

---

# Utilising Supervised Parametric Classification to Assess the Quality of the UK Rural Road Network using Aerial LiDAR Data

**201374125**

## Abstract

---

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nulla consectetur turpis ut libero efficitur, ullamcorper ultrices nulla elementum. Proin enim lorem, tempus sed tempus eget, fermentum vitae odio. Aenean congue urna id metus tincidunt efficitur. Donec finibus nec elit vitae tempor. Nullam eget vestibulum arcu. Nullam pulvinar sed quam nec ullamcorper. Curabitur at faucibus augue. Nulla scelerisque pretium sodales. Nulla facilisi. In id ipsum orci. Donec ex magna, sodales quis tempus a, sollicitudin sit amet odio. Proin nec molestie ante. Fusce ut hendrerit risus. Nunc eleifend magna dui, sit amet tempus velit tincidunt et. Duis iaculis convallis mi sed scelerisque. (Wang et al. 2015)

---

**Keywords:** LiDAR; Aerial Imagery; Linear Probability Classification; Road Quality

In Partial Fulfillment of the Requirements for the Degree of  
Geographic Data Science MSc



---

### Acknowledgements

---

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nulla consectetur turpis ut libero efficitur, ullamcorper ultrices nulla elementum. Proin enim lorem, tempus sed tempus eget, fermentum vitae odio. Aenean congue urna id metus tincidunt efficitur. Donec finibus nec elit vitae tempor. Nullam eget vestibulum arcu. Nullam pulvinar sed quam nec ullamcorper. Curabitur at faucibus augue. Nulla scelerisque pretium sodales. Nulla facilisi. In id ipsum orci. Donec ex magna, sodales quis tempus a, sollicitudin sit amet odio. Proin nec molestie ante. Fusce ut hendrerit risus. Nunc eleifend magna dui, sit amet tempus velit tincidunt et. Duis iaculis convallis mi sed scelerisque. (Wang et al. 2015)

---

## Contents

---

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Introduction to LiDAR . . . . .	2
1.1.1	Description of LiDAR . . . . .	2
1.1.2	Benefits of LiDAR . . . . .	3
1.1.3	Flaws with LiDAR . . . . .	3
1.2	Objective of this paper . . . . .	3
1.2.1	Site Selection . . . . .	3
1.3	sign posting . . . . .	3
<b>2</b>	<b>Literature</b>	<b>4</b>
2.1	Literature Review . . . . .	4
2.1.1	Road Extraction Techniques . . . . .	4
2.1.2	National Speed Limits . . . . .	4
2.1.3	Accessibility . . . . .	4
2.2	Feasibility for Road Extraction in the UK . . . . .	4
2.3	Rural Roads and Speed . . . . .	6
<b>3</b>	<b>Methods</b>	<b>8</b>
3.1	Data . . . . .	8
3.2	LiDAR Preprocessing . . . . .	8
3.2.1	Last Pulse . . . . .	9
3.2.2	Normalisation . . . . .	9
3.2.3	Points Extent . . . . .	9
3.3	Road Analysis . . . . .	10
3.3.1	Road Sampling . . . . .	10
3.4	Aerial Imagery . . . . .	10
3.4.1	Linear Probability Model . . . . .	11
3.5	Road Angles . . . . .	11
3.5.1	Road Node Elevation Change . . . . .	12
3.5.2	Surface Quality . . . . .	12
<b>4</b>	<b>Results</b>	<b>13</b>
4.1	Overview of LiDAR Data . . . . .	13
4.2	Overview of Roads . . . . .	13
4.3	Data Preparation . . . . .	13
4.4	Perpendicular Sampling . . . . .	14
4.5	Linear Probability Model . . . . .	14
4.6	Widths and Road Quality . . . . .	14
4.7	Road Assessment . . . . .	14

---

**CONTENTS**

---

4.8 ( . . . . .	14
4.9 Noise Filtering . . . . .	15
4.10 Further Analysis . . . . .	15
<b>5 Discussion</b>	<b>16</b>
<b>A Code Appendix</b>	<b>19</b>
A.1 Session Information . . . . .	19
A.2 Functions . . . . .	21
A.2.1 Formatting . . . . .	21
A.2.2 Catalog to Dataframe . . . . .	21
A.2.3 LiDAR Clean . . . . .	21
A.2.4 Filter Noise . . . . .	21
A.2.5 Extract Buffer . . . . .	21
A.2.6 Euclidean Distance . . . . .	22
A.2.7 Perpendicular Sampling . . . . .	22
A.2.8 Combine Catalog . . . . .	23
A.2.9 Compute Samples . . . . .	23
A.2.10 Greyscale . . . . .	23
A.2.11 Compute Linear Model . . . . .	23

---

## List of Figures

3.1	Road LiDAR points at maximum distance apart for each sample location. Showing two example sample locations ( <i>A</i> and <i>B</i> ), road centreline represented by the thick grey line. True road width is indicated by the dashed lines $O_A$ and $O_B$ .	11
3.2	Bearing Angle Between Road Segments	12
3.3	Bearing Angle Between two Road Segments	12
4.1	LiDAR point clouds for one selected road section aggregated into $1m^2$ grids	13

---

## List of Tables

3.1 LiDAR Point Cloud Summary Data . . . . .	9
3.2 OS Roads Data Summary . . . . .	10
4.1 Correlation Matrix . . . . .	14

## 1. Introduction

---

ROAD usage within the United Kingdom has been steadily increasing by year with the highest ever billion vehicle miles travelled in 2018 (255 million) (Department for Transport 2019). Characterised by tall hedgerows and winding turns, rural roads in the UK are often unsuited higher traffic flow, due to the obstruction of view due to the protected hedgerows, narrow lanes and often poor condition. The abundance of these roads, with "Unclassified" local network roads making up 60% of all roads in the UK (Department for Transport 2012), and their varying nature, the national assessment of these roads into appropriate speed limits on an individual basis has been considered impractical. Due to this, there have been no individual assessments for the majority of rural roads, and given their nature are classified as unlit, single carriageway roads and thus assigned a default speed of 60mph (?). Highways England manages the motorways and trunk roads within the UK, but local road networks are maintained by Local Authorities, and as such have no comprehensive information regarding these smaller road networks (England 2019). Rural roads in the UK are often cited as by far the most dangerous road type with studies suggesting that up to two thirds of vehicle accidents occur on rural roads (Corben et al. 2005).

The Rural Urban Classification defines a rural area as one outside a settlements with more than 10,000 resident population (Government 2011), therefore a road could be considered rural, if either connecting or present within small settlements in the UK. this study will focus particualrly on rural connecting roads.

A Governmental review of speed policy considered the need for the role of speed and accidents on rural roads to be further considered (Department of the Environment, Transport and the 2000), suggesting a framework for individua classification of roads,

taking into account local considerations of the road to implement more suitable speed limits. In 2012, draft guidance for rural roads was presented by the Department for Transport suggesting a blanket reduction in rural single carriageway road speed limits from 60mph to 40mph including a reduction to 50mph for lower quality A and B roads (BBC 2012). However, this draft guidance has yet to be implemented, and the report archived.

// likely due to costs involved, blanket implementation likely doesn't justify the cost. Need to rely on more quantiative analysis of individua roads, cost estimated as \$... or cost of changing all speed cameras to metric estimated as \$80/400m. Resulting in the current state of a zombified combination of metric and british imperial measruements.

