Manson (2006) argues, one should evaluate the design against the expectations, the behavioural model of this dissertation will be compared to the base model from the work of Joubert et al. (2016). This is done to see if there is any difference when road grade is introduced as a contextual variable. The expectation is reached if there is indeed a noticeable difference in risk profiles of drivers with the addition of slope. For example, if a risk space from the behavioural model of Joubert et al. (2016) is set up with the contextual variable being a steep downhill, one expects to see more erratic behaviour and more speeding.

The relationship between road grades and driver behaviour is analysed at the end of the study, but this is not the only conclusion. The concept of adding road grade to the model of Joubert et al. (2016) is also discussed and the impact it may have in industries such as insurance.

1.5 Document outline

In Chapter 2, the significance of analysing driver behaviour is investigated in conjunction with literature on the importance of road grade. The smoothing of digital signals as well as the smoothing of elevation profiles are also investigated in this chapter.

Chapter 3 explores how the two sets of positioning data, namely GPS traces and a DEM, can be merged. To find a reference road profile, *true* elevation values are collected from a sample road. This chapter discusses the collection process in detail. After the sample road is set up, the road grade calculation method is finalised.

The behavioural model is replicated in Chapter 4 after the truck data has been appended with road grade values. An analysis is done to see if an incline has an impact on driver behaviour and whether it is at all worth it to add road grade to behavioural models.

Finally, in Chapter 5, a conclusion is reached on the success of the project. Thereafter the appendix contains all relevant documentation for the proposed model.

Chapter 2

Literature review

Literature is reviewed in three main sections. Firstly, the importance of driver behaviour is analysed, and how road grade might have an impact on it. Different road grade calculation methods are investigated thereafter. The final section is on how one can go about data smoothing, making a decision on the best method for both acceleration and elevation smoothing.

2.1 Driver behaviour

Insurance companies have two main categories to assess a driver's risk: Pay As You Drive (PAYD) and Pay How You Drive (PHYD). A PAYD strategy focusses on how someone is travelling over distances, for example which roads they use and how often. PHYD insurance strategies base their analysis purely on driver's personal behaviour, looking at aspects such as speeding, harsh braking and acceleration (Winlaw et al., 2019). In this dissertation, we use a PHYD strategy when profiling the risk of a driver.

The behavioural model that forms the base of this study is of Joubert et al. (2016) who classify risk profiles of drivers on three levels: good, average and bad performers. Their model does not predefine extreme events, rather, the entire envelope of data is observed and the model adapts to the pool of vehicles. The valuable information from these datasets are in the following parameters: Global Positioning System (GPS) location, time stamp, speed, lateral acceleration, longitudinal acceleration, speed limit. They show that their model defines extreme events different than most literature. The big difference is the discretisation since it does not consider a predefined threshold that is exceeded as an extreme case of dangerous driver behaviour.

Data is discretised into a three dimensional risk space. Joubert et al. (2016) then use horizontal slices to visualise how the three different risk levels look within the three dimensional space. These horizontal slices are taken at a z-acceleration (or vertical acceleration) of $1009mg$ where the most data entries are available.

Celaya-Padilla et al. (2018) state that road anomalies such as potholes and speed bumps have a significant effect on accident rates. Speed bumps can also cause drivers to change direction or even swerve to avoid reducing their speed on the bump (Xu et al., 2015). These obstacles are not always labelled correctly on road maps, but can be recorded with the use of sensors in a vehicle. Using the outputs of a gyroscope, an accelerometer and a GPS sensor, Celaya-Padilla et al. (2018) developed a genetic algorithm that detects road abnormalities. In this report the vertical acceleration from such sensors is examined, but is not used as part of the risk profiling of drivers. It is only analysed to see how one might be able to use road surface information in the context of driver behaviour – ultimately

one can improve fuel consumption as well as the safety of drivers and pedestrians.

While the impact of longitudinal behaviour impacts the overall behaviour of a driver, other studies such as that of Hamdar et al. (2016) show how the characteristics of a roadway influences behaviour. These studies mostly focus on characteristics such as horizontal alignment, road and shoulder width, and median barriers. This dissertation, however, looks deeper into a road's vertical gradient since it is known that accident rates are higher with grade sections compared to level sections; steep grades compared to mild grades; and downgrades compared to upgrades (Glennon, 1987).

While this dissertation does not investigate the performance (or vehicle specific power) of a vehicle, it can be mentioned that it can have an influence on a driver's behaviour. By investigating speed and driving violations, Horswill and Coster (2002) show how there is a significant relationship between a vehicle's performance and the driver's choice of speed - a higher performance is associated with a higher choice of speed. This relationship is ultimately shown in their study to influence the risk-taking behaviour of a driver - the more power a vehicle has, the greater the risks a driver would usually take. The choice within car-buying also correlates with risk-taking, i.e., buyers with risk-taking propensities will place more value on a vehicle with increased performance. Horswill and Coster (2002) also discuss how other factors such as smoothness, safety, and noise would influence the risks a driver would take. Here they did not only look into speed and driving violations, but also added gap acceptance and preferred car following distance. Gap acceptance refers to the amount of space someone will be willing to pull into in a stream of traffic.

Because this study has a homogeneous dataset with heavy vehicles that have more or less the same specifications, the impact of vehicle performance is not analysed - the scope of analysis is kept at road geometry.

2.2 Elevation and road grade

Road grade is the change in elevation divided by the horizontal distance in the form of percentage (Boroujeni and Frey, 2014). It quantifies the slope of a road, i.e. a negative road grade value would indicate a downhill.

While this study only focusses on road grade's impact on driver behaviour, it is quite significant in other areas as well, especially in vehicle economics. Sentoff et al. (2015) found that road grade has a notable effect on emissions when studying vehicle specific power. The authors argue that neglecting road grade underestimates energy and emissions, with an error of between 10% and 20% and fuel economy improves by 15% to 20% when a vehicle is on a flat route. One can therefore investigate more than just a driver's behaviour with road grade.

To find road grade, one needs elevation and distance as mentioned above, but there are other (instantaneous) methods as well. Different on-board equipment can be used to measure the instantaneous elevation or even the road grade of a vehicle directly. Accelerometers can provide the latter, but interference with unrelated vehicle movement creates a profile that is very noisy (Rogers and Trayford, 1984). Another commonly used measurement tool is a GPS logger (Boroujeni and Frey, 2014). However, the accuracy once again presents a limitation with low resolution. Although horizontal errors with GPS are low, vertical errors are much higher since buildings and trees interfere with signalling. Changes in satellite locations will also affect the accuracy and resolution of the elevation. Condequently, GPS is most accurate when paired with other sensors and equipment. When a GPS logger is used with a barometric altimeter, only then it is deemed accurate enough (Boroujeni et al., 2013).

Outside of a vehicle's sensors, road grade can be found on original or revised road design drawings, but these are only available for large building projects like a highways or bridges. These elevation profiles might also differ from the actual roads since the drawings sometimes only specify the original construction design and not all the changes that were made during construction (Zhang and Frey, 2006).

There are various open-source elevation data sets available that do not use the above mentioned collection methods. Digital Elevation Models (DEMs) can be used to estimate elevation and will provide the most accurate road profile (Liu et al., 2018). Spaceborne Interferometric Synthetic Aperture Radar (InSAR) is the most popular method used to create DEMs and is the technology behind the Shuttle Radar Topography Mission (SRTM); the biggest open-access global DEM. By 2015 the SRTMGL1 version covered Africa, Europe, North America, South America, Asia, and Australia in $1^\circ \times 1^\circ$ tiles at 1 Arc second (about 30 meters) resolution. To date, this is the most accurate DEM that covers South Africa and most of the Southern Africa area.

DEMs are not a perfect representation of topography and like most measurement systems it has a few errors. These errors can be grouped in categories of cause: deficient and/or old data; processing and numerical; faulty measurement errors from faulty equipment or poor placement accuracy (Wise, 2000). Errors can have a ripple affect in the accuracy of slope and curvature of the surface, and need to be tracked when elevation values are used in calculations such as that of road grade. To combat the accuracy uncertainty, this dissertation attempts to filter out DEM errors as well as smooth out the spatial resolution errors of the 30 m SRTM.

The most accurate road grade calculation method according to literature is in the form of a derivative of the elevation profile, but this report examines another option of using the triple-axis data from an accelerometer to find road grade. Acceleration-based and velocity-based slope measurements both fall into this option.

The three main parameters of a triple axis accelerometer are the lateral, longitudinal, and vertical (x, y, and z) accelerations. Furthermore, the unit provides the measurements of yaw, roll, and pitch angles when a vehicle is free to operate in three dimensions and have motion on the orthogonal axes, all centred around its center of gravity. Yaw, roll, and pitch are the motions about the perpendicular, longitudinal, and lateral axes respectively. Yaw dynamics are present in vehicles and can be improved by the improvement of rear wheel steering and control of brake torque. The roll dynamics are mainly present when a vehicle's suspension is analysed. Pitch is the motion affecting the vehicles considered in this report the most and is a measurement of how far the nose of a car is tilted up or down. It is present when there are motions of positive and negative acceleration or when driving on irregular road conditions. Yaw and roll are more important in the analyses of submarines or aeroplanes and will not be considered here.

When road grade is derived via the acceleration method as mentioned earlier, one does not need knowledge of the vehicle and the method is independent of the signal from its GPS. The accelerometer is fixed to the vehicle, and the road slope is added to the pitch to measure the grade angle. Road grade, Θ_{acc}, is then calculated by on-board vehicle velocity and acceleration as follows:

$$\Theta_{acc} = \sin^{-1}\left(\frac{a_{acc} - \frac{d}{dt}v_{wheel}}{g}\right) \tag{2.1}$$

where a_{acc} denotes the longitudinal vehicle acceleration, v_{wheel} the longitudinal wheel-speed, and g the gravitational acceleration.

The second method uses GPS-velocity-based measurements and is not sensitive to

motion that changes the vehicle pitch. Road grade, Θ_{GPS}, calculated by the velocity of the vehicle, as measured by GPS traces is as follows:

$$\Theta_{GPS} = \tan^{-1}\left(\frac{v_z}{\sqrt{v_x^2 + v_y^2}}\right) \tag{2.2}$$

where v_x, v_y and v_z denote the velocities on all three planes.

The acceleration-based estimation can only work in moving conditions, whereas the GPS-based estimation can be used after recording the vehicle's movement. Although the latter is a better option when doing postprocessing on previously recorded data, it can have an error in the road grade estimate that can increase by too much in low speed regions of below 5m/s, or approximately 18km/h (Jo et al., 2013). With this difference in mind, the two calculation methods are examined and compared further in Chapter 3.

The last two road grade estimation methods refer back to the fixed road profile and are formulated by simple derivatives and linear regression of the DEM, respectively. Magrath and Brady (2017) suggest that an 80 m interval is sufficient to measure road grade as a derivative of distance travelled, but this argument will be investigated in Section 3.5 using other interval length options that work best with the specific dataset used in this study.

2.3 Data smoothing techniques

For the risk model there are two parameters that need some kind of filtering or smoothing. These parameters are elevation and acceleration. The latter will be discussed in the second part of this section, and literature on elevation smoothing is investigated first.

To derive road grade, one needs elevation values that accurately represent the profile of a road. To find this accurate representation, the stepwise elevations extracted from the SRTM dataset need to be smoothed. The stepwise nature of the road profile can cause derivatives to have completely incorrect slope estimations, but a smoothing model that is fitted to the profile will remove these sudden spikes or drops.

Wood et al. (2014b) smooth elevations with a Savitzky-Golay (SG) filter. The method uses local regression instead of the popular moving average,000 thereby flattening peaks less than the moving average filter would. The SG filter is also a low-pass filter and uses a local least-squares polynomial approximation (Schafer, 2011). It is another way to reduce noise while keeping the shape and height of waveform peaks. SG filters are thereby known to properly preserve the shape of peaks and will, in turn, preserve a road profile's peak and not remove it. Bromba and Ziegler (1981) define the filter operator, $Af(k)$, using (2.3),

$$Af(k) = \sum_{n=-\infty}^{\infty} a(n) f(k-n) \tag{2.3}$$

where f denotes the original dataset and a denotes the filter function. When a smoothing degree of $2M$ (an even number) is desired, the smoothing filter is written as

$$D_{2M} f(k) = \sum_{n=-N}^{N} a_{2M}(n) f(k-n) \tag{2.4}$$

with $a_{2M}(n) = 0$ if $|n| > N$.

Figure 2.1 illustrates how an original dataset is smoothed with a SG filter. It can be seen how the shape of the peak is preserved during smoothing.

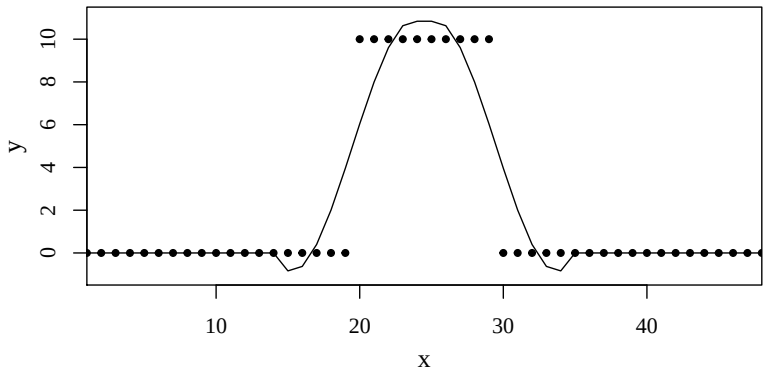


Figure 2.1: Example of an SG filter on arbitrary points. Dots represent the original dataset and the line represents the smoothed function fitted.

Another smoothing method that uses the square error of an approximation, is a *natural cubic spline*. A spline function is a curve constructed out of independent polynomials that are continuous at their joining points (Burden et al., 2011). If f is defined at $a = x_0 < x_1, x_1 < \cdots < x_n = b$, then f has a unique natural spline interpolant, s, on the nodes $x_0, x_1, \ldots, x_n$; that is, a spline interpolant that satisfies the natural boundary conditions $s''(a) = 0$ and $s''(b) = 0$. Figure 2.2 demonstrate how a spline is fitted to an original dataset.