National speed limits have seen little variation for a number of years, with the majority of roads following the broad criteria for the three main roads types. The three national speed limits are:

- the 30 mph speed limit on roads with street lighting (sometimes referred to as Restricted Roads)
- the national speed limit of 60 mph on single carriageway roads
- the national speed limit of 70 mph on dual carriageways and motorways.

(?)

These limits are set by the UK government, however local speed limits are determined by local councils. Department for Transport (2013a) outline in Setting Local Speed Limits, that national speed limits are not appropriate for all roads, where local road conditions present the requirement for alternative speed limits. The majority of the rural road network in the

UK follows the national speed limit of 60mph for single carriageway roads, and 70mph for dual carriage-way roads, despite driver speed often being far below the speed limit. Noted as especially common on C and Unclassified roads due to the narrow roads, bends, junctions and access. In 2011, an estimated 66% of total road deaths in Britain occurred on rural roads, with 51% on single carriageway rural roads with the national speed limit of 60mph.

Department for Transport (2013a) suggest that selecting alternative speed limits for single carriageway rural roads should consider:

- History of collisions
- The road's function
- Existing mean traffic speed
- Use by vulnerable road users
- The road's geometry and function
- The road environment, including road-side development

This paper will particularly focus on the requirement of road geometry //

The Road Safety Management Capacity Review (Department for Transport 2018) outlines the current limitations with road safety management, with the lack of defined and measurable safety performance framework, noting that such a framework should set out the long term goal of total prevention of road deaths and injuries, achieve this through a reducing in average speeds on different road types, and an improvement in emergency response times. This review states that at present there is a distinct lack of both urban and rural road hierarchies, which could be used to better match appropriate speed limits, with function, layout and design. Again, this review notes that posted speed limits often allow for speed far in excess of the design limits of single carriageway rural roads, with inappropriate but allowable speed often a contributing factor in rural accidents. Finally the report calls for a review on national speed limits as soon as possible.

// limitations of the newly introduced speed limit appraisal tool?

This paper will present a method for rural road extraction for a region in the North West of England. The methodology is produced in order to ensure scalability and automation, allowing for replication for any area where data is available. Data used will include road centreline geometries, LiDAR point cloud, and aerial imagery to extract road widths through linear likelihood models. Additionally, this paper aims to extract other features of roads such as elevation changes, surface quality, and the sharpness of bends. These key features aim to provide information that may inform the selection of appropriate speed limits on rural roads, where the 60mph national speed limit is inappropriate.

## 1.1. Introduction to LiDAR

### 1.1.1. Description of LiDAR

LiDAR data is collected by emitting rapid laser pulses from an aircraft towards the ground which are reflected back, measuring the distance between the aircraft and surface objects at up to 500,000 measurements per second (Agency 2019). This method produces a set of highly accurate three dimensional points which collectively are known as a LiDAR point 'cloud'. As LiDAR data detects all surface objects, the resultant point cloud produced will include all natural and man made structures, including buildings, roads, trees in addition to the natural variation in the terrain height, known as a digital surface model (?).

The main features unique to LiDAR, as opposed to similar aerial data collection techniques such as true colour imagery are outlined below:

- *Pulses:* LiDAR systems record the data by recording emitting a laser pulse which is reflected back at the aircraft by ground objects. If the laser hits a solid object such as ground or a building roof, this laser pulse is entirely reflected back towards the aircraft, giving a single point. However, if the laser pulse hits a soft object such as a tree canopy, the pulse may be partially returned, giving multiple return pulses (Rottensteiner et al. 2003). Therefore, these multiple pulse returns give information regarding objects at an exact *x y* location, but with varying *z* locations.

- *Intensity:* LiDAR systems also give intensity values for return pulses, which gives information regarding the reflectance of the surface of objects that are hit by the laser pulses. If intensity is given  $I$  then reflectance  $R$  may be represented as  $\frac{I}{E_T}$  where  $E_T$  refers to the first pulse signal intensity (Charaniya et al. 2004).
- *Elevation:* In addition to  $x$  and  $y$  coordinates, the distance between the plane and the reflected ground or object is recorded and assigned a  $z$  value.

### 1.1.2. Benefits of LiDAR

Rural roads in the UK are often characterised by dense hedgerows either side, often containing large oak trees which extend over the road surface. In addition to the reduction in corner visibility on these roads, standard aerial imagery suffers from the road surface being obscured by shadows from these trees, and the tree canopy itself. Aerial imagery also often suffers from obstruction due to clouds (Li et al. 2016). Due to the inclusion of pulses with modern LiDAR data, the road surface can often be detected through the canopy by selecting the final pulse values, additionally the infrared laser pulses have smaller shadows, due to the narrow scanning angle of LiDAR (?). Non LiDAR imagery often suffers from scene complexity, where road patterns, vehicles and lane markings reduce road heterogeneity (Li et al. 2016).

The 3D  $z$  value information provided by LiDAR data allows for the separation of ground and objects on the surface, meaning roads and buildings are often easily separated, despite having similar reflectance (or intensity) (?). Additionally, the reflectance of roads is often homogeneous, and distinctly separate from vegetation (Clode et al. 2004).

### 1.1.3. Flaws with LiDAR

LiDAR lacks any texture or spectral information, and often studies in road detection have combined LiDAR with imagery to alleviate this issue (?), with the inclusion of luminescence information to aid with road classification (e.g. Charaniya et al. 2004). LiDAR points are distributed irregularly and with varying density, with point density often higher where flight

strips overlap (Li et al. 2016), and tall objects can occlude points, leaving limited data surrounding trees or buildings.

Often studies use LiDAR height data to identify kerbs to separate streets from pavement (?), however rural roads often have no kerb, and are at the same level as the surrounding vegetation if they are managed grass verges (Yadav et al. 2018).

LiDAR data often requires a large amount of processing due to the irregular distribution of points, presence of noise and the number of variables that have to be considered, Yadav et al. (2018) note that often papers do not include the computational time for processing this data, which can be time consuming.

```
Error in .local(x, ...): invalid layer names
Error in fortify(data): object 'bg' not found
```

## 1.2. Objective of this paper

- Produce an automated method using LiDAR and aerial imagery to determine the true width of roads within the chosen study area.
- Using OS Road and LiDAR Data produce an automated method for determining the characteristic of roads that relate to overall road quality outline above.
- Assess the potential for speed reduction on rural roads, with the assumption that at present the roads chosen are 60mph.

### 1.2.1. Site Selection

The site selected in this study was chosen to include a selection of A, B, and Unclassified local roads within a rural setting. Particularly the roads chosen are often partially obscured by tree canopies and do not have visible kerbs, both key features in rural British roads.

**figure 1a.** Additionally the area selection was limited by the availability of LiDAR point cloud data, which is shown on **figure1b**.

## 1.3. sign posting

## 2. Literature

---

### 2.1. Literature Review

#### 2.1.1. Road Extraction Techniques

TYPICAL road extraction techniques have focused purely on rural road networks, where techniques involving LiDAR often rely on the kerb and change in height to determine road edges.

#### 2.1.2. National Speed Limits

Accidents on rural roads often occur within the 60 mph speed limit meaning a distinction between what is an appropriate speed should be made that does not relate to a given speed limit. Baruya (1998) suggest a distinction between both *excess* and *inappropriate* speed. *Excess* speed being driving above the speed limit, and therefore directly breaking the law, *inappropriate* speed; driving too fast for the conditions of the road, not necessarily above the speed limit.

Lowering the speed limit on roads has been shown to result in an overall reduction in the average speed of vehicles (Finch et al. 1994), and subsequently the number of road traffic accidents (Taylor et al. 2002). Additionally, ? found that accident frequency is linked with traffic and pedestrian flow, vehicle speed, and particular features of the road geometry.

As noted with the MASTER European speed study, access to quality sources of data which includes the road geometry for use in individual road assessments has been limited. ? note that the periodic assessment of roads is greatly important due to the increasing traffic load, and new automated techniques will enable this in areas where in the past it had not been feasible.

#### 2.1.3. Accessibility

Transport accessibility for rural communities is far more limited than for urban communities, where often rural areas have limited or no public transport, meaning there is a heavy reliance on personal transport (Gray et al. 2001). Accessibility in this context can be defined as the transport facility or opportunities with which basic services can be reached from a given location by using a certain transport system (Gutiérrez 2009).

There has been little focus on the improvement of transport technologies in rural areas, with the potential for new technologies implemented into urban areas improving rural accessibility (Velaga et al. 2012). A key area to address is the level of accessibility to hospital services for rural communities, where recent centralisation of these services has negatively impacted the level of access for rural communities (?). This also impacts the level of access for hospital services to reach rural areas, where distance and time taken to a hospital directly correlates with a patients mortality (?). Emergency vehicles are often larger than personal vehicles and as such it can be assumed that accessibility for these types of services is often more limited depending on the quality of rural roads.

### 2.2. Feasibility for Road Extraction in the UK

Traditional road extraction techniques from aerial imagery relied on human extraction that is both time consuming and prone to inaccuracy (Wang et al. 2015). However, recent improvements in remote sensing technology now mean that automated extraction techniques have been explored which rely on the much higher resolution imagery available (Xu et al. 2018). This paper will use Light Detection

and Ranging (LiDAR) data, made available by the UK government for much of England. LiDAR is collected by measuring the distance from an aircraft to the ground, emitting laser pulses at regular intervals (Agency 2019), allowing for very high resolution terrain models to be produced, with a resolution up to 25cm. Additionally the 'point clouds' produced by the laser pulses contain information that may be utilised to determine the qualities that are indicative of road locations (e.g. Clode et al. 2004).

Additionally, luminescence information extracted from 25cm resolution aerial true colour imagery is available for research purposes for the entirety of the UK.

// timeline of aerial image quality /lidar

See (Kumar et al. 2013):

vosselman2009: segmenting lidar data identification of planar or smooth surfaces and classification of point cloud data based on attributes. clode2004: segmentation of airborne lidar to road and non-road objects using hierarchical classification techniques based on elevation and intensity information, accuracy impaired due to car parks and private roads. goulette2006: methodd for segmenting road, trees, and facades from terrestrial mobile lidar data. road sgemented as horizontal plane high density of points in the histogram + usedt o compute road width and curvature. Trees and facades identified as vertical planes and disconnected elements in the histogram. yuan2008: algoritm using fuzzy clusting method to cluster LiDAR points, straight lines then fitted to the linearly clustered data using slope information for extracting the road. elberink2009: automated method for 3d modelling highway infrastructure using airborne LiDAR + 2d topographic. Road polygons were extracted from the topographic map data using a map based seed-growing algorithm combined with a Hough transformation. points added to the corre- spondng road polygons using a LiDAR based seed-growing algorithm then mapped polygons

Most of the proposed methods (Ferchichi and Wang, 2005; Wan et al., 2007; Mokhtarzade and Valadan Zoj, 2007; Mena and Malpica, 2005; Mohammazadeh et al., 2006; Wang et al., 2005) using satellite images and aerial photographs only provide road pixels and its two- dimensional (2D) location information.

- Extraction includes detection of planar or smooth surfaces, and the classification of points or point clusters based on local point patterns, echo intensity and echo count information (?)

2. Lidar points distibute irregularly with non uniformity density. the point density in the overlapping area between fight strips is greater than non-overlapping regions. There is no point in areas occluded by tall objects. can be soled interpolating into intensity image (e.g. Zhao,2012, zhu 2009) binary image, Clode 2007. others see Samadzadegan 2009, Jiangui 2011.

(Yadav et al. 2018): Major limitation of existing methods to extract rural road surface as mostly without curb. Methodologies such as Yang et al., 2013 which deals with curbs and boundaries characterized as ash-palt/soil and ashpalt vegetation and ashpalt/grassy bank. Exact edge location difficult if point density less and sufficient elevatio jump not present at boundary of carriageway (Mc Elhinney et al., 2010). Wu et al., 2013 dependent on road range which cannot be calculated correctly if height diffrrnce between road surface and shoulder not obvious.

- from (Vosselman & Liang 2009): clode 2004a, pulse reflectance strength + dtm to detect roads . Later improved including buildings and vegetation.

- Akel et al. 2005, road areas classifying smooth segments based on size and shape characteristics - Rieger 1999 used surface slope in a 20cm elevation grid to detect roads in mountainous areas - Hatger and Brenner 2003,2005, laser scanning data at 0.5m to extract properties of roads for which centrelines arer known from a road database. Including height longitudinal and transversal slope, curvature, and width. Height profiles across road split into straight sgements. A RANSAC procedure on the end points of the sements with low slope values is used to fit straight lines to the road sides. (?)

Look at (Taylor et al. 2002) for a full assessment of rural roads. Includes bends, hills etc etc.

Clode 2004:

Note that as they use point density of 1/2m, and intensity has a footprint of 20 - 30cm the intensity if not typically used. (Rottensteiner, 2003). \*\*with my method we use a resolution of 25cm, much higher point density\*\*

// list varying resolution increases per papers?

From: (?) - Segmentation through discontinuity, prominent points or edges, or areas fulfilling certain criteria in continuity. - Point operators which try to find isolated points, corners, or points being part of a one dimensional curve (Haralick and Shapiro, 1992, Canny, 1986 (old)) - Second step building contour chains from individual points. - alternatively, discontinuity perpendicular to a linear structure can be defined in terms of the continuous areas to the left and right of the structure (Bruegelmann, 2000, Wild and Krystek 1996. + region growing)

- inclusion of as much information e.g. minimum width requirements etc (actually not defined in the UK however) - Continuous surfaces approximated by planes, usually horizontal

Techniques from Hatger 2005:o General Planar Region Growing Segmentation Iterating following: 1. Find the best seed region which fulfills the desired predicate, 2. Add elements to the seed region (i.e. grow it) as long as they are connected to it, and they too fulfill the predicate 3. If the region cannot be grown anymore, accept it and go to 1. using the remaining elements. They use estimation of local planes, and look for the smallest residuals. The predicate is a certain maximum distance  $e$  of the points to the plane given by its Hesse normal form

The scan line grouping approach Less computationally expensive than region growing. - Presented by (?). - Raster data: plane  $z = ax + by + d$  all points along the line  $y = y_0$  fulfill the equation  $z = as + by_0 + d$  the line equation. add entire scan lines Steps: 1. Partitioning of each scan line  $y = y_0$  into linear segments fulfilling corresponding line equations 2. Search for a seed region by investigating overlapping linear segments of three successive lines 3. Growing the best seed region by adding neighbouring line segments, as long as they still are part of the same plane 4. Postprocessing:

(RANSAC), introduced by (Fischler and Bolles, 1981). fixes errors

from Clode 2004:

Road pixel must lie on or near DTM, and have a certain intensity and normalized local point density.