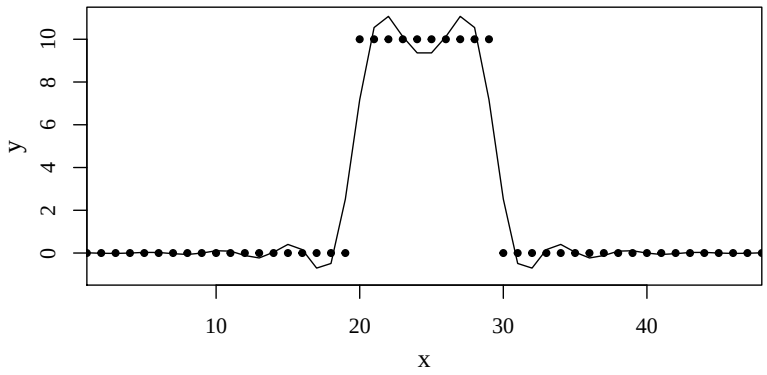


Figure 2.2: Example of a cubic spline on arbitrary points. Dots represent the original dataset and the line represents the smoothed function fitted.

Liu et al. (2018) use the spline $s''(x_i)$ for the smoothing of an elevation profile of a road. Since the curve is fitted on nodes, the degree of smoothness influences the actual fit of the data points. To minimise the trade-off between the goodness-of-fit and smoothness, the minimum trade-off, L, is calculated using (2.5),

$$L = \alpha \sum_{i=0}^{x_n} w_i (y_i - s(x_i))^2 + (1 - \alpha) \int_{x_0}^{x_n} (s''(x))^2 dx \tag{2.5}$$

where $s(x_i)$ denotes the smoothed elevation value at point x_i; w_i the weighting of each individual point starting all equal to 1; and α the weight factor with $\alpha \in [0, 1]$.

A method that is worth noting, but is not relevant in this dissertation, is the one by Chen et al. (2017). The authors present a local enhancement mechanism integrated

with a Butterworth filter (BWF) to remove noises in raw elevation data. They compare the statistical methods of Mean Standard Deviation (MSD) and Mean Average Deviation (MAD) to a revised BWF and the results suggest that the BWF method remove noises in the data most effectively. The filter is often referred to as a *maximally flat magnitude* filter. Meaning, it always chooses the most flat response surface. The concept sounds attractive when working with road profiles since most roads are flat, but the filter overshoots at high orders. Chen et al. (2017) use a local enhancement mechanism to accompany the de-noising of overshooting, but the enhancement only provides more computational challenges. As this method is more popular in the analysis of audio circuits, it is disregarded as a possible smoothing method in this report.

The acceleration data is considered more noisy as the elevation data and needs a smoothing that can accommodate sudden spikes and turbulence of seemingly-straight lines.

Noise can be removed on various degrees, the most common methods to reduce such noise are the popular moving average and exponentially-weighted moving average. They are both sufficient when a process is close to a steady-state, but lack in capabilities when a more dynamic state is present. A more sophisticated model that can handle dynamic states is a Kalman filter – a filter widely used in the aerospace industry.

Kalman filtering is specialised to sensor data, adjusting current data to past data to reduce noise in the measured value (Park et al., 2019). It is the first step to predict the next value to be measured based on the previous readings. It then refines the estimated point with the actual measurement to find a value closer to the latter. One of the input parameters of a Kalman filter is the measurement noise variance, which an ordinary user will find challenging to find without in-depth knowledge on the type of sensor. It often happens that someone chooses a variance arbitrarily, resulting in a filter with poor performance. Park et al. (2019) developed a framework to calculate the noise variance from an original set of data. From their two methods to predict values, the use of a wavelet transform is applicable here, since the second method is more inclined to neural networks.

The Kalman filter aims to correct noisy time-series data (Kalman, 1960). Probabilistic estimations are applied to a state-space model that uses past data recursively. The accuracy of the results are higher when compared to a generic smoothing filter using only incoming values. The state equations are set up as follows:

$$x_t = F_t x_{t-1} + w_t \tag{2.6}$$
$$z_t = H_t x_t + v_t \tag{2.7}$$

to predict:

$$\bar{\hat{x}}_t = F_t \hat{x}_{t-1} \tag{2.8}$$
$$\bar{P}_t = F_t P_{t-1} F_t^T + Q_t \tag{2.9}$$

for the correction:

$$K_t = \bar{P}_t H_t^T (H_t \bar{P}_t H_t^T + R_t)^{-1} \tag{2.10}$$
$$\hat{x}_t = \bar{\hat{x}}_t + K_t(z_t - H_t \bar{\hat{x}}_t) \tag{2.11}$$
$$P_t = (I - K_t H_t)\bar{P}_t(I - K_t H_t)^T + K_t R_t K_t^T \tag{2.12}$$

where x_t is the true state at time t; z_t is the measurement at time t; F_t is the state transition model applied to x_{t-1}; H_t is the observation model; and w_t and v_t are the process and measurement noise variance respectively. Figure 2.3 illustrates how the Kalman filter ultimately updates its predicted values to filter out the noise in the vertical acceleration.

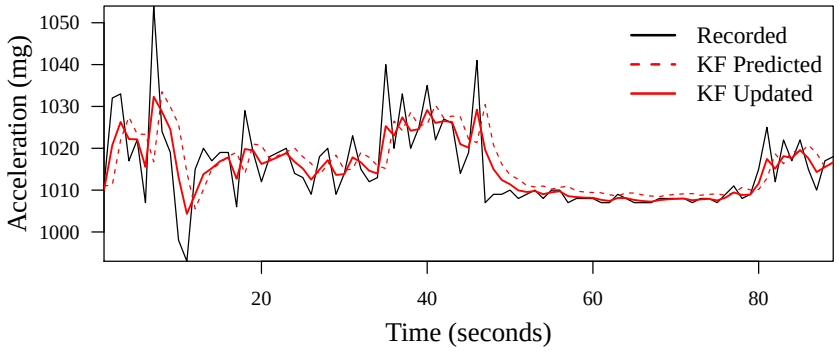


Figure 2.3: Example of applying a Kalman Filter on z-acceleration with predicted values and final updated values.

Table 2.1 summarises the smoothing techniques reviewed. Since the accuracy of the elevation profile is more important than that of the acceleration data, two methods will be implemented and compared before calculating road grade profiles in Section 3.5. Only the Kalman filter will be used in the smoothing of the acceleration data.

Table 2.1: Comparison of smoothing filters.

Filter	**Description**	**Article**	**Chosen for**
Kalman	State-space filter predicting and correcting the prediction per time interval for linear systems	Park et al. (2019)	z-acceleration
SG	Least squares polynomial approximation preserving peak shapes	Wood et al. (2014b)	DEM
Spline	Least squares polynomial approximation with independent polynomials	Liu et al. (2018)	DEM
Dynamic Data Reconciliation (DDR)	Kalman alternative, but with prefiltering before reconciliation	Bai et al. (2006)	-
BWF	Used when a maximally flat pass band is required and can handle non-linear systems.	Chen et al. (2017)	-

Chapter 3

Model

This chapter sets out the process for the generation of the smoothed elevation model with road grade estimations. It explains how the sensor and Digital Elevation Model (DEM) data is extracted, processed and used in calculating road grade. Before the elevations can be used in the road grade calculations, they are smoothed using two smoothing techniques: the Savitzky-Golay (SG) filter and the cubic spline. After the smoothed elevation profile is found with the most accurate technique, the method for determining road grade is discussed and concluded.

3.1 Sample road

To test the elevation smoothing techniques, a *true* elevation profile can be established to act as a gold standard. This is achieved by driving a road segment and measuring the elevations of the road using high-accuracy, on-board equipment. A sample route is chosen in Pretoria with a length of $28km$, with 60% on highways and the rest on urban roads. The absolute ground truth for the elevation profile is unknown.

To find the gold standard elevations, two sets of elevation profiles are found using two different methods. Different vehicles are equipped with different measurement tools that measure altitude differently. Sections 3.2 and 3.3 detail how these road profiles were found and how they compare against each other. Both gold standard estimation methods have the same base coordinates and their latitude and longitude points are generated $10m$ apart along the sample route. The total of 2800 points are visualised on the route[1] in Figure 3.1.

A base list of coordinates are now established and ready to be added onto with elevation values. When a reliable set of elevations are available, only then can it be compared to the DEM data with its applied smoothing techniques. Road grades can then ultimately be calculated as a function of the DEM elevations and be used as a contextual variable in analysing driver behaviour.

3.2 Road profile generation with VBOX

The first *gold standard* road profile is estimated using a *Racelogic* electronic measurement system fitted to a highly sensorised Land Rover Defender, courtesy of Dr Herman Hamersma of the Vehicle Dynamics Group, Department of Mechanical and Aeronautical

[1]Coordinate Reference System (CRS) of Hartbeesthoek94-Lo29, EPSG:2055, WGS84, revision date: 2019-01-14.

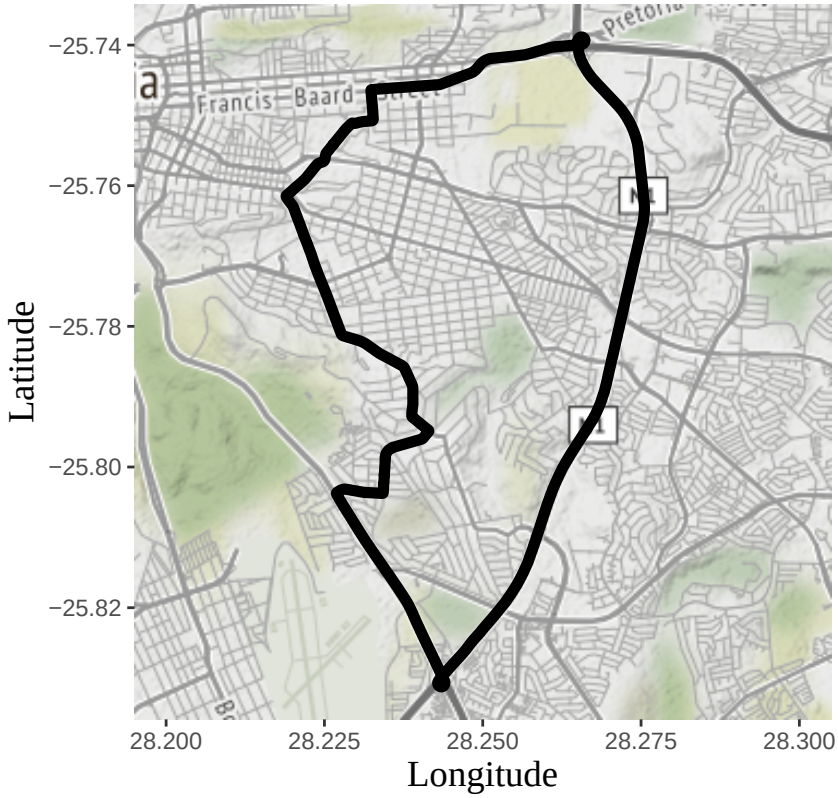


Figure 3.1: 28km sample road route in Pretoria.

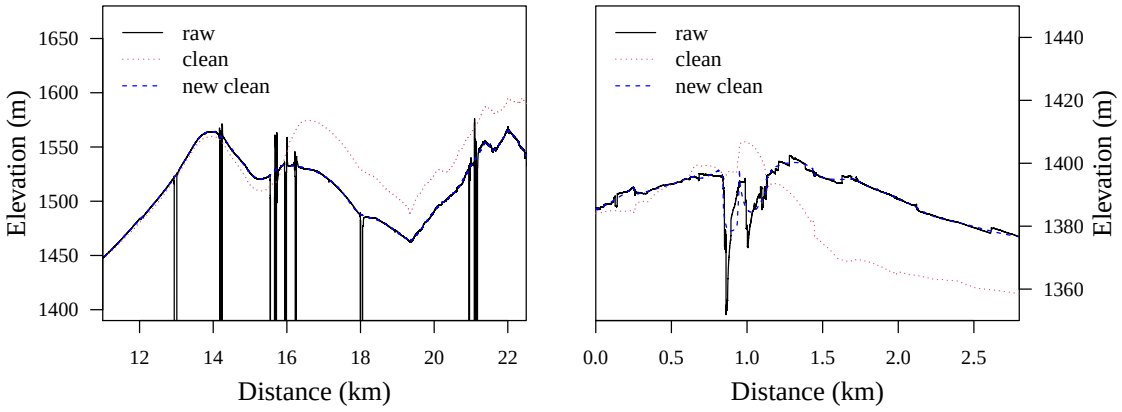
Engineering at the University of Pretoria. The vehicle was driven on the sample route once to record the elevation profile. The measurement system also recorded some more parameters that will be introduced later in this report.

Racelogic's VBOX has an Inertial Measurement Unit (IMU) coupled with their VBOX 3i Global Positioning System (GPS) data logger to improve GPS accuracy when the sky's visibility is unfavourable. This allows for smoother velocity data and a pitch and roll angle accurate to 0.06°. The yaw rate and acceleration resolution is 0.0137°/s and 1.5 mg respectively. The IMU04 used by Racelogic gives pitch, roll, and yaw rates using three gyroscopes, as well as providing acceleration on all three axes via three accelerometers.

The VBOX 3i data logger records at a rate of 100 Hz. With this data logger integrated with the IMU, it uses a Kalman filter on the measurements made in real-time. Dynamic overshoot is removed through lever arm correction by the addition of a dual antenna system.

Examining the parameters of the system's output, it was found that a Kalman filter is used on the raw elevation values to produce the *clean* output elevation values. A Kalman filter should primarily be used as a digital filter and a distorted elevation profile was created because of it. The filter, therefore, should only be used on parameters such as the three axis raw accelerations and pitch angles. The distorted elevation profile called *clean* can be seen in Figure 3.2a. The *raw* values are the values of the VBOX before any corrections took place.

To fix the big difference in the *clean* elevation profile, the elevations are re-estimated from the raw values. Outliers are removed and high frequency recordings of the unit are smoothed out by fitting a spline to the curve. The spline is fitted to smooth the raw values



(a) 12km showing distorted clean values.

(b) First 3km with a big error in recordings.

Figure 3.2: Two extracts showing the raw, clean, and new values of the VBOX elevations.

as close to the raw profile as possible. These elevation values are now the *new clean* values and are also plotted in Figure 3.2a.