\*\*My paper will extend this, inclusion of road shapefiles to remove known non roads, noted as a limitation in Clode therefore, less focus on automated extraction, (as it isn't needed generally), more focus on

automated identification of road features, i.e. width etc.\*\*

Intensity identifies the type of road material (bitumen in their paper)

Create DTM from last pulse DSM \*\*may not be need for my work if I am already filtering onto known approximate locations of roads\*\*

Intensity identifies the type of road material (bitumen in their paper)

Charanya gives information which is utilised in this paper for the modelling. i.e. Lum and Intensity are most useful for road detection

## 2.3. Rural Roads and Speed

112. Rural roads account for 66% of all road deaths, and 82% of car occupant deaths in particular, but only around 42% of the distance travelled. Of all road deaths in Britain in 2011, 51% occurred on National Speed Limit rural single carriageway roads (DfT, 2011). The reduction in road casualties and especially deaths on rural roads is one of the key road safety challenges. Research has assessed the risk of death in collisions at various impact speeds for typical collision types on rural roads. This research suggests that the risk of a driver dying in a head on collision involving two cars travelling at 60 mph is around 90%, but that this drops rapidly with speed, so that it is around 50% at 48 mph (Richards and Cuerden, 2009).

113. Inappropriate speed, at levels below the legal limit but above those appropriate for the road at the time (for example, because of the weather conditions or because vulnerable road users are present), is a particular problem for rural roads. Exceeding the speed limit or travelling too fast for the conditions are reported as contributory factors in 16% of collisions on rural roads. Specifically, inappropriate speed is recorded as a contributory factor in 20% of crashes on minor rural roads with a 60 mph limit. 114. Speed limit changes are therefore unlikely to fully address this problem

34. On rural roads there is often a difference of opinion as to what constitutes a reasonable balance between the risk of a collision, journey efficiency and environmental impact. Higher speed is often perceived to bring benefits in terms of shorter travel

times for people and goods. However, evidence suggests that when traffic is travelling at constant speeds, even at a lower level, it may result in shorter and more reliable overall journey times, and that journey time savings from higher speed are often overestimated (Stradling et al., 2008). The objective should be to seek an acceptable balance between costs and benefits, so that speed- management policies take account of environmental, economic and social effects as well as the reduction in casualties they are aiming to achieve.

(Department for Transport 2013b) Straightforward method for determining appropriate speed limits. Forecasting speed reduction and therefore accident reduction with revised speed limits. Does not take into account road geometry etc.

Traffic Advisory Leaflet

VERALL PATTERNS OF SPEED ON DIFFERENT ROADSSpeeds on single carriageway rural roads are generally wellwithin the national 60mph speed limit. Observations from270 sites around England show a wide distribution of meanspeeds on roads with speed limits of 60mph. They also show that, for the majority of roads, mean speeds arealready below the posted limit or would typically only needto be reduced by a few miles per hour, where lower limitsmight be re-

quired.

The Transport Research Laboratory (TRL) have produced extensive evidence that confirms the correlation between speed and the likelihood of road accidents, demonstrating that faster speeds do directly increase the chance that a driver is involved in an accident (Taylor et al. 2002). Additional assessments roads within the EU have confirmed this relationship particularly focusing on rural roads with the production of the EURO model (Baruya 1998), however Taylor et al. (2002) note that the data used in this study was very limited for the United Kingdom due to the time consuming extraction of road features, overall the study found that 20% of fatal accidents occurred on rural roads.

Key factors in determining poor road quality:

- Width (Taylor et al. 2002) (Aarts & van Schagen 2006)
- Surface quality
- Blind corners/winding roads (Aarts & van Schagen 2006)
- Junction Sharpness (Aarts & van Schagen 2006)

// read these papers and identify why these should be considered when assessing road quality

### 3. Methods

---

This paper primarily makes use of the free open source statistical language **R** (R Core Team 2019). Managing the large LiDAR datasets from my personal computer was made possible through the **lidR R** package (Roussel & Auty 2019). Further details regarding the **R** environment and computer setup used for this paper is given in **Appendix A**. All content was written using **L<sup>A</sup>T<sub>E</sub>X** combined with the **.rnroweb** file type (Ihaka 2011), for *Literate Programming*<sup>1</sup>. The template is built from scratch but takes much inspiration (and code) from the **R-LaTeX-Template**.

All code is hosted on my personal GitHub account. Also hosted are my complete dotfiles, used in conjunction with the Linux distribution Manjaro, with the i3 window manager. All writing and code was produced using **neovim** with my personal configuration to implement integrated development environment (IDE) style features for writing R code, while also providing essential features for writing in **L<sup>A</sup>T<sub>E</sub>X**. **neovim** has the benefit of being both highly customisable, and lightweight, which allows for much lower system utilisation compared with R Studio when working with large datasets. One essential **vim** package to mention is the **Nvim-R vim** package.

Given in **Appendix A** are the code snippets utilised in this methodology, for many equations, the relevant code is given as a reference to the appendix location.

#### 3.1. Data

LiDAR point cloud data was downloaded through the UK Government's open data repository which uses the Open Government Licence, allowing for:

- Copying, publishing, distributing and transmission of the data
- Adaptation of the data

<sup>1</sup>See Knuth (1984)

- Commercial and Non-commercial use of the information

LiDAR data used in this paper is available [HERE](#) under this licence (UK Government 2019). This data was given as a compressed LAS file format (.laz), the **R** package **lidR** provided the function **catalog()** which enabled each separate .laz to be combined into one object of class **LAScatalog**. Analysis on this object could then be split into chunks (selected as 500m<sup>2</sup>), allowing for multi-core threading to speed up analysis, and a reduction in the memory overhead when reading in data, often a limitation of the **R** language as objects are stored entirely into memory when read (Wickham 2014). The **LAScatalog** object did not require the compressed .laz files to be read into memory as .las files, meaning memory limitations were far less of a problem.

Aerial imagery was downloaded through **Digimap®** which uses the *Aerial Digimap Educational User Licence*, allowing for free use of the data for educational purposes (The University of Edinborough 2019).

Road centreline geometries were accessed through the **Ordnance Survey Open Data** repository which shares the Open Government licence. These were downloaded in the GeoPackage format (.gpkg) nationally and cropped to the extent of the LiDAR point cloud data.

#### 3.2. LiDAR Preprocessing

The total number of LiDAR points used in this study is 9,419,272. All LIDAR data has a vertical accuracy of +/-15cm Root mean square error (RMSE). An overview of the LiDAR data selected for this study is given on Table ???. The variables of primary interest

are:

- **Z:** The distance a laser pulse is reflected back to scanner, calculated by the time taken for a return pulse to be detected.
- **Intensity:** The amplitude of the return pulse, reflected back by the surface terrain or objects.
- **ReturnNumber:** A number of range 1-5, indicating for a point, the corresponding order of a reflected laser pulse. A ReturnNumber of 1 indicates the first return for a pulse (and highest  $z$  value), a return number of 5 indicates the last return (and lowest  $z$  value).
- **NumberOfReturns:** The number of return pulses for a single laser pulse (maximum of 5).
- **Classification:** A number given to a point indicating a specific numeric classification. Of interest in this study is a classification of 2, indicating a ground point. More information can be found [here](#), which outlines numerical classifications for various vegetation types and man made structures.