While the spline sits neatly on the raw elevation profile, a strange occurrence is found in the first few kilometres of the trip taken with the Land Rover. This error is visualised in Figure 3.2b. The strange elevation values cause the fitted spline to be greatly misshaped. The incorrect raw values might be because of a loss of the GPS signal or caused by obstructions on or next to the road.

The VBOX records the number of visible satellites on its trip and this data might indicate whether there was indeed a loss of signal or if the strange values are caused by something else. The number of visible satellites are plotted in Figure 3.3. The recommended minimum number of satellites for a reliable GPS connection is three, and the figure shows that there were a few points (1.8% of the entire trip) where the number of visible satellites went below three, but only momentarily. In the problematic area of the first two kilometres however, no big drop in the GPS signal was experienced.

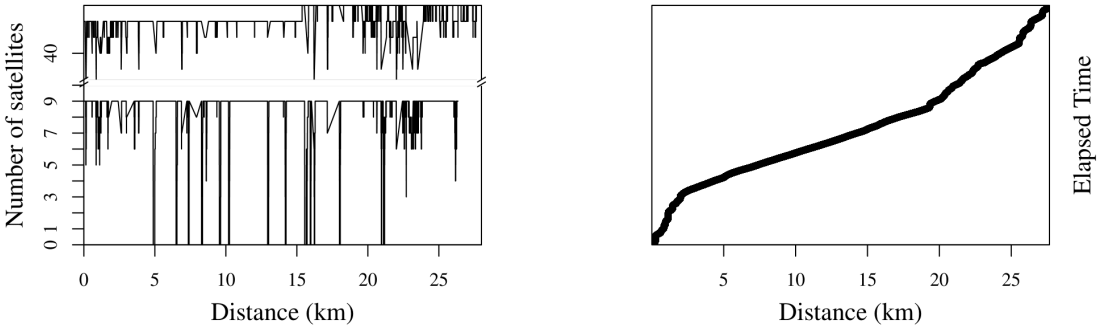


Figure 3.3: GPS Signal investigation on trip.

In the figure on the right, the times of the VBOX recordings are plotted. If any break in the recordings was present, it would show that there was a loss of signal or a momentary break in the VBOX's recordings, but the line is continuous with no signal complications on the trip. This leads one to believe it could have been an error within the measurement system, and not the signal. Other reasons might include bad satellite positioning, multi-paths caused by buildings or bridges, or ellipsoid model errors of the GPS chip. Since one

cannot conclude what the reason might be, these error sections will just not be used in the determination of smoothing techniques.

The *new clean* values are now used as the gold standard elevations from the VBOX. This Data Logger records longitude and latitude with an accuracy of 3 m with a 95% Circular Error Probable (CEP), meaning that the readings will fall within a circle of 3 m radius 95% of the time. The height accuracy however, is lower at 6 m with a 95% CEP. The new road profile is illustrated in Figure 3.4 with its CEP. Figure 3.1 is updated with coloured elevation values and shown in Figure 3.5. Red and purple are the lowest and highest points on the route, respectively.

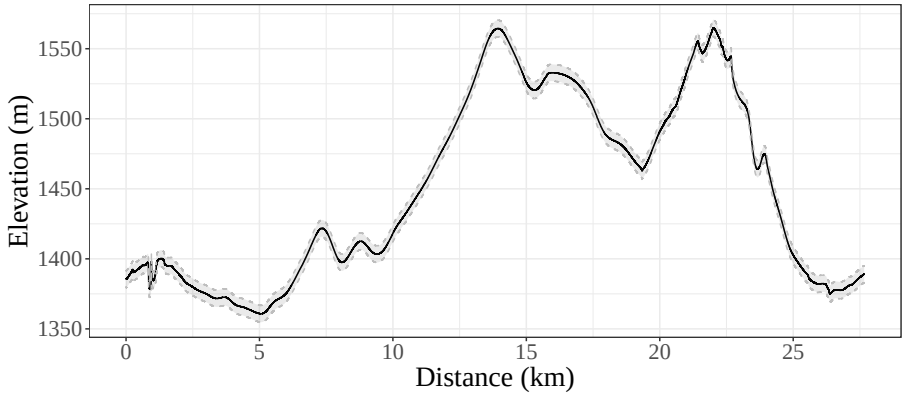


Figure 3.4: Elevation profile of trip with 95% CEP in light grey.

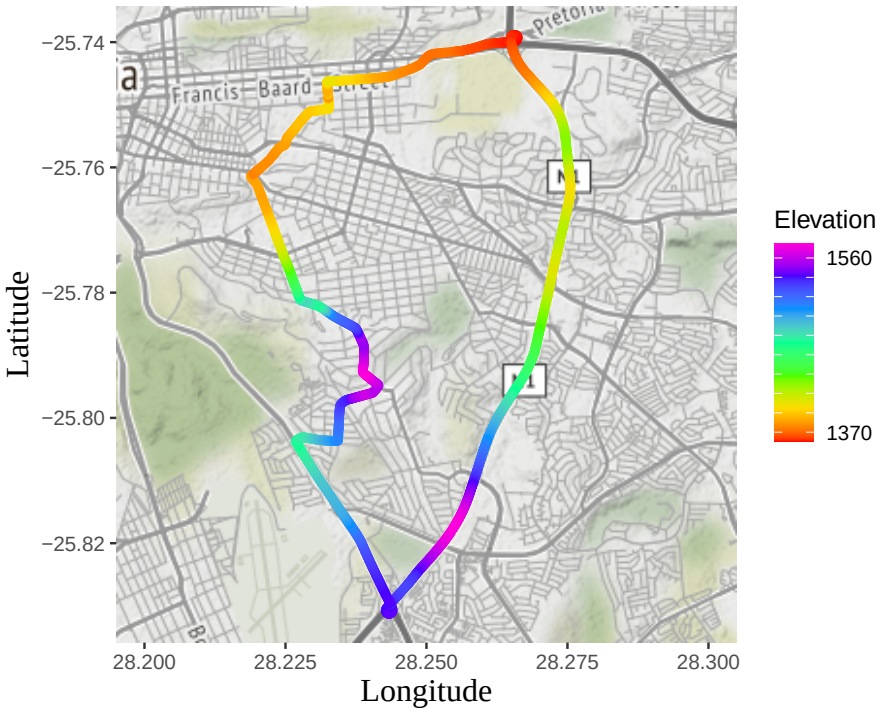


Figure 3.5: Sample road route with elevation values.

With the confirmed road profile, road grade can now be calculated to test the profile since additional data is available from the VBOX and its IMU. The 100 Hz measured

parameters are: time, position (latitude and longitude), velocity, heading, height, vertical velocity, lateral acceleration, longitudinal acceleration, radius of turn, centreline deviation, pitch angle, roll angle, and yaw angle.

An assumption is that the travelled distance (interval) has a constant slope. The accelerometer is installed in the center of the vehicle as illustrated in Figure 3.6. The figure indicates all the important parameters that will be used in this study.

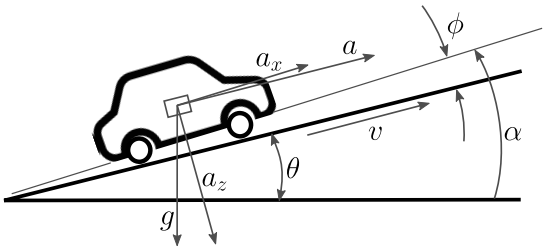


Figure 3.6: Vehicle movement on an incline with all its relevant measurements.

The longitudinal and vertical acceleration are denoted by a_x and a_z, where a represents the vehicle acceleration on the plane. The velocity, v, and acceleration, a, are parallel to the road surface. The pitch angle, ϕ, is not necessary to use in this study since the only angle that is needed is the road grade, but the vehicle inclination angle, α, is divided into the vehicle pitch angle and the road gradient angle, θ,

$$\alpha = \phi + \theta. \tag{3.1}$$

After determining that the filters used by the VBOX concerning the dynamic data are valid, the clean data is used in two ways to calculate road grade. The first method uses the on-board vehicle acceleration and velocity as in Equation (2.1). The second method is chosen by using the GPS velocity and on-board velocities as in Equation (2.2).

These two methods are calculated with the VBOX parameters. Their grade distributions are illustrated in Figure 3.7. It can be seen that the methods do not follow the same distribution, with the second having too many values out of the tolerance of -0.15. Between 8% and 5% a road is marginally steep and between 8% and 15% it is very steep. Beyond 15% (or 0.15) is considered outside of reasonable bounds on a roadway (TxDot, 2018).

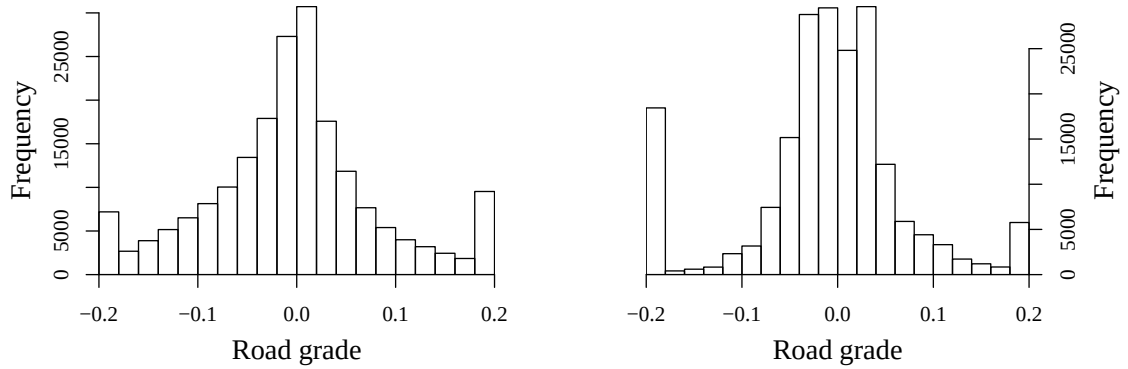
With both the velocity and acceleration methods there are shoulders in the far ends of the distributions. While promoted in literature for certain applications, this dissertation's results indicate that they are not reliable for the the 100 Hz data because of the unrealistic road grade estimates they produce.

To distinguish if this is in fact the equations that do not work well because of its instantaneous nature, or if it might be the VBOX data's unreliability, a second approach is used to find an elevation profile of the sample road. This second approach is estimated with a different device and does not only use one run on the route.

3.3 Road profile generation with 15 runs

The next method to find a *true* representation of the sample road's elevation profile is done by driving the same route 15 times. A commercial barometric (pressure-based) altimeter[2]

[2]Garmin Oregon 650.



(a) IMU Acceleration.

(b) GPS Velocities.

Figure 3.7: Histogram comparison of the first two road grade calculation methods.

was installed in the passenger vehicle to record all 15 runs' GPS traces, along with the elevations of all the points. These points, however, are not recorded as frequently as with the VBOX, but combining data from 15 trips help increase the number of values and minimise variability on the $28km$ route.

The variation in elevation recordings are illustrated in Figure 3.8. Here one can see

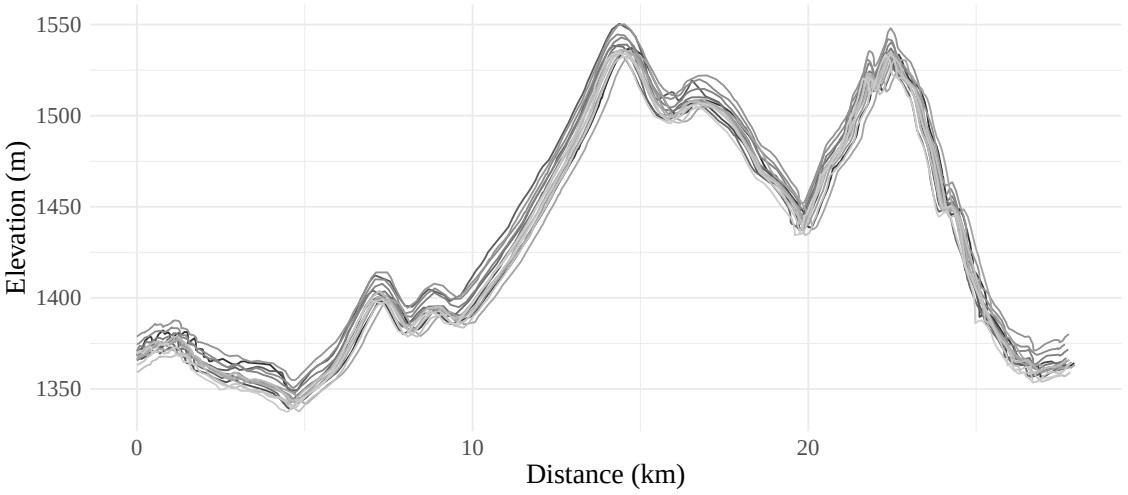


Figure 3.8: Road profile as with recorded elevations of 15 runs.

that there is a range in elevation values of about 20 meters between all 15 recordings and simply combining them would produce consecutive differences in elevations that are too high to represent a credible road grade values. The combined data should thus be weighted in order to find a smooth road profile. Removing the big jumps in elevations can be done by the implementation of a k nearest-neighbour method. The sample route's $28km$ with $\pm 10m$ points from Section 3.1 are used again, where an elevation value is added to each point using its $k = 6$ nearest-neighbours out of the combined trips. To find this *nearest* value, one needs to use multivariate interpolation because there is more than one variable that needs to be estimated namely, latitude, longitude, and height. This multivariate interpolation is called Inverse Distance Weighting (IDW).

With IDW, the heights at the assigned points on the sample road are calculated with a weighted average, using the values available from the known neighbours. The set of known

points are those of all 15 trips together. The IDW creates a smooth continuous function around the set of points different from the known points and the function then provides the third variable of height for every sample point, leaving a set of 2800 points as needed.

To estimate road grade, Rogers and Trayford (1984) collected elevations with 10 trips on a road segment and used the mean values to determine the final road grade. Here the number of trips are more than 10, therefore using the median value can also be considered. A mean is an average of the points selected, while the median is the middle value of all the points. Both are used to find an elevation profile and to calculate road grade as derivatives of 50 m intervals. The road grades for the trip are plotted in Figure 3.9 showing that the median values in light grey create a more erratic road profile and the erratic nature does not change if the IDW interval size is altered. The mean (as used by Rogers and Trayford

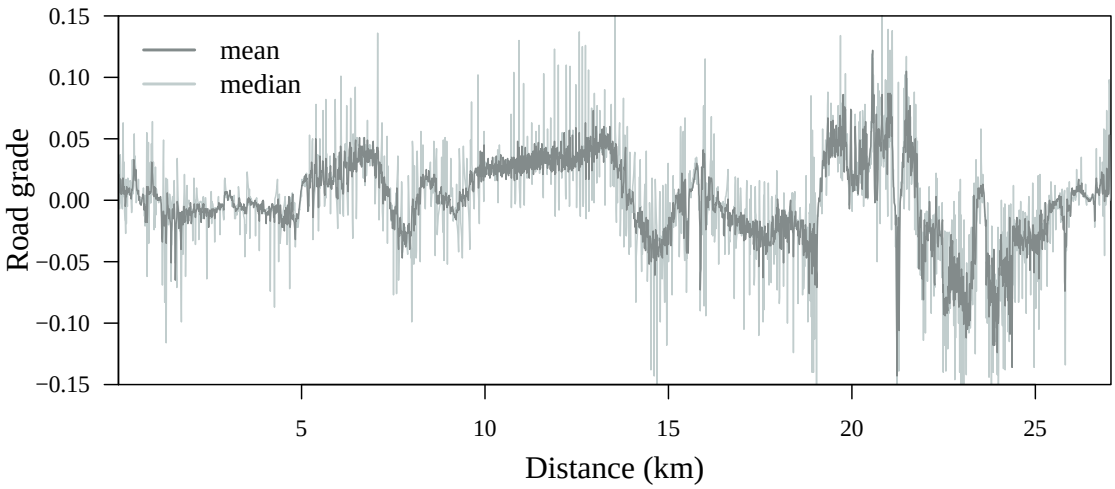


Figure 3.9: Road grade values using the mean and median elevations of the runs.

(1984)), therefore, proves to be the better of the two options when constructing the road profile.

What needs to be decided is whether the above method of multiple runs and their mean elevation values are the best way to estimate the gold standard elevation profile of the sample route, or if the VBOX (with all of its alterations) is better. The VBOX data had to be adjusted quite a lot to fit into what is believed to be the *true* road, and the multiple runs might seem redundant to find an accurate profile using a mean of 15 runs. They both have quite an uncertainty to them.

Using more than 10 runs will remove any unwanted errors in values, which was achieved with the multiple runs method, but was not attempted with the VBOX. It was assumed that because of the unit's accuracy and frequency of recordings, that it would be sufficient to only drive one trip on the route. One more trip can be run on the same route with the same Land Rover Defender and the same VBOX device, but it will need to be recalibrated since it wasn't known what caused the errors in the first few kilometres.

On the second run, all the same variables were measured and a gold standard elevation profile was estimated on the sample points of the route. The profile is visualised in Figure 3.10 and the result is almost identical to the first run. There are no unrealistic dips in the first few kilometres as before and one can now have more confidence in the CEP and its height accuracy.

At this point, there are two estimations but it is still unclear as to which of the two are closest to the *true* road profile. The mean values from the multiple runs can be compared

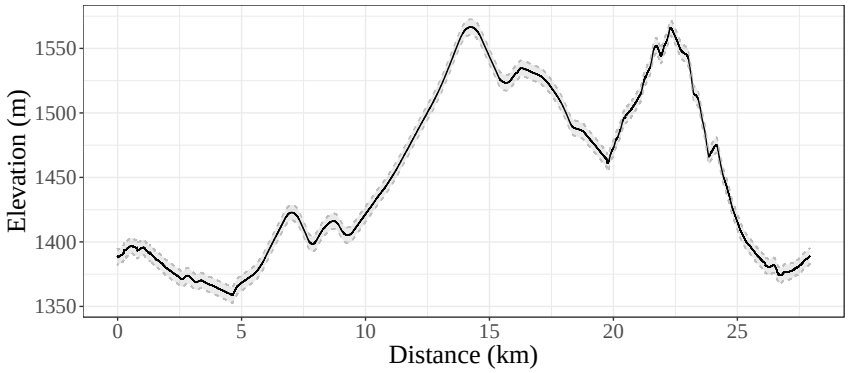


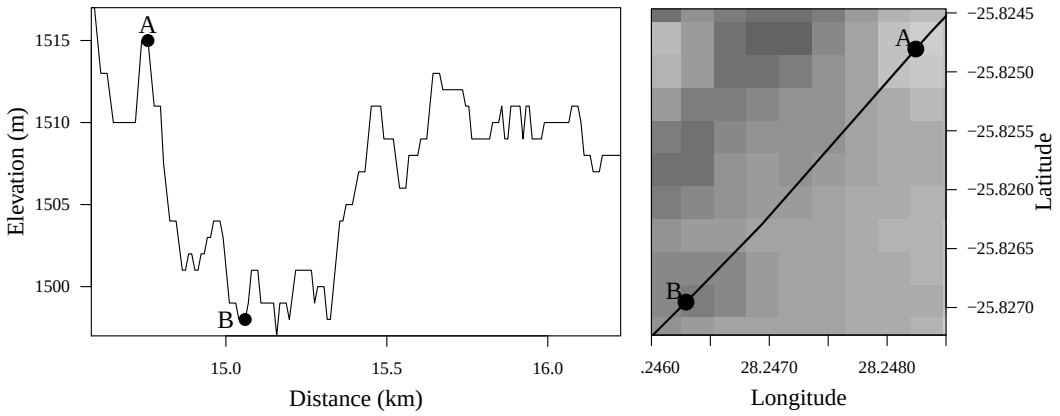
Figure 3.10: Second run elevation profile of trip with 95% CEP in light grey.

to the VBOX values (the second run) using a DEM as reference. The sample route's points can again be used to add DEM height values onto to build a reference road profile.

3.4 Road profile generation with DEM and smoothing

The Shuttle Radar Topography Mission (SRTM) DEM elevations are extracted from the open source United States Geological Survey (USGS) database, and a topographical visual of a single coordinate block is shown in Figure A.1 in the Appendix. From the SRTM raster the 30 m×30 m elevation block (pixel) is extracted for each coordinate of the sample road. The elevation value for each block is the lowest measured elevation within the 900 m^2 bounds, creating a stepwise nature when using roads that flow though these blocks.

To visualise the stepwise profile of a road, Figure 3.11 shows an arbitrary 2 km piece of the sample route. Figure 3.11a on the left is the side profiles with the elevations as a function of the 2 km distance. Figure 3.11b on the right is there to illustrate how the 30 m×30 m blocks can look from above with each block shaded by its respective elevation value. The letters A and B denote points on the road to show the difference of two extreme heights.



(a) Elevation profile as with SRTM elevations.

(b) Elevation blocks from above.

Figure 3.11: Stepwise SRTM DEM of a piece of the sample road.

Although a DEM creates a realistic road profile, it can sometimes provide incorect values when one has a bridge or overpass on the surface. At these points, the lowest

elevation is wrongly presented (Wood et al., 2014a). Unrealistic drops in these cases can cause incorrect road grades to be calculated.

To get rid of unwanted errors such as incorrect road grades from large elevation differences, one can use a maximum difference in elevation to set a boundary of what is an acceptable elevation value for a given point. TxDot (2018) provides a workable maximum road grade since roads more or less do not exceed threshold road grades of 15% on local roads and 8% on highways (not including ramps). The road grade Θ is calculated using (3.2)

$$\Theta = \left| \frac{\Delta Elv_t}{\Delta Dist_t} \right| < \text{threshold} \tag{3.2}$$

where ΔElv_t is the elevation difference between point t and its preceding point $t-1$; $\Delta Dist_t$ is the geographical distance between point t and $t-1$.

The entire process to end up with smoothed DEM data follows the process illustrated in Figure 3.12. After the SRTM DEM data is appended to the GPS traces, the raw elevations are downsampled to have uniformly spaced intervals. Each point is represented by a mean value since this will decrease the complexity of coordinates that are not equally spaced. For example, the captured coordinates will have unnecessarily too much information when a vehicle is moving slow or in rest.

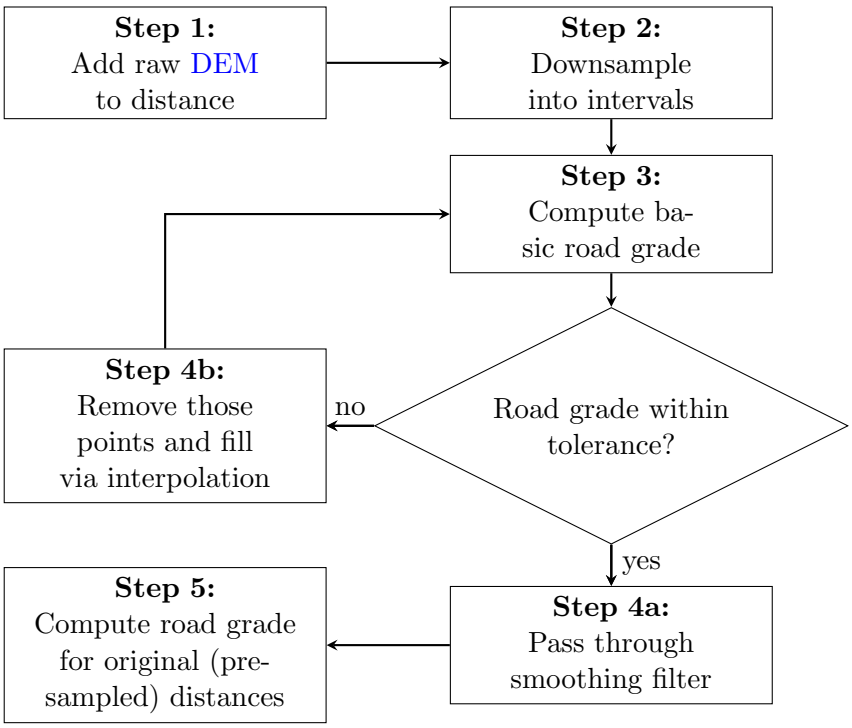


Figure 3.12: Step-by-step process of appending the smoothed DEM to coordinates.

Equation (3.2) provides an easy road grade calculation method as a derivative of elevation and can by used to easily find the erroneous points on the road. The loop from Step 4b is there to account for these instances where the tolerance is exceeded. Such points are then discarded and filled via interpolation.

A quick examination of the sample road's SRTM road grades finds 19.3% of the points steeper than the 0.15 threshold. These occurrences may be because of bridges and overpasses (both which occur a number of times along the 28km route) or caused by errors in

the DEM capturing. A liberal 0.2 is used as the tolerance value in this report and Figure 3.13 shows how new elevations (in red) were added to the coordinates via interpolation in Step 4b.

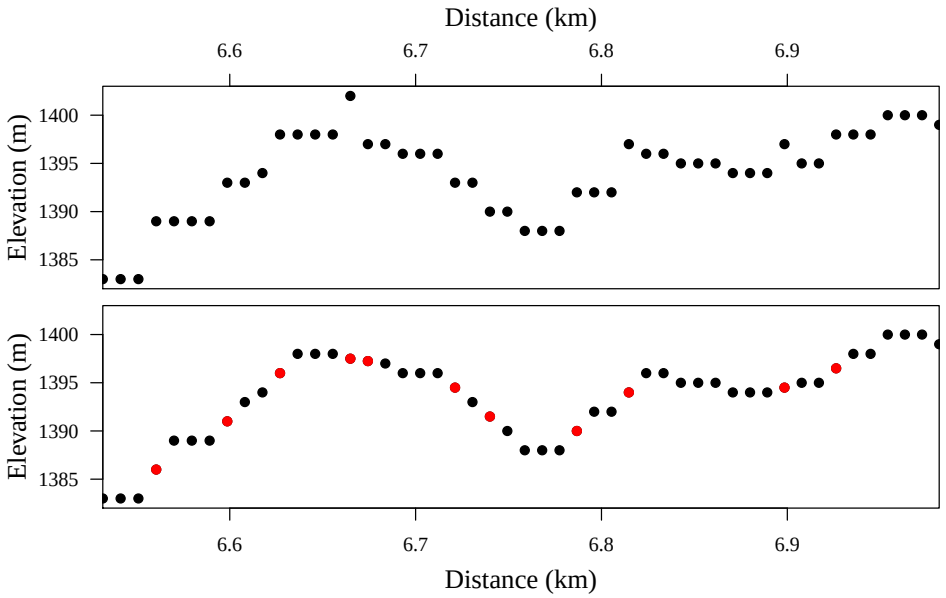


Figure 3.13: Removing of bad points and filling via interpolation.

Before the two smoothing methods can be compared, the reference profile to which the accuracy of the filter is tested should be determined. The VBOX and multiple runs both have elevation profiles, but one of them is closer to the SRTM elevation profile than the other. The VBOX is visually the closest to the SRTM profile as seen in Figure 3.14 and is used to test the parameters of the two smoothing filters.

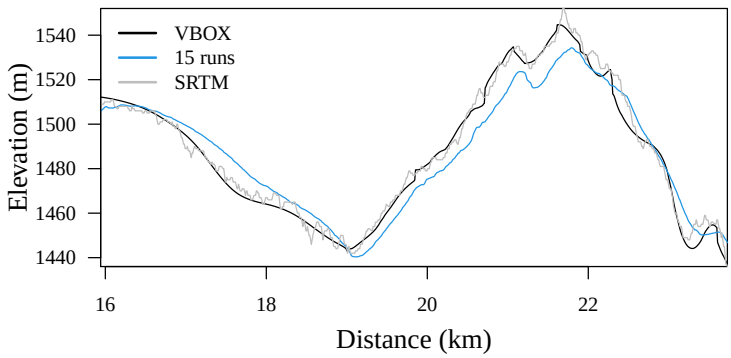


Figure 3.14: Comparing the elevation profile of the VBOX, multiple runs, and SRTM on a road segment.

The SG and spline are compared in Step 4a. With both these techniques, a polynomial order of three is used. A third order polynomial is the lowest order that can approximate linear, quadratic, and cubic relationships. For cubic splines, estimating elevation changes between the top and bottom of roadways can be supported by the capability of a low third order as well. The filter length is the tested parameter with the SG filter and the degrees of freedom is the spline's parameter.

Figure 3.15 illustrates how three different settings of parameters are tested against the

Table 3.1: Parameter testing of three different levels of smoothing.