**Table 3.1: LiDAR Point Cloud Summary Data**

	Mean	SD	Min	Max
Z	80.55	5.95	64.85	115.79
Intensity	176.57	125.08	1.00	4064.00
ReturnNumber	1.47	0.95	1.00	5.00
NumberOfReturns	1.95	1.42	1.00	5.00
ScanDirectionFlag	0.50	0.50	0.00	1.00
EdgeOffFlightline	0.00	0.03	0.00	1.00
Classification	3.04	1.70	1.00	8.00
ScanAngleRank	-1.97	13.18	-22.00	22.00

### 3.2.1. Last Pulse

The LiDAR point cloud data used in this paper gives the values for 5 pulse returns. The canopy above roads may be excluded through ignoring early pulses (higher Z values), therefore only the last pulse values for any point are selected, this can be expressed as;

$$\mathbf{p}_i = (lpx, lpy, lpz, lpi),$$

A.2.3

where  $\mathbf{p}_i$  is a single instance of a LiDAR point within the chosen point cloud,  $lpx$  is the last pulse

$x$  coordinate,  $lpy$  the last pulse  $y$  coordinate,  $lpz$  the last pulse  $z$  coordinate, and  $lpi$  the last pulse intensity value.

### 3.2.2. Normalisation

Ground points were classified using the Cloth Simulation Filtering (CSF) algorithm, as described in [Zhang et al. \(2016\)](#). Points were already classified in the data provided, however, as the classification technique was unknown, reclassification was considered necessary. The general implementation simulates the movements of a piece of cloth lying over the inverse of a point cloud, as the point cloud is flipped, the cloth settles beneath ground points, while covering points that lie separate to the ground, essentially forming a digital terrain model (DTM), cloth simulations are described in more detail in [Bridson et al. \(2005\)](#). The CSF algorithm is given;

$$X(t + \Delta t) = 2X(t) - X(t - \Delta t) + \frac{G}{m} \Delta t^2,$$

A.2.3

where  $m$  is the mass of a single LiDAR point (set to 1),  $\Delta t$  is the time step between points and  $G$  represents the gravity constant.

With the classification of ground points, (given  $Classification = 2$ ), a full DTM may be produced through spatial interpolation of the classified points. This process is called normalisation, and ensures that when extracting height information, any observed values are due to objects on the surface of the terrain, and not due to the lie of the terrain itself. Interpolation uses the inverse distance weighting and  $k$  nearest neighbours algorithms to produce the DTM. Nearest neighbours were selected as  $k = 10$ , with  $p = 2$  for the inverse weighting, and used to produce a DTM with a resolution of 1m. This particular technique was selected over more comprehensive methods such as kriging as the number of points is very high, and the small benefit was considered minimal compared with the increase in computational load. The  $z$  values from the DTM were then subtracted from the LiDAR point cloud, leaving a normalised point cloud.

### 3.2.3. Points Extent

With the normalised last pulse point cloud, the point cloud was clipped to within a 30m extent of each known road location, using the OS road shapefiles. Selecting a 30m extent ensured that even with slight inaccuracy in road location, the road LiDAR points would likely not be excluded. A large number of unimportant points were therefore removed, saving on computational resources. Additionally this extent ensured that both road and non road points were included, but reduced the chance of false positives from occurring as fewer non road points were now included in the analysis.

## 3.3. Road Analysis

This section combines data extracted through the OS road shapefiles preprocessed LiDAR data, and aerial imagery to obtain a set of criteria to assess the chosen road network. These criteria are;

- Road Width
- Bend Sharpness
- Road Steepness
- Surface Quality

**Table 3.2: OS Roads Data Summary**

id	idE381337E-E88D-4232-8CAD-F543F178EBE4
endNode	id42B6F387-D838-445C-AA7A-6558362B7B9F
startNode	idC8EE8B4C-D965-436A-BA02-A0925A6EA1B8
roadNumberTOID	osgb4000000013398492
roadNameTOID	
fictitious	FALSE
roadClassification	B Road
roadFunction	B Road
formOfWay	Single Carriageway
length	241
loop	FALSE
primaryRoute	FALSE
trunkRoad	FALSE
roadClassificationNumber	B5392
name1	
name2	
roadStructure	

### 3.3.1. Road Sampling

The LiDAR point cloud data was sampled at regular 10 meter intervals for each road, perpendicular to

the road direction, ensuring that when road direction changed, the sampling locations remained perpendicular. Each road was first split into nodes at which road direction changes, and from this, points at 10 meter intervals between two connected nodes were calculated;

$$p_k = p_1 + |p_2 - p_1| \times I_k$$

where  $I_k$  is the interval value which increases by 10 meters until the length of the total distance of the node, given  $I_1, I_2, \dots, I_k > 0$  and  $k \geq 2$ .

$$I_k = (L_k + 20) - T_k$$

$$L_1 = 20$$

$$T_k = T_{k-1} + E$$

where  $L_k$  is the distance from the start of the node to the current sample line position,  $T_k$  is total length of all nodes within a road from the start of the road and  $E$  is the euclidean distance between two nodes.

To calculate the sample locations, perpendicular to the roads, first the  $x$  and  $y$  euclidean distances from a reference point  $p_k$  and the end of the current node  $n_2$  were calculated;

$$\text{len} = \sqrt{(n_{k+1} - p_k)}$$

$$\text{len}_n = (n/\text{len}) \times (n_{k+1} - p_k)$$

A.2.7

from this, the lengths were inverted, giving two points of distance  $n$  from  $p_k$ . Inversion was achieved by subtracting the given  $\text{len}_n$   $x$  value from  $p_k$ , while adding the  $\text{len}_n$   $y$  value, and vice versa. These perpendicular sample lines were buffered to a width of 1m, giving a total area of 20m<sup>2</sup> per sample and spatially joined to the existing LiDAR point cloud data, removing any points that fell outside the sample lines.

A.2.9

## 3.4. Aerial Imagery

With the perpendicular sample lines extracted for the length of every road, to assist with the prediction of correct road locations, true colour aerial imagery was included. This imagery was first converted from three band RGB raster images, to a single-band grey-scale raster brick with values ranging 0 to 255, considered to be luminescence information.

$$\text{luminescence} = (\text{Band}_1 + \text{Band}_2 + \text{Band}_3) \div 3$$

A.2.11

To exclude as much noise as possible from both the LiDAR point cloud intensity information, and aerial luminescence, only points with the ground classification (*Classification* = 2), and points with only a single laser return pulse were included in the models (*NumberOf Returns* = 1).

### 3.4.1. Linear Probability Model

Individual linear probability models were constructed for each sample location for each road, giving results for a cluster of points across a perpendicular segment of each road every 20m. Road centrelines were selected as the outcome variable, provided by the OS road network line shapefiles, buffered to 1m to ensure a three dimensional cross section of points. Additionally, global linear probability models were constructed, and compared against the individual linear models.

Models were constructed using a maximal approach, selecting all appropriate predictor variables, iterating through models by removing variables that did not significantly influence the model outcome, or created noise.

An additional variable *Dist* was included, representing the shortest distance from a point to the centreline of the road it is associated with, considering that road points should be weighted more towards points that are closer to the centre-point of the road.

Predictions for each point were then ran using each model, which gave the likelihood a point was either road or non road. As the likelihood values from the predictions gave a range of numerical values, points that fell below a certain threshold were removed, leaving only points that were most likely correctly identified as road points. This threshold was selected as the 95% quantile for each sample point;

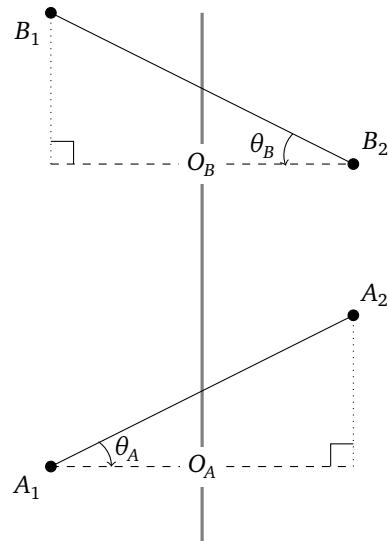
$$\mathbf{S} = \left( \mathbf{p}_i \in \left[ \frac{95}{100} \times \mathbf{p}_{l_{pi}} \right] \right)$$

Where  $\mathbf{S}$  is the total point cloud.

Some points considered to be noise were still present, but often isolated. To ensure no isolated

points were present, the time interval between two neighbouring points was checked, if a large time interval was found, the isolated point was removed.

$\mathbf{S}_i$  now consisted of a collection of predicted road points for each sample line along a road segment, excluding sample lines partially or fully obscured by tree canopy. To obtain road widths from these points, the maximum distance between two points in a particular sample was determined, these points were kept and all others removed. A linear section of road with two samples resembles ??;



**Fig. 3.1:** Road LiDAR points at maximum distance apart for each sample location. Showing two example sample locations (A and B), road centreline represented by the thick grey line. True road width is indicated by the dashed lines  $O_A$  and  $O_B$ .

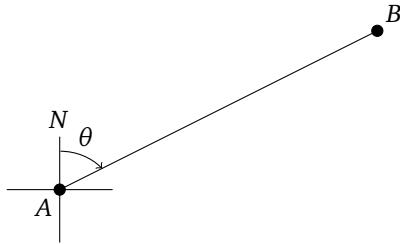
To determine the width of the road section, according to the final two selected points at every node, pythagoras could be used to find the opposite line length, perpendicular to the road segment, considering the distance between the two points to be the hypotenuse of a right angled triangle (??). The average width for each road identifier was then found.

$|K_1 K_2| = \text{Hypotenuse}$ ,  $O_K = \text{Opposite}$ , therefore  $O_K = |K_1 K_2| \times \cos(\theta_K)$

### 3.5. Road Angles

The angle of bends in roads were identified through the nodes produced in the creation of the road shapefiles. First the road linestrings were split into points, with coordinates representing each node within a road, a point at which the orientation of the linestring is altered.

The bend in a road was considered to be the *bearing angle*  $\theta$ , from a point  $A(a_x, a_y)$  to a point  $B(b_x, b_y)$ , with the angle measured in a clockwise direction from north. This can be represented as a figure;



**Fig. 3.2: Bearing Angle Between Road Segments**

To find the angle  $\theta$ , the point  $B$  can be represented into relation to point  $A$  as;

$$(b_1, b_2) = (a_1 + r \sin \theta, a_2 + r \cos \theta)$$

Where  $r$  is the length of the line segment  $AB$ . Rearranging the equation for  $\theta$  gives;

$$\tan \theta = \frac{b_1 - a_1}{b_2 - a_2}$$

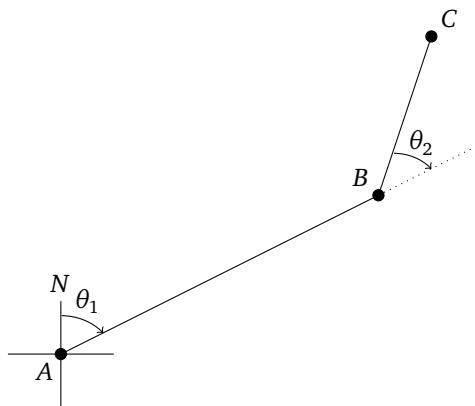
This equation can be rewritten to calculate the value of  $\theta$  using the *atan2* function;

$$\hat{\theta} = \text{atan2}(b_1 - a_1, b_2 - a_2) \in (-\pi, \pi]$$

Finally the bearing angle  $\theta \in [0, 2\pi)$  is given as;

$$\theta = \begin{cases} \hat{\theta}, & \hat{\theta} \geq 0 \\ 2\pi + \hat{\theta}, & \hat{\theta} < 0 \end{cases} \quad ??$$

With the bearing angle of the first line segment  $AB$  for a particular road, the change in orientation of the second road segment  $BC$  may be given as  $\theta_2 = \theta_{BC} - \theta_{AB}$ , with additional nodes following the pattern  $\theta_k = \theta_{N(N+1)} - \theta_{(N-1)N}$ .



**Fig. 3.3: Bearing Angle Between two Road Segments**

For each road the maximum bearing angle between two nodes was selected, as well as the average bearing angle for a certain road.

### 3.5.1. Road Node Elevation Change

The elevation change between two road node points was calculated by first selecting non-normalised LiDAR points at a node points within a  $1m^2$  area. Points were then filtered by those only classified as ground, and with only a single return, to reduce the chance of inaccurate  $z$  values. With several filtered points for each node, the mean  $z$  value was found for each node, and elevation change between each node was calculated. For each road the maximum elevation change was calculated, alongside the mean elevation change.

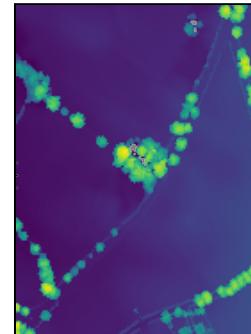
### 3.5.2. Surface Quality

Surface quality was assessed roughly through the range in intensity values found in each known road point, and the average number of returns for a road.

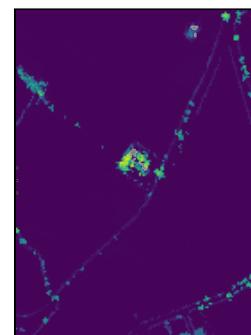
## 4. Results

### 4.1. Overview of LiDAR Data

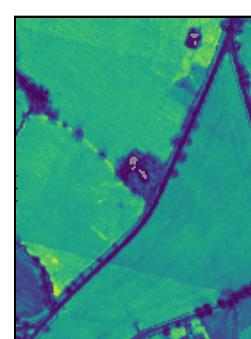
Table ?? indicates that there are likely some points with noise, particularly reflected by the highest intensity value (4063) relative to the mean value (187). Noise exclusion techniques are described in Fang & Huang (2004), this paper takes a simplistic noise filtering technique that aims to solely remove extreme outliers from the observed intensity values in the chosen data.



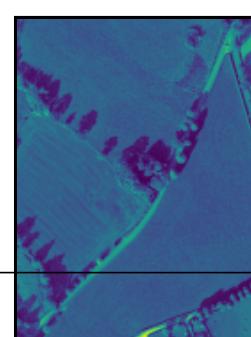
(a) Base point cloud Z values



(b) Normalised Point cloud Z values for only last returns (lpz)



(c) Aerial Data combined to 1 band



(d) Base point cloud Z values

### 4.2. Overview of Roads

Class differences, all singlecarriageway roads, therefore likely 60mph. Removed private roads. A roads + B roads + unclassified. Many unnamed roads, only identification possible through `identifier`, which splits roads by junctions.

Histograms?