	SG filter			Cubic Spline		
	light	**medium**	**high**	**light**	**medium**	**high**
Parameter value	31	101	201	200	90	20
RSS/1 000	41.688	**36.613**	41.825	41.991	**35.660**	44.087

VBOX elevations, producing three different levels of smoothing and Table 3.1 shows the filter length or degrees of freedom versus the Residual Sum of Squares (RSS) it produces. The RSS is calculated as the summed squares of all the points' absolute differences between the VBOX elevations and the smoothed DEM.

The lowest RSS with the SG filter is found with an interval length of 101. This interval length can also be noted as being about 28 times smaller than the total length of the points on the sample road. When the curves are visually compared, the same SG filter length also produces the best curve next to the SRTM elevations.

The cubic spline gives the lowest RSS value overall, but the parameter is more sensitive than that of the SG filter. This sensitivity can be seen in how much the spline curves in Figure 3.15 differ when the amount of smoothing is altered.

Because of the cubic spline's sensitivity, the better option would be the SG filter. A medium amount of smoothing with the filter would also produce a curve that represents the *true* elevation profile of a road most accurately.

3.5 Road grade calculation

Figure 3.12 ends with Step 5, determining road grade, now that the smoothing parameters for the SRTM elevations are established. Two methods are discussed in this section to find the most efficient technique. The interval sizes of the methods are evaluated aiming to find the best size for road grade calculations. The first method is to derive a grade profile as a simple derivative, similar to equation (3.2) and linear regression is implemented for the second method.

3.5.1 Road grade as a simple derivative of elevation

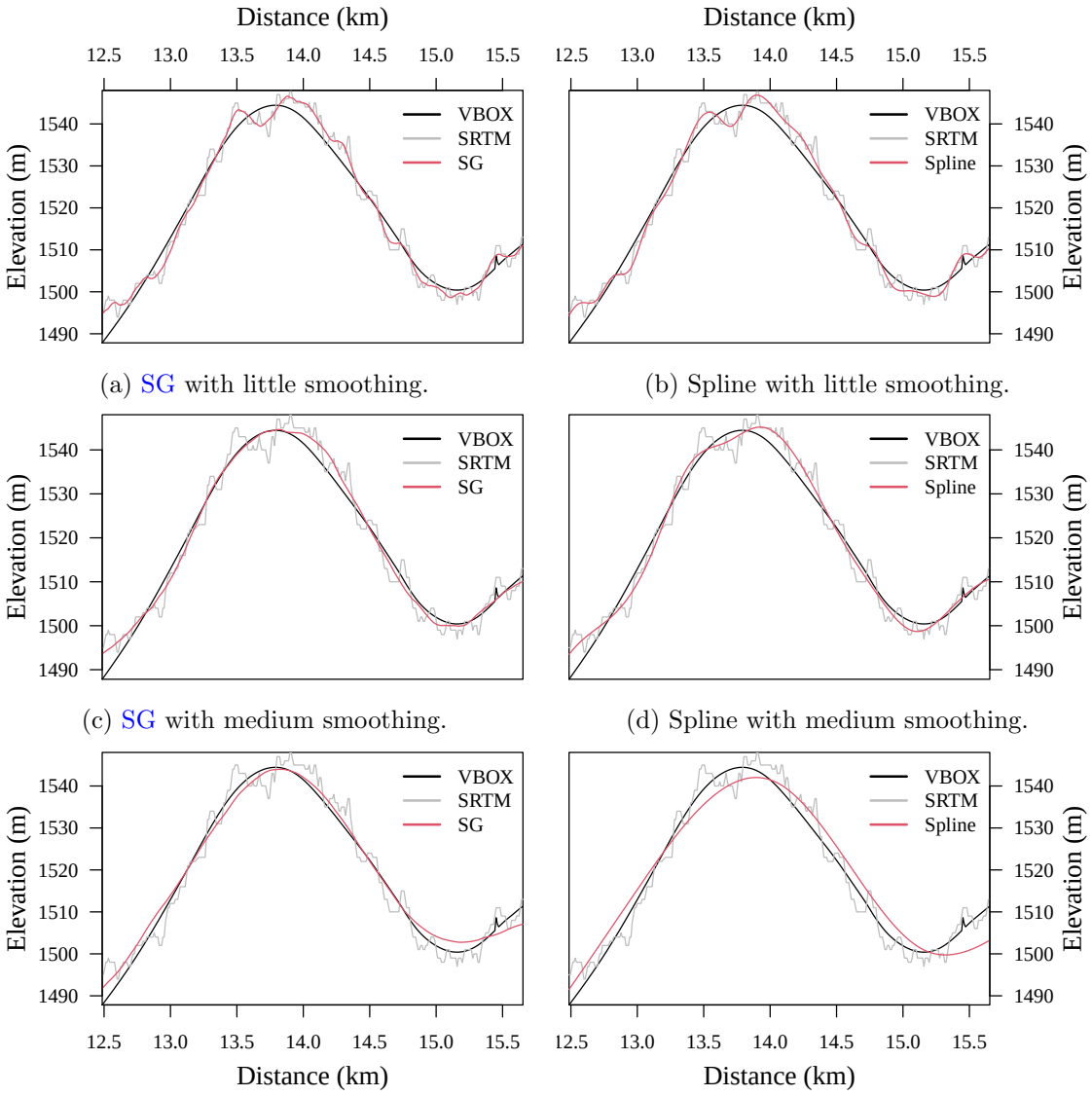
Using the VBOX elevations, road grade can be calculated as a simple derivative of the elevation as a function of the distance travelled between points. Studies such as the one by Magrath and Brady (2017) suggest that $80m$ is the best interval size, but what happens when it is decreased? Increasing the size will distort the derivative and valuable road geometry might be lost, but the sample road's dataset has a coordinate on every $10m$ so a decreasing of the interval size might work well.

Equation (3.2) sets out the basic road grade calculation as a basic derivative of elevation and distance. To test different distances (interval sizes denoted by int), the equation can be adjusted using (3.3)

$$\Theta_i = \frac{e_f - e_i}{int} \qquad \forall \quad i \in \{1, 2, \ldots, 2800\}, int \in \{30m, 40m, \ldots, 80m\} \tag{3.3}$$

where Θ_i is the road grade and e_i is the elevation at the point of concern, i, and e_f is the elevation of the point on the road, int meters on.

All the points between i and f are given the same road grade value as point i. For example, an interval size of $40m$ will provide a set of road grade data that look as follows:



(a) SG with little smoothing.

(b) Spline with little smoothing.

(c) SG with medium smoothing.

(d) Spline with medium smoothing.

(e) SG with a lot of smoothing.

(f) Spline with a lot of smoothing.

Figure 3.15: SG filter versus cubic spline. Both are compared to the VBOX and SRTM elevations with a small, medium, and large amount of smoothing.

$[\Theta_1, \Theta_1, \Theta_1, \Theta_1, \Theta_5, \Theta_5, \Theta_5, \Theta_5, \Theta_9, \dots]$. The tested intervals are in steps of $10m$ and range from $30m$ to $80m$. For all six of these intervals, road grade values are calculated using (3.3).

The lowest number of points (at about 2%) fall in the *flat* road grade category when the interval size is $80m$. A road is regarded as more or less flat when its road grade is between -0.05 and 0.05, typically represented by the highways. The sample road's highways account for 60% of the route and an interval of 80 m only having 2% within the flat group proves that this derivative method significantly misrepresents the grade profile of the road.

Other distances are tested in the same manner and only two interval sizes prove to accurately estimate the route's elevation profile: 50 m and 60 m. Either of these will be useful in road grade calculations.

3.5.2 Road grade with linear regression

Linear regression is similar to the derivative at a point, but is not so instantaneous in nature. More than one point can be used to fit a regression line and then draw the derivative off from that line for a road grade value. Here, a point is used with two points before and two points after, each 10 m apart. A regression expression is fit to these five points and the coefficient (road grade) is extracted. This is done for every equidistant point on the road, not like the simple derivative where the original points were given the same road grade as the interval they were in.

The set of all values is found in $\{1, \dots, \mathrm{I}\}$, but the first two and last two grade values are not calculated, therefore, a set is given as $\mathbf{I}$, where $\mathbf{I} = \{3, 4, \dots, \mathrm{I}-3, \mathrm{I}-2\}$. For each elevation point in $\mathbf{I}$ a set of values $\mathbf{X}$ are used:

$$\mathbf{X}_i = \{x_{i-2}, x_{i-1}, x_i, x_{i+1}, x_{i+2}\} \qquad \forall i \in \mathbf{I} \tag{3.4}$$

Each linear regression y_i is then calculated using (3.5)

$$y_i = B_i + C_i \mathbf{X}_i \qquad \forall i \in \mathbf{I} \tag{3.5}$$

where B_i and C_i denote the intercept and coefficient, respectively. The road grade is represented by the coefficient C_i, providing a value for each coordinate with its elevation.

To decide if linear regression is a better option compared to a simple derivative, the interval sizes are considered. A simple derivative will average out an interval's road grade, where the linear regression will provide each point with its own road grade value. This provides a smoother and more precise road grade curve when surrounding points are not averaged out, but rather used in a regression expression that changes. The choice is thus to have linear regression be the method to use from this point onwards and will be used on a different dataset in Chapter 4 hereafter.

To visualise how the final road grades pair with their respective elevations, the sample road's smoothed elevation profile is plotted in Figure 3.16. The smoothing is done by an SG filter and the road grade is derived by linear regression using sampled points of 10 m apart and afterwards interpolated for all the points in-between. To inspect the accuracy, a quick glance at the zero-line of the road grade plot shows that downhill grades do indeed produce negative road grade values and the same other way around. This road grade model can now be used with any set of data that contains consecutive lateral and longitudinal coordinates.

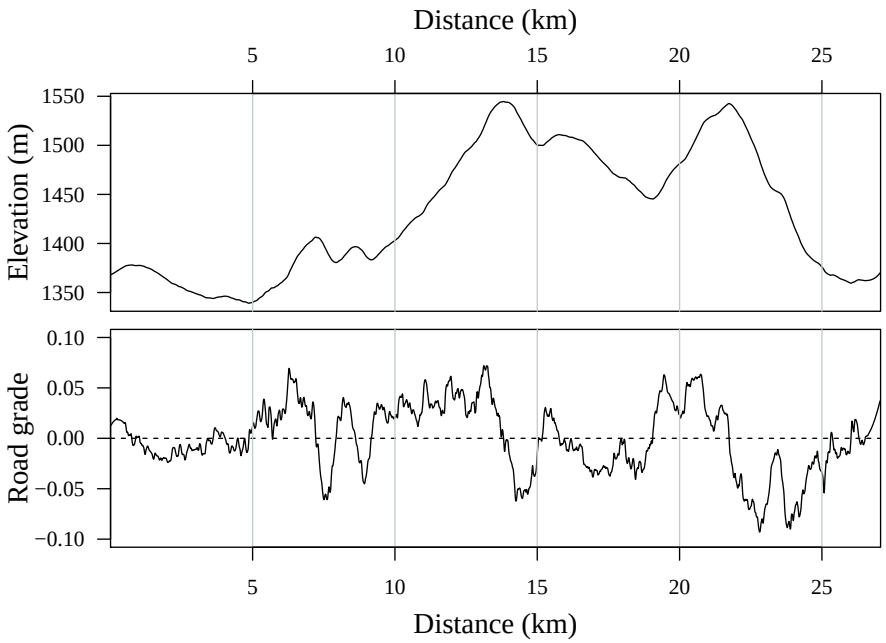


Figure 3.16: Final smoothed SRTM elevation profile and road grades of the sample road.

Chapter 4

Results and discussion

This chapter serves as an analysis of the results of the addition of road grade to a driver behaviour model. The model is first discussed and ran as-is on the current truck accelerometer data. After the road grade model has provided the coordinate traces with slope values, the model is run again, but this time in categories of road grades. These categories allows one to study the differences in behaviour on different road steepness levels. Individual driving performance scores and rankings are also calculated on the different grade levels to analyse how performance can change when the road geometry changes.

4.1 Driver behaviour model from literature

The model of Joubert et al. (2016) measures relative performance of drivers, comparing a driver's behaviour to a pool of drivers and thereby defining their risk level. The authors use speed and horizontal acceleration as determining factors to classify these risk levels. Four risk levels are set out using a quantile vector as $\mathbf{Q} = \{q^{none}, q^{low}, q^{med}, q^{high}\}$ and range from 0 to 1. The *risk space* is the discretised volume using the three dimensions of acceleration. In their model, a rhombic dodecahedron tessellation is used as the discretisation of choice. Each vehicle's records are individually sorted into the discretised cell in which its acceleration values are contained. When done, the cells are ranked from highest (most records) to lowest (fewest records) and the cumulative percentage of the representation records is calculated.

There is a comparison of a single driver to the population of drivers. The total number of records is represented as $R = \sum_{i=1}^{C} r_i$, where r_i is the risk of cell i, or simply the number of observed records in cell i. The argument holds that the more common a specific acceleration profile (a high number of records in that cell), the lower the risk associated with that cell. A decreasing order of risk is set, i.e. $r_1 \geqslant r_2 \geqslant \ldots \geqslant r_C$.

Each risk category is split at a, b, and c, all within the bigger set, $\mathbf{C}$, that contains all the drivers in the dataset. To find these three points, the cumulative number of records for each risk space are set out as follows.