### 4.3. Data Preparation

Covers script 00.

from ?? (b) intensities values for ground points below the tree are much lower than those that are not below trees, essentially if including returns with higher than 1 return the intensity values are more unreliable.

```
// find Average intensity values for first and last returns, i.e. show the shadow below trees
```

#### 4.4. Perpendicular Sampling

Covers 01

Show figure of sample lines. Show the reduction in number of points (memory saving)

#### 4.5. Linear Probability Model

Covers 02

#### 4.6. Widths and Road Quality

#### 4.7. Road Assessment

First model (maximal):

//NOTES// 1. Unfiltered vs filtered global maximal model.. Filters include no samples with multiple returns (partial canopy obstruction). excludes far too many points

2. Compare global models 1/2/3, and generalised linear model using f1.

3. Compare global model 1 (best fit), to individual model using f1.

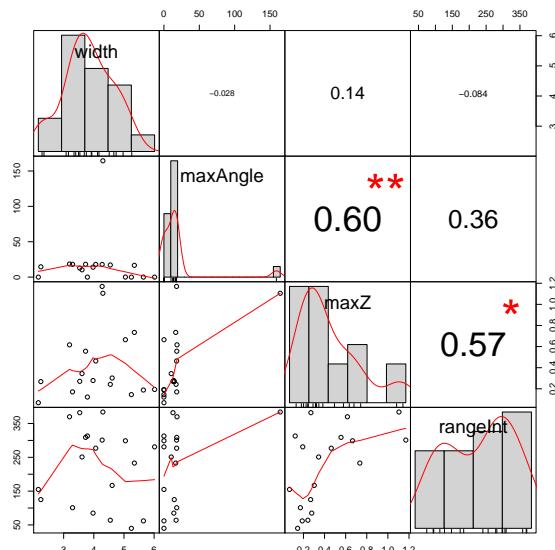
**Table 4.1: Correlation Matrix**

Variable	Width	Max Angle	Max Z	Int Range
Width	-0.03	0.14	-0.08	
Max Angle	-0.03	0.60 **	0.36	
Max Z	0.14	0.60 **	0.57 *	
Int Range	-0.08	0.36	0.57 *	

\*\*\* Significant at the 0.001 level

\*\* Significant at the 0.005 level

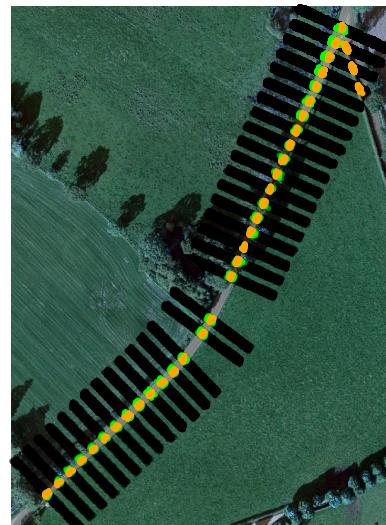
\* Significant at the 0.01 level

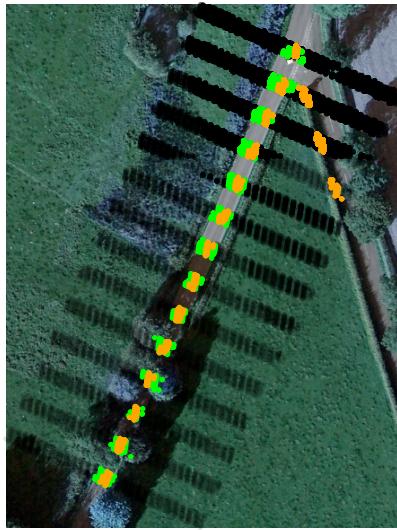


#### 4.8. (

Linear Probability Model)

```
[1] 0.9469083
[1] 0.9464282
[1] 0.957181
[1] 0.9579088
[1] 0.9583517
```





$$\begin{aligned} \text{Road}_t = & \alpha + \beta_1 \text{Intensity}_t \\ & + \beta_2 \text{Luminescence}_t \\ & + \beta_3 Z_t \\ & + \beta_4 \text{Dist}_t + \epsilon \end{aligned} \quad (4.1)$$

//for this section see 450 assess2, details on validation etc.

#### 4.9. Noise Filtering

// Manually measure a road width for a reference, i.e. how close to accurate different models are. GLM LM and individual filtered and unfiltered. Find max and min widths and mean, describe what likely causes the differences.

#### 4.10. Further Analysis

// if there is time:

Use width and max bend sharpness to estimate required speed for required stopping distance. Some maths could be involved. Use to produce speed limit assessment, include steepness somehow (speed limits due to stwwpwnss etc)

---

## 5. Discussion

==== notes ==== (?) The idea now is to use this information in a more sensitive segmentation, trying to extract the true road extents. Of course, when there is no C0 (height) or C1 (inclination) discontinuity at the road boundaries, there is no way detecting it using laser scan data, and other data sources such as aerial images have to be used. However, the question is how reliable even small discontinuities can be detected. For example, a road may be bounded by an embankment, which usually is a relatively large structure. It may on the other hand be separated from the

pavement or a traffic island by a kerb of only 15 cm in height. This seems to be hopeless, since it is close to the expected noise of the laser measurement. However, if one considers profiles perpendicular to the road, the point is that a 10 m wide road, scanned with 1 m density, will yield 10 measurements to estimate the road surface (assumed to be planar), leading to a standard deviation of about 15 cm / 105 cm.

// Mention logistic regression, considered but not used as results not better and requires a payoff.