- No risk:

$$\frac{\sum_{i=1}^{a} r_i}{R} \leqslant q^{none} \text{ and } \frac{\sum_{i=1}^{a+1} r_i}{R} > q^{none} \tag{4.1}$$

  with $i \in \mathbf{C}^{none} = \{1 \ldots a\}, \mathbf{C}^{none} \subseteq \mathbf{C}$

- Low risk:

$$\frac{\sum_{i=1}^{b} r_i}{R} \leqslant q^{low} \text{ and } \frac{\sum_{i=1}^{b+1} r_i}{R} > q^{low} \tag{4.2}$$

with $i \in \mathbf{C}^{low} = \{a+1, a+2, \ldots, b\}, \mathbf{C}^{low} \subseteq \mathbf{C}$

- Medium risk:

$$\frac{\sum_{i=1}^{c} r_i}{R} \leqslant q^{med} \text{ and } \frac{\sum_{i=1}^{c+1} r_i}{R} > q^{med} \tag{4.3}$$

with $i \in \mathbf{C}^{med} = \{b+1, b+2, \ldots, c\}, \mathbf{C}^{med} \subseteq \mathbf{C}$

- High risk:
  $i \in \mathbf{C}^{high} = \{c+1, c+2, \ldots, C\}, \mathbf{C}^{high} \subseteq \mathbf{C}$

The levels $\mathbf{Q}$ mentioned before were tested by the authors to find a visual balance of all four categories. They chose $\mathbf{Q} = \{0.960, 0.994, 0.999, 1.000\}$ because of its appropriate balance between the levels. So, one can interpret that as "96% of records fall in those (green) cells that are most commonly observed". That means that 96% of all the records make up the *no risk* portion of the risk space. Another parameter they had to choose was the cell size since it affects the discretisation of the risk space, and $10mg$ was the size they found to be ideal to have clear separation between risk categories yet capture the true shape of the risk (sub)spaces. Figure 4.1 is extracted from their paper where the risk space was constructed using light vehicle accelerometer data and a horizontal slice taken at $1009mg$ for visualisation. Green, yellow, orange, and red represent no risk, low risk, medium risk, and high risk records respectively.

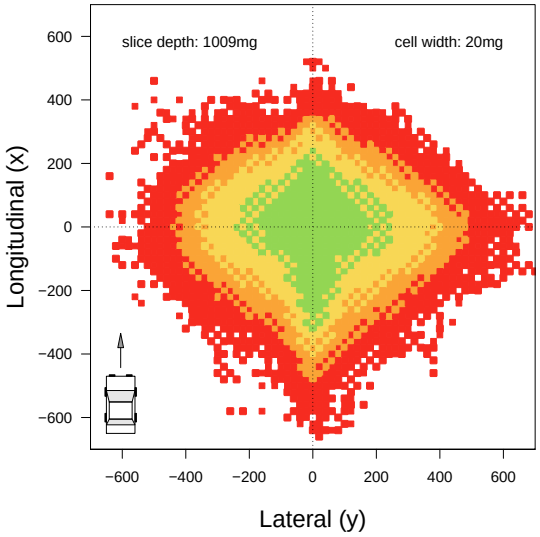


Figure 4.1: Risk space of light vehicles by Joubert et al. (2016).

Joubert et al. (2016) use the risk quantiles in $\mathbf{Q}$ to determine risk categories based on acceleration. The authors do, however, also use speeding as a second determining factor, but with specific speed thresholds, not quantiles. They give speed and acceleration each a weight (for example $40\% : 60\%$) and show that a weight below 20% for speed does not affect any overall risk scoring of drivers. The opposite is also true when the weight of acceleration is below 25%. Finding the risk space is ultimately a trade-off between speed and acceleration, it is up to the insurer (as in the article) to decide what weight to give each aspect. This dissertation only focusses on acceleration and does not incorporate speed as well. It is done to investigate a behavioural change (if any) of a driver when the slope

of a road changes, and erratic lateral and longitudinal movement is what will provide the best indication of this behaviour. Horizontal speed can be distorted when a steep slope is investigated since the vertical component can be lost when only using relative Global Positioning System (GPS) speed.

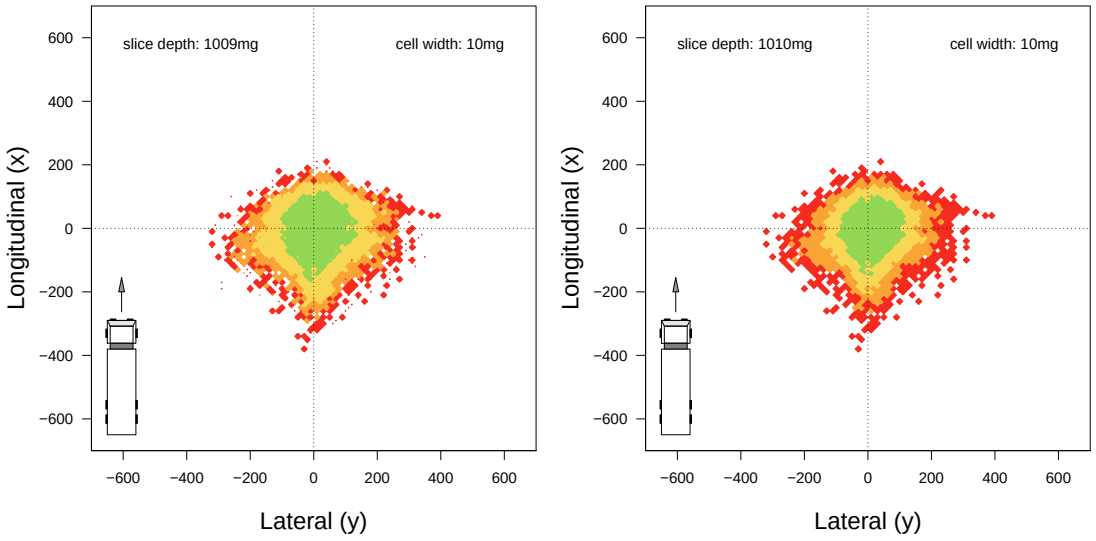
When one has the accelerometer data and the risk quantiles and cell sizes, an individual's proportion of records within each risk category can be determined. These proportions within each category are represented relative to the total population as $p^{none}, p^{low}, p^{med}, p^{high}$. The weighting of each of these categories are then denoted by $w^{none} = 0, w^{low} = 1, w^{med} = 2, w^{high} = 3$.

Using the weightings and proportion of records, a score, s_i, is calculated for each individual i as follows.

$$s_i = 3 - (p^{none}w^{none} + p^{low}w^{low} + p^{med}w^{med} + p^{high}w^{high}) \quad \forall i \in \mathbf{P} \tag{4.4}$$

Each driver's risk score is normalised to a range within [0,1] and thereafter sorted against all the scores of the population from best to worst. The scoring on the data from this report is done in Section 4.5.

The existing model of Joubert et al. (2016) is run on the truck dataset and the risk space is visualised in Figure 4.2a. To visually balance the new risk space, the cancellation of gravitational acceleration is changed by setting the slice height to $z = 1\,010mg$ and the quantiles are changed to $\mathbf{Q} = \{0.940, 0.985, 0.997, 1.000\}$. This new risk space of all the records are seen in Figure 4.2b.



(a) $\mathbf{Q} = \{0.960, 0.994, 0.999, 1.000\}$ (b) $\mathbf{Q} = \{0.940, 0.985, 0.997, 1.000\}$

Figure 4.2: The changes in the risk space quantiles and slice depth to fit this study's vehicle data.

When the new risk space visualisation is compared to the one in Figure 4.1, one can see that the extent of truck acceleration records is smaller and less spread out than that of light vehicles. This is quite plausible given the size and weight of a truck. There is much less harsh movement from a heavy vehicle than one would observe with a passenger vehicle. The more inertia, the more power is needed to slow down or speed up.

Another reason for a smaller extent of the risk space is the homogeneous dataset. The vehicles from the dataset are from the same company and are more or less the same age.

Recorded behaviour would, therefore, be mostly impacted by the drivers and not by the type of vehicle. A dataset of passenger vehicles would produce a dataset with a lot of different types of vehicles, all with different capabilities of speed and acceleration.

These factors can all impact the risk profile of a group of drivers, or fleet of trucks in this case. To find how much, if at all, road geometry impacts driver behaviour, one can add another variable to the dataset: road grade.

4.2 Appending the smoothed Digital Elevation Model (DEM) to the GPS traces from truck accelerometers

With the goal of having more information of the road geometry, in the form of slopes, one can add elevations to the pool of variables in the dataset. This dissertation uses a dataset with 45 trucks and their movement within a month. The useful variables of the dataset at this point are: vehicle ID, time-stamp, longitude, latitude, x-acceleration, y-acceleration, and z-acceleration.

Finding the road grades of the roads driven by the vehicles, one needs to first extract elevation values from the digital Shuttle Radar Topography Mission (SRTM) dataset by using the latitude and longitude of the traces. The final estimated elevation values are found when the SRTM elevations are smoothed following the Savitzky-Golay (SG) filter and its smoothing parameters of the previous chapter. When each truck has a set of final smoothed elevations and coordinates, only then can road grades be derived.

The roads travelled are separated for each truck since start and end points should not be mistaken between the trucks. For each road of each truck, the road grade values are derived by linear regression as previously detailed in Section 3.5.2. The regression expression is fitted to the elevation relative to the distance travelled, with two points before and two points after each point.

After road grade values are appended to the original dataset, the following parameters are available to formulate the drivers risk model: vehicle ID, time-stamp, longitude, latitude, x-acceleration, y-acceleration, z-acceleration, and road grade.

4.3 Road grade as a variable

The behavioural model of Joubert et al. (2016) allows for one to analyse contextual variables and the impact they might have on the overall risk space of a dataset. Road grade is used as the contextual variable here, and five different categories are split up to analyse the results.

The five sets are set up as follows: steep downhill, medium downhill, flat, medium uphill, and steep uphill. Because the dataset used in this study is from truck accelerometers, one needs to consider how the steepness categories will be split up with this type of vehicle, also remembering that these vehicles drove on highways for the most part. It cannot simply be accepted that the threshold values given by literature (as in Equation (3.2)) would be sufficient. TxDot (2018) states that anything between -8% and 8% can be considered as medium steep and between 8% and 15% can be considered as steep, on the negative side as well. What is considered as steep is not the same when the dataset is mostly on highways, therefore, the threshold values of 8% and 15% are not applicable for this study and new ones need to be defined.

A useful and common method to split a variable up into categories, is to separate the variable at three standard deviations from the mean (which would be 0 in this case). This

method is known as the empirical rule or the 68-95-99.7 rule. Rounding the standard deviation of the dataset's road grade values to a useful decimal, a value of 0.02 is found to be sufficient to determine the different levels of steepness. Table 4.1 sets out the road grade values and the percentages of records within those limits. The overall distribution

Table 4.1: Five categories of road grade on truck dataset.

		Percentage of records	
Category	**Road grades**	**Medium smoothing**	**High smoothing**
Steep down	$\theta \leq -0.04$	2.6%	2.2%
Medium down	$-0.04 < \theta \leq -0.02$	8.1%	9.9%
Flat	$-0.02 < \theta \leq 0.02$	72.6%	73.5%
Medium up	$0.02 < \theta \leq 0.04$	13.7%	11.1%
Steep up	$0.04 < \theta$	3.0%	3.3%

of records between the different road grade categories might be impacted when the initial smoothing parameters are changed within the SG filter. To test the model's sensitivity to the smoothing filter, the level of smoothing is changed to *high*, i.e. the filter length is shortened. The distribution is shown in the last column of Table 4.1 and shows how much the percentages of records change when the level of smoothing is adjusted too high. Reducing the amount of smoothing is not added to the comparison as it proved, in the previous chapter, to not be accurate enough when estimating the road profile.

An average difference of percentage of records of the two columns is 1.2% and is not enough to conclude that the smoothing parameter has a significant effect on the distribution. Since the distribution is not impacted as much, it is sufficient to continue using a medium level of smoothing. Changing it to a high level of smoothing would mostly just increase the computational cost of the model without being worth it.

4.4 Driver behaviour model with road grade as a contextual variable

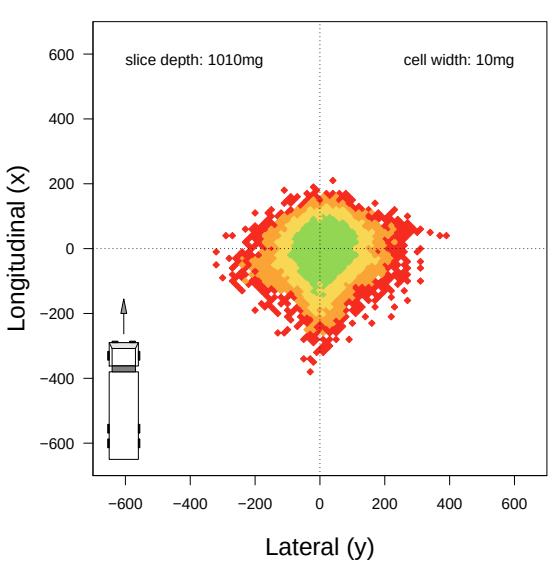
With five categories defined, they can be used within the risk model set out in Section 4.1. The risk model filters the records to only use those that fall within each of the grade categories, providing a risk space for each.

4.4.1 Flat roads

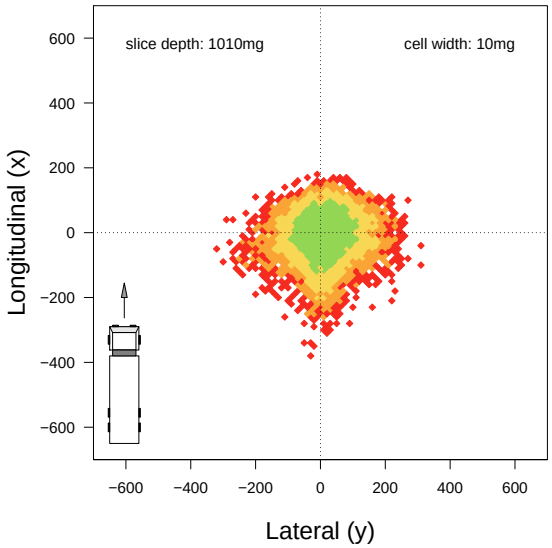
The first road grade level to be visualized is flat roads in Figure 4.3b. It is the largest category and does not differ much from the risk space of entire population of records shown in Figure 4.3a. The outer, higher risk rings of these two groups are seemingly identical which is expected with 72.6% of records in the flat category.

4.4.2 Downhill roads

When a risk space is built from the records that associate with downhill roads, one sees more braking and, consequently, a space that has more negative longitudinal acceleration in Figure 4.4. For steep downhills, the majority of records lie in the negative acceleration area and for medium steep road segments, there are even more records in this area. The data, therefore, shows that the truck drivers are braking downhill for the most part. Again,

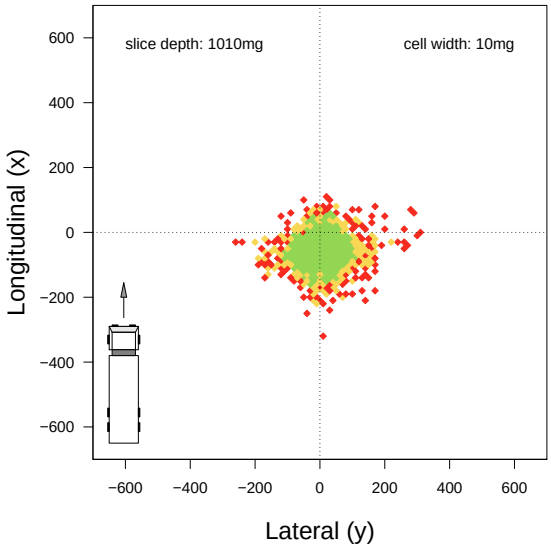


(a) All records.

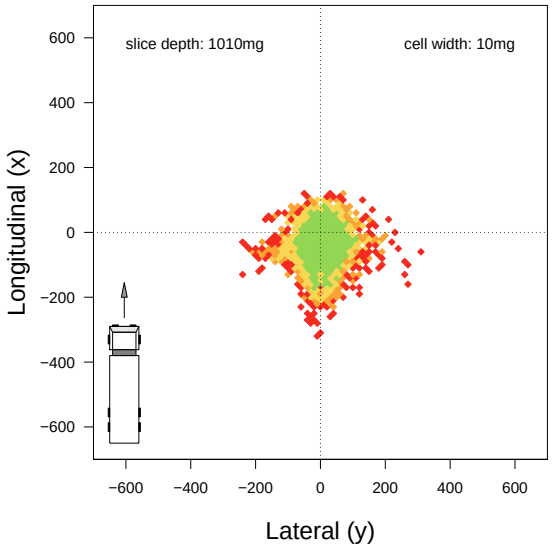


(b) Risk profiles of entire population and flat grades.

Figure 4.3: All vs flat.



(a) Steep downhill.



(b) Medium steep downhill.

Figure 4.4: Risk profiles of steep and medium steep downhill grades.

this is quite plausible since heavy vehicles rely on engine braking and lower speeds. In fact, in many cases it is regulatory that heavy goods vehicles change to lower gears when going downhill.

Anticipation is a basic aspect for any driver, especially when you are more experienced and braking with a heavy vehicle downhill can be wise when one anticipates a fast downhill drive. Reducing your speed rather than trying to keep it constant might be an overcompensation, but it can be the safer option.

Lots of trucks, if not all these days, are fitted with an endurance brake, that is designed to keep the vehicle slow when going downhill over a longer period of time. This is to reduce *brake fade* that is caused by manual foot-braking as brake-fade can cause a driver to loose control when driving with a load downhill. Professional drivers also make use of engine braking, keeping the truck at a lower gear at high engine revolutions that also limits the speed. All of this compensates for the possible forward acceleration that can take place downhill, but does not provide a reason for such a large number of records with negative acceleration.

The braking is most likely due to overcompensation for the safety risks of a driving downhill at high speeds with a heavy vehicle. Drivers might be decelerating even more to start reducing their speed, rather than trying to keep it constant.

For the lateral movement of the vehicles, there are a few high risk records here and there around the central *blob* and these records can be cases of drivers swerving sideways to avoid collisions. They could have swerved around slower trucks to pass them downhill, but passing downhill can cause problematic instances such as speeding or overheating of the brakes. Both of these cases are unsafe with a large vehicle, especially if it has a heavy load, and could result in a loss of control and an accident.

4.4.3 Uphill roads

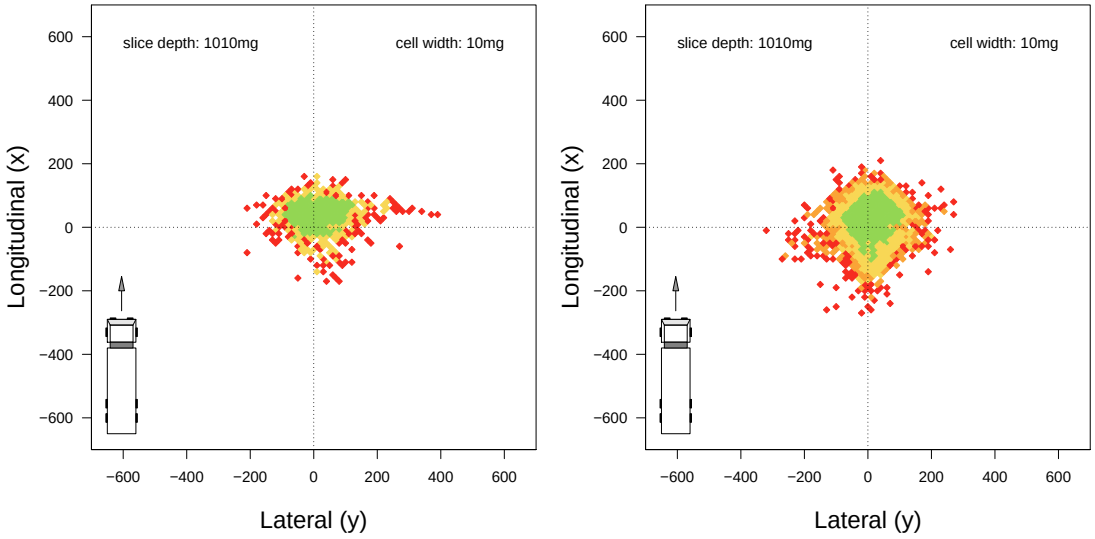
The last behavioural comparison that is studied is that of drivers on steep and medium steep uphill roads. Figure 4.5 illustrates the risk spaces of these two different types of uphill grades and it is visible that both risk spaces have the majority of their records in the positive forward acceleration area.

Sometimes it is wise to get out of optimal fuel or speed ranges for a while to anticipate an event, in order to save fuel for an entire trip. For example, a driver can gather speed before an important peak on the hill s/he is on and this will save the extra power that would have been used to go over the peak if s/he was driving at a constant speed. This idea of increasing speed before a very steep road segment is reflected in Figure 4.5b as the number of forward accelerations are significantly higher.

When analysing the lateral movement, a tail of observations is found in Figure 4.5a. These records come from vehicles that were accelerating faster to the right and might be there because of them taking a gap in traffic to overtake other vehicles and, more specifically, other trucks. A lighter or more powerful truck can overtake an older, heavier, or less powerful truck. Passing on an uphill is much safer than on a downhill, so one would expect more records that look like swerving in this section than that of downhills.

4.5 Scoring and ranking of drivers

In the previous section it was shown how road geometry impacts the overall behaviour of drivers and that, indeed, the driver behaviour profile differs for different road grade categories. This section focusses on the behaviour and performance of individual drivers



(a) Steep uphill

(b) Medium steep uphill

Figure 4.5: Risk profiles of steep and medium steep uphill grades.

on different road grades. Their individual driving performance is scored and ranked against one another and is then investigated to see how, if at all, their performance changes when driving on different road grades.

Each driver is scored using Equation (4.4) and their scores are then normalised on a range of $[0; 1]$, where the individual with the lowest score and worst performance receives a zero, and the individual with the highest score and best performance receives a 1. The scores are then ranked from best to worst and the drivers each get a ranking number. This scoring and ranking is done for the entire population of records, but the records are also filtered on all five road grade categories and then modelled, scored, and ranked again. All the ranking values are found in the appendix in Table A.1. By ranking drivers on the different categories, one can see if a driver's rank changes against the pool of drivers after road grade is taken into consideration.

To examine the overall changes in performance, the rankings of individuals on the base model (denoted by the x-value) are plotted against the rankings of the same person when filtered on the five different road grade categories (denoted by the y-value), as with the previous section.

The first comparison is between the base model and flat roads, as plotted in Figure 4.6. Points on the diagonal line are those individuals who have constant ranking positions. The points above the line represent those who have gained ranking by not applying any filtering on road grades and using all the records. The points below the line represent those who have gained ranking by including road grade in the risk model. The correlation coefficient, R, in the comparison with flat grades is 0.89. This high R value shows that the majority of drivers do not experience a change in rank positions. With more than 70% of the population of records on flat roads, this can be expected. The notable outlier is the individual who, when we disregard the road grade, ranked 16th. When road grade is taken into account, the same vehicle's ranking drops to 43rd on flat road grades.

The same is done for the rankings on downhill road segments and the base rankings as plotted in Figure 4.7. The R value is calculated as 0.58 and 0.38 on steep and medium steep downhill grades, respectively. Both of these coefficient values are too low to represent a relationship between the two ranking axes. For steep downhills, there are however little

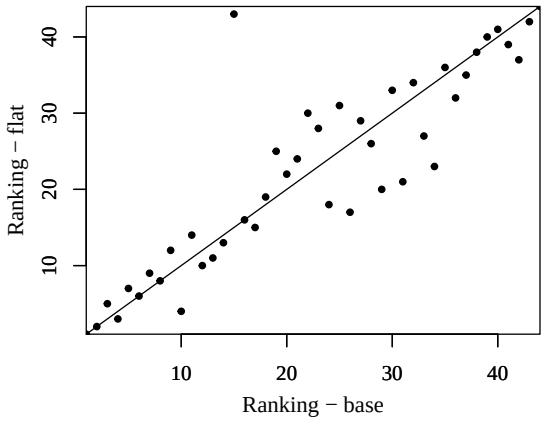
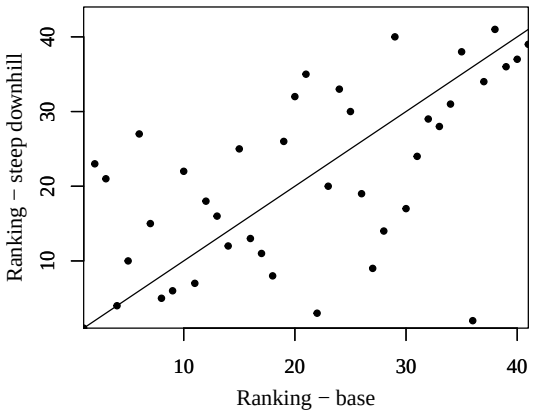
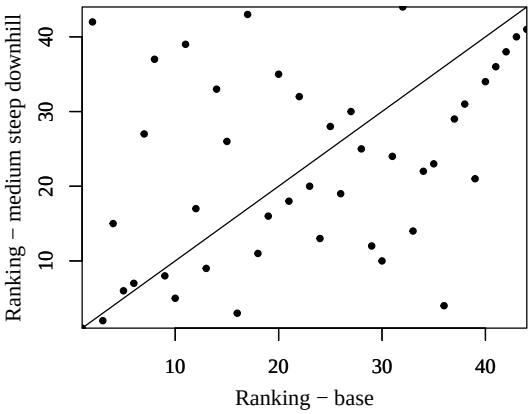


Figure 4.6: Comparing change when filtering on flat road grades.



(a) Steep downhill grades.



(b) Medium steep downhill grades.

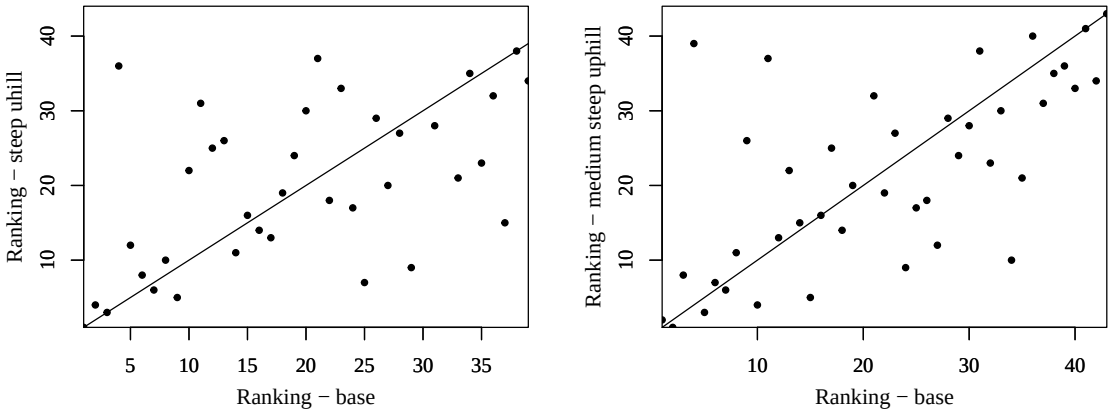
Figure 4.7: Comparing change when filtering on downhill road grades.

to no points where an individual has a drastic change in position. On medium grades, quite a few drastic changes are observed. Overall, one can see that an individual is not bound to have the same ranking on downhill grades that they would have had on the base risk model.

On uphill road segments, the R value becomes 0.54 and 0.66 on steep and medium steep grades, respectively, and the plotted correlation is visualised in Figure 4.8. The plots are very similar to that of downhill grades, showing almost no correlation between the axes. The R values of uphill grades are however a bit higher than the values found with downhill grades. Although a few individuals have a constant ranking, most of their positions change.

While the above plots are visual representations of the individuals' changes in rankings, some people can be singled out for discussion. Table 4.2 sets out 10 drivers and their rankings within all categories. In the table it can be seen how some drivers have one or two categories that are significantly different to the rest.

For about a third of the drivers, their rankings would change when road grade is taken into account during the ranking calculations. For example, individual 17 has a ranking of 27 on the base model, but their rankings drop significantly on upward slopes. This indicates that they drove worse than the rest of the drivers on uphill roads, and this



(a) Steep uphill grades.

(b) Medium steep uphill grades.

Figure 4.8: Comparing change when filtering on uphill road grades.

Table 4.2: Ranking of drivers in normal risk model and in all road grade categories.

		Road grade considered				
Id	**Road grade ignored**	**steep down**	**medium down**	**flat**	**medium up**	**steep up**
1	9	27	13	7	24	29
17	27	6	7	33	23	28
21	15	7	4	17	14	37
26	7	11	6	5	6	25
28	10	5	30	12	34	8
33	4	4	36	10	10	2
35	21	3	39	31	35	33
37	3	22	16	4	5	3
39	36	39	41	35	39	4
40	37	40	8	42	11	21

information is lost when only using the base model.

A few more examples of drastic changes in rankings are listed below:

- Individual 1 only performs well on flat roads, but receives a good overall base ranking;
- Individuals 28 and 33 have big differences in rankings, mostly because of their poor performance on anything other than steep up- and downhill roads;
- Individual 35 performs exceptionally bad on all roads except steep downhills, and 39 the same, but for steep uphills;
- Individual 40 receives a very poor overall ranking, but has improved rankings on medium steep up- and downhill roads

Again, it is seen that road grade impacts a driver's risk score and ranking against other drivers. This is not only reflected in the combined records of performance rankings, but also when examining individual ranking differences when the model is run on data that was filtered according to road grades.

Chapter 5

Conclusion

The motivation to investigate the impact of slopes on driver behaviour comes from the understanding that road geometry can influence a vehicle and its performance. Existing literature does not show how much road grades can influence *behaviour*, but it does show how it influences a vehicle's performance. This performance can be anything from higher emission rates to higher accident rates on inclines, leading to the question: how much, if at all, do inclines influence the behaviour and risk of a driver?

To estimate the road grade values for a given road, Shuttle Radar Topography Mission (SRTM) elevation data is appended to a set of truck accelerometer traces from the movement of 45 trucks over the course of a day. The step-wise nature of the SRTM elevation profile is smoothed by a Savitzky-Golay (SG) filter and then derived by linear regression to find road grade values.

The road grade values are used within an existing risk model from the literature to analyse the behaviour of truck drivers on different types of inclines. The input accelerometer records of the risk model were filtered on different categories of road grade: flat, medium steep uphills, steep uphills, medium steep downhills, and steep downhills. In doing this, behavioural data was found for the five different grade types and individual risk scores were determined for both the base model and for all five categories.

The visual risk spaces of all the records showed that drivers (of heavy goods vehicles) tend to decelerate when on downhill roads and accelerate when on uphill roads. A reason for these two occurrences can be because drivers anticipate the safety risk of speeding downhill and would rather brake, and they anticipate the extra energy needed to push through a steep peak on an uphill coming ahead and would rather pick up speed earlier to avoid using more power on the peak. Section 5.1 sets out how one can approach the study of lighter vehicles – one might expect the same behavioural occurrences as with heavy vehicles, but the wide variety of vehicle types might influence this assumption.

Moving on from the behaviour of the entire group, individual driving performance was also calculated in the form of rankings. An individual is ranked against all the other drivers within the dataset to examine performance relative to the performance of others. The rankings are calculated on the base model from literature as well as on all the different road grade categories. The results show that the majority of drivers have relative constant rankings, but that some drivers behave differently with changes in gradient. Therefore, road grade does not only influence the overall behaviour of a group of drivers, but it can also influence the behaviour of individuals in different ways.

The objective of this dissertation is not to decide how the risk of drivers on inclines should be dealt with when profiling their risk in an industry such as insurance. Instead, it shows successfully that discriminating on different road grades *indeed* affects the outcome

of a driver's risk behaviour. Ignoring road grade, therefore, implies that insurers may be estimating driver risk incorrectly in a Pay How You Drive (PHYD) schema.

5.1 Future work

To find more opportunities of study within the areas of behaviour and road grade, one can continue the investigation by adding more data, either by using the same vehicles over a longer period of time or by using a larger number of vehicles. This dissertation only used a day's travel of 45 heavy goods vehicles, but using more data can be beneficial to better understand the *scale* of the influence of inclines on a driver's performance. One can perhaps quantify this influence to a point where an insurer is able to include it in their risk analysis. After finding the scale of influence of all the different types of inclines, the question still remains: Should a driver be penalised/rewarded for their behaviour on different slopes, or should they only be judged on their *overall* behaviour? This implementation is of course up to an insurer, but can definitely change the way they quantify the risk of their drivers.

Another area that was not investigated in this dissertation is that of accidents and how driver behaviour and/or road grade might be causes of accidents. Knowing the potential correlation and relationship can be highly valuable for not only the insurance industry, but for local road safety.

When accelerometer data of various types of drivers can be paired with accident statistics, dangerous roads (or road types) can be identified. For example, if one finds a road that has a very steep downhill where a lot of vehicles experience accidents, one can investigate their behaviour and determine if the cause lies with the drivers or if the road geometry is the main cause. This steep road can then be flagged as potentially dangerous and further safety measures can be taken if necessary.

Bibliography

Bai, S., Thibault, J., and Mclean, D. D. (2006). Dynamic data reconciliation: Alternative to Kalman filter. pages 485–498.

Bener, A., Jadaan, K., Crundall, D., and Calvi, A. (2020). The effect of aggressive driver behaviour, violation and error on vehicle crashes involvement in Jordan. *International Journal of Crashworthiness*, 25(3):276–283.

Boroujeni, B. Y. and Frey, H. C. (2014). Road grade quantification based on global positioning system data obtained from real-world vehicle fuel use and emissions measurements. *Atmospheric Environment*, 85:179–186.

Boroujeni, B. Y., Frey, H. C., and Sandhu, G. S. (2013). Road Grade Measurement Using In-Vehicle, Stand-Alone GPS with Barometric Altimeter. *Journal of Transportation Engineering*, 139(6):605–611.

Bromba, M. U. A. and Ziegler, H. (1981). Application Hints for Savitzky-Holay Digital Smoothing Filters. Technical report.

Burden, R. L., Faires, J. D., Brooks/Cole Publishing Company., and Cengage Learning (Firm) (2011). *Numerical analysis*. Brooks/Cole, Cengage Learning, 9 edition.

Celaya-Padilla, J. M., Galván-Tejada, C. E., López-Monteagudo, F. E., Alonso-González, O., Moreno-Báez, A., Martínez-Torteya, A., Galván-Tejada, J. I., Arceo-Olague, J. G., Luna-García, H., and Gamboa-Rosales, H. (2018). Speed bump detection using accelerometric features: A genetic algorithm approach. *Sensors (Switzerland)*, 18(2).

Chen, X., Li, Z., Wang, Y., Shi, C., Wu, H., and Wang, S. (2017). Highway Elevation Data Smoothing Using Local Enhancement Mechanism and Butterworth Filter. Technical Report 6.

Davis, T. J. and Keller, C. P. (1997). Modelling uncertainty in natural resource analysis using fuzzy sets and monte carlo simulation: Slope stability prediction. *International Journal of Geographical Information Science*, 11(5):409–434.

Glennon, J. C. (1987). Effect of Alignment on Highway Safety. Technical report, Washington DC.

Hamdar, S. H., Qin, L., and Talebpour, A. (2016). Weather and road geometry impact on longitudinal driving behavior: Exploratory analysis using an empirically supported acceleration modeling framework q. *Transportation Research Part C*, 67:193–213.

Hawker, L., Bates, P., Neal, J., and Rougier, J. (2018). Perspectives on Digital Elevation Model (DEM) Simulation for Flood Modeling in the Absence of a High-Accuracy Open Access Global DEM. *Frontiers in Earth Science*, 6:233.

Horswill, M. and Coster, M. (2002). The effect of vehicle characteristics on drivers' risk-taking behavior. *Ergonomics*, 45:85–104.

Hu, X., Zhu, X., Ma, Y. L., Chiu, Y. C., and Tang, Q. (2018). Advancing usage-based insurance - A contextual driving risk modelling and analysis approach. *IET Intelligent Transport Systems*, 13(3):453–460.

Jo, K., Kim, J., and Sunwoo, M. (2013). Real-Time Road-Slope Estimation Based on Integration of Onboard Sensors With GPS Using an IMMPDA Filter. *IEEE Transactions on Intelligent Transportation Systems*, 14(4):1718–1732.

Joubert, J. W., De Beer, D., and De Koker, N. (2016). Combining accelerometer data and contextual variables to evaluate the risk of driver behaviour. *Transportation Research Part F*, 41:80–96.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering, Transactions of the ASME*, 82(1):35–45.

Litman, T. (2002). Evaluating Transportation Equity. *World Transport Policy & Practice*, 8(2):50–65.

Liu, H., Li, H., Rodgers, M. O., and Guensler, R. (2018). Development of road grade data using the United States geological survey digital elevation model. *Transportation Research Part C*, 92:243–257.

Magrath, J. and Brady, M. (2017). Evaluating different methods for determining road grade best suited to advanced bus transportation systems. Technical report.

Manson, N. J. (2006). Is operations research really research ? *ORiON*, 22(2):155–180.

Park, S., Gil, M.-S., Im, H., and Moon, Y.-S. (2019). Measurement Noise Recommendation for Efficient Kalman Filtering over a Large Amount of Sensor Data. *Sensors*, 19(5):1168.

Rogers, K. J. and Trayford, R. S. (1984). Grade Measurement With an Instrumented Car. Technical Report 3.

Rolison, J. J., Regev, S., Moutari, S., and Feeney, A. (2018). What are the factors that contribute to road accidents? an assessment of law enforcement views, ordinary drivers' opinions, and road accident records. *Accident Analysis Prevention*, 115:11–24.

Schafer, R. W. (2011). What Is a Savitzky-Golay Filter? *IEEE Signal Processing Magazine*, pages 111–117.

Sentoff, K. M., Aultman-Hall, L., and Holmén, B. A. (2015). Implications of driving style and road grade for accurate vehicle activity data and emissions estimates. *Transportation Research Part D*, 35:175–188.

Shinar, D. and Oppenheim, I. (2011). Review of Models of Driver Behaviour and Development of a Unified Driver Behaviour Model for Driving in Safety Critical Situations. *Human Modelling in Assisted Transportation*, pages 215–223.

TxDot (2018). Roadway Design Manual. Technical report, TxDot manuals.

Wechsler, S. P. (2003). Perceptions of digital elevation model uncertainty by DEM users. *URISA*, 15(2):57–64.

Winlaw, M., Steiner, S. H., MacKay, R. J., and Hilal, A. R. (2019). Using telematics data to find risky driver behaviour. *Accident Analysis & Prevention*, 131:131–136.

Wise, S. (2000). Assessing the quality for hydrological applications of digital elevation models derived from contours. *Hydrological Processes*, 14(11-12):1909–1929.

Wood, E., Burton, E., Duran, A., and Gonder, J. (2014a). Contribution of road grade to the energy use of modern automobiles across large datasets of real-world drive cycles. *SAE International*, 1.

Wood, E., Burton, E., Duran, A., and Gonder, J. (2014b). Contribution of Road Grade to the Energy Use of Modern Automobiles Across Large Datasets of Real-World Drive Cycles. volume 1.

Wood, E., Duran, A., Burton, E., Gonder, J., and Kelly, K. (2015). EPA GHG Certification of Medium-and Heavy-Duty Vehicles: Development of Road Grade Profiles Representative of US Controlled Access Highways Strategic Partnership Project Report. Technical report.

Xu, Q., Shao, F., Guo, T., and Gong, L. (2015). The effects of campus bump on drivers' fixation dispersion and speed reduction. *Mathematical Problems in Engineering*, 2015.

Yu, R. and Abdel-Aty, M. (2014). Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data. *Safety Science*, 63:50–56.

Zhang, K. and Frey, H. C. (2006). Road grade estimation for on-road vehicle emissions modeling using light detection and ranging data. *Journal of the Air and Waste Management Association*, 56(6):777–788.

Appendix A

Data as appendix

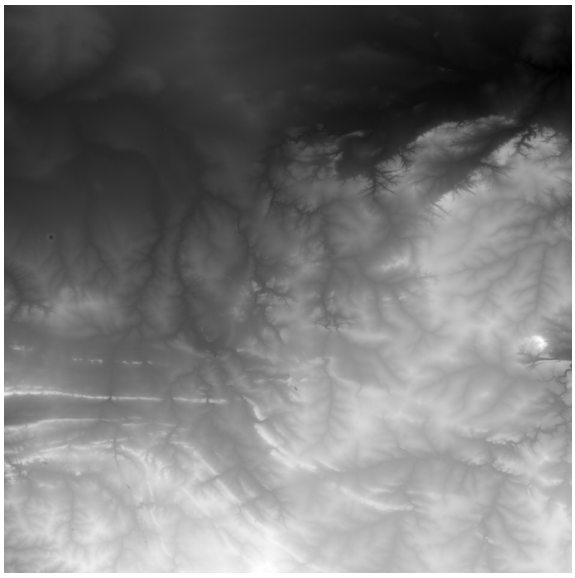


Figure A.1: S26E028 coordinate block with SRTM elevations extracted from United States Geological Survey (USGS)

Table A.1: Ranking of drivers in base model and in all road grade categories.

ID	no road grade	steep down	medium down	flat	medium up	steep up	st dev
1	9	27	13	7	24	29	9.54
10	30	25	27	22	33	20	5.03
11	11	17	18	13	8	14	3.94
12	14	28	33	11	18	16	9.04
13	6	9	5	6	7	7	1.48
14	38	35	42	38	31	38	4.09
15	23	2	35	34	32	35	14.36
16	22	10	34	20	13	10	10.14
17	27	6	7	33	23	28	12.3
18	24	31	31	21	29	19	5.76
19	2	36	3	2	1	-	-
2	13	16	24	14	12	17	4.56
21	15	7	4	17	14	37	12.95
22	26	19	15	28	9	13	7.22
23	12	14	29	9	27	5	10.73
24	29	32	37	27	28	26	4.53
25	34	37	40	32	42	39	3.81
26	7	11	6	5	6	25	8.38
27	41	38	44	40	41	32	4.47
28	10	5	30	12	34	8	13.27
29	31	34	38	25	37	11	11.29
3	20	23	23	29	19	27	3.9
30	35	21	20	37	38	34	8.8
31	8	18	9	8	3	6	5.63
32	40	29	43	39	36	30	5.94
33	4	4	36	10	10	2	13.67
34	18	12	21	24	26	22	5.39
35	21	3	39	31	35	33	14.39
36	17	30	12	26	25	24	6.77
37	3	22	16	4	5	3	8.51
38	42	-	2	43	-	-	-
39	36	39	41	35	39	4	15.58
4	25	15	28	19	17	12	6.06
40	37	40	8	42	11	21	15.92
41	28	33	25	23	30	31	4.22
42	16	13	19	16	16	15	2.17
43	19	26	26	18	22	18	4
44	32	20	22	36	21	36	8.25
45	44	-	32	44	20	-	-
5	33	24	14	30	40	-	-
6	1	1	1	1	2	-	-
7	39	-	11	41	4	-	-
8	5	-	10	3	15	-	-
9	43	-	17	15	-	-	-