---

## Bibliography

- Aarts, L. & van Schagen, I. (2006), 'Driving speed and the risk of road crashes: A review', *Accident Analysis & Prevention* **38**(2), 215–224. ZSCC: 0000992.
- Agency, E. (2019), 'LiDAR', <https://data.gov.uk/dataset/977a4ca4-1759-4f26-baa7-b566bd7ca7bf/lidar-point-cloud>. ZSCC: NoCitationData[s0].
- Baruya, A. (1998), 'MASTER: Speed-accident relationship on European roads'. ZSCC: 0000014.
- BBC (2012), 'Plan for 40mph country road limit', *BBC News*. ZSCC: 0000000[s0].
- Bridson, R., Marino, S. & Fedkiw, R. (2005), Simulation of clothing with folds and wrinkles, in 'ACM SIGGRAPH 2005 Courses on - SIGGRAPH '05', ACM Press, Los Angeles, California, p. 3. ZSCC: 0000550.
- Charaniya, A., Manduchi, R. & Lodha, S. (2004), Supervised Parametric Classification of Aerial LiDAR Data, in '2004 Conference on Computer Vision and Pattern Recognition Workshop', IEEE, Washington, DC, USA, pp. 30–30. ZSCC: 0000166.
- Clode, S., Kootsookos, P & Rottensteiner, F. (2004), 'The Automatic Extractin of Roads from LiDAR Data', p. 7. ZSCC: NoCitationData[s0].
- Corben, B., Oxley, J., Koppel, S. & Johnston, I. (2005), 'Cost-effective measures to improve crash and injury risk at rural intersections.', p. 10. ZSCC: 0000004.
- Department for Transport (2012), 'Guidance on road classification and the primary route network', p. 26. ZSCC: 0000005.
- Department for Transport (2013a), 'Setting local speed limits', p. 42. ZSCC: 0000004[s0].
- Department for Transport (2013b), 'The Speed Limit Appraisal Tool: User Guidance', p. 93. ZSCC: 0000000[s0].
- Department for Transport (2018), 'Road Safety Management Capacity Review'. ZSCC: NoCitationData[s0].
- Department for Transport (2019), 'Road traffic statistics - Summary statistics', <https://roadtraffic.dft.gov.uk/summary>. ZSCC: 0000000[s0].
- Department of the Environment, Transport and the (2000), 'New directions in speed management: A review of policy'. ZSCC: 0000002.
- England, H. (2019), 'Network management'. ZSCC: 0000000.
- Fang, H.-T. & Huang, D.-S. (2004), 'Noise reduction in lidar signal based on discrete wavelet transform', *Optics Communications* **233**(1-3), 67–76. ZSCC: 0000106.
- Finch, D., Kompfner, P., Lockwood, C. & Maycock, G. (1994), 'Speed, Speed Limits and Accidents', <https://TRL.co.uk/sites/default/files/PR058.pdf>. ZSCC: 0000232.
- Government, U. (2011), 'Rural Urban Classification', <https://www.gov.uk/government/collections/rural-urban-classification>. ZSCC: 0000347[s0].
- Gray, D., Farrington, J., Shaw, J., Martin, S. & Roberts, D. (2001), 'Car dependence in rural Scotland: Transport policy, devolution and the impact of the fuel duty escalator', *Journal of Rural Studies* **17**(1), 113–125. 00075.
- Gutiérrez, J. (2009), 'Transport and accessibility'. ZSCC: 0000038.
- Ihaka, R. (2011), 'Rneweb: Literate Programming with and for R', p. 5. ZSCC: 0000000.
- Knuth, D. E. (1984), 'Literate programming', *The Computer Journal* **27**(2), 97–111. ZSCC: 0002155.
- Kumar, P., McElhinney, C. P., Lewis, P. & McCarthy, T. (2013), 'An automated algorithm for extracting road edges from terrestrial mobile LiDAR data', *ISPRS Journal of Photogrammetry and Remote Sensing* **85**, 44–55. ZSCC: 0000088.
- Li, Y., Hu, X., Guan, H. & Liu, P. (2016), 'AN EFFICIENT METHOD FOR AUTOMATIC ROAD EXTRACTION BASED ON MULTIPLE FEATURES FROM LiDAR DATA', *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **XLI-B3**, 289–293.
- R Core Team (2019), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
- Rottensteiner, F., Trinder, J., Clode, S. & Kubik, K. (2003), 'Building Detection Using LiDAR Data and Multi-spectral Images', p. 10. ZSCC: NoCitationData[s0].
- Roussel, J.-R. & Auty, D. (2019), *lidR: Airborne LiDAR Data Manipulation and Visualization for Forestry Applications*. ZSCC: NoCitationData[s0].
- Taylor, M. C., Baruya, A. & Kennedy, J. V. (2002), 'The relationship between speed and accidents on rural single-carriageway roads', p. 32. ZSCC: 0000125.
- The University of Edinburgh (2019), 'Aerial Digimap Educational User Licence', [https://digimap.edina.ac.uk/webhelp/aerial/licence/digimap\\_aerial\\_eula.pdf](https://digimap.edina.ac.uk/webhelp/aerial/licence/digimap_aerial_eula.pdf). ZSCC: 0000000[s0].
- UK Government (2019), 'Find open data - data.gov.uk', <https://data.gov.uk/>.
- Velaga, N. R., Beecroft, M., Nelson, J. D., Corsar, D. & Edwards, P. (2012), 'Transport poverty meets the digital divide: Accessibility and connectivity in rural communities', *Journal of Transport*

- Geography* **21**, 102–112. ZSCC: 0000141.
- Vosselman, G. & Liang, Z. (2009), ‘DETECTION OF CURBSTONES IN AIRBORNE LASER SCANNING DATA’, p. 6. ZSCC: 0000030.
- Wang, J., Song, J., Chen, M. & Yang, Z. (2015), ‘Road network extraction: A neural-dynamic framework based on deep learning and a finite state machine’, *International Journal of Remote Sensing* **36**(12), 3144–3169. ZSCC: 0000062.
- Wickham, H. (2014), *Advanced r*, Chapman and Hall/CRC. ZSCC: 0000183.
- Xu, Y., Xie, Z., Feng, Y. & Chen, Z. (2018), ‘Road Extraction from High-Resolution Remote Sensing Imagery Using Deep Learning’, *Remote Sensing* **10**(9), 1461. ZSCC: 0000007.
- Yadav, M., Lohani, B. & Singh, A. K. (2018), ‘Road Surface Detection from Mobile LiDAR Data’, *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* **IV-5**, 95–101. ZSCC: 0000000.
- Zhang, W., Qi, J., Wan, P., Wang, H., Xie, D., Wang, X. & Yan, G. (2016), ‘An Easy-to-Use Airborne LiDAR Data Filtering Method Based on Cloth Simulation’, *Remote Sensing* **8**(6), 501. ZSCC: 0000112.

Word Count: 3427

---

## A. Code Appendix

### A.1. Session Information

```
R version 3.6.1 (2019-07-05)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Manjaro Linux

Matrix products: default
BLAS:   /usr/lib/libopenblas-p0.3.6.so
LAPACK: /usr/lib/liblapack.so.3.8.0

locale:
[1] LC_CTYPE=en_GB.UTF-8          LC_NUMERIC=C
[3] LC_TIME=en_GB.UTF-8          LC_COLLATE=en_GB.UTF-8
[5] LC_MONETARY=en_GB.UTF-8      LC_MESSAGES=en_GB.UTF-8
[7] LC_PAPER=en_GB.UTF-8         LC_NAME=C
[9] LC_ADDRESS=C                 LC_TELEPHONE=C
[11] LC_MEASUREMENT=en_GB.UTF-8   LC_IDENTIFICATION=C

attached base packages:
[1] stats      graphics    grDevices utils      datasets   methods    base

other attached packages:
[1] pbapply_1.4-1            rgdal_1.4-4
[3] future_1.14.0           varhandle_2.0.3
[5]forcats_0.4.0            stringr_1.4.0
[7] dplyr_0.8.3              purrr_0.3.2
[9] readr_1.3.1              tidyverse_1.2.1
[11] tibble_2.1.3             raster_2.9-23
[13] lidR_2.1.2               scales_1.0.0
[15] sp_1.3-1                 data.table_1.12.2
[17] kableExtra_1.1.0          ggpubr_0.2.2
[19] sf_0.7-7                 cowplot_1.0.0
[21] magrittr_1.5              viridisLite_0.3.0
[23] viridis_0.5.1             ggthemes_4.2.0
[25] broom_0.5.2               PerformanceAnalytics_1.5.3
[27] RStoolbox_0.2.6           zoo_1.8-6
[29] xts_0.11-2                ggplot2_3.2.0
[31] Hmisc_4.2-0               survival_2.44-1.1
[33] Formula_1.2-3            pacman_0.5.1
[35] lattice_0.20-38           nvimcom_0.9-83
[37] knitr_1.24                colorout_1.2-1
[39] colorout_1.2-1

loaded via a namespace (and not attached):
[1] colorspace_1.4-1     ggsignif_0.6.0    class_7.3-15
[4] htmlTable_1.13.1    base64enc_0.1-3   rstudioapi_0.10
[7] listenv_0.7.0       rlas_1.3.2       prodlim_2018.04.18
[10] lubridate_1.7.4     xml2_1.2.1       codetools_0.2-16
[13] splines_3.6.1      doParallel_1.0.15 zeallot_0.1.0
[16] jsonlite_1.6        caret_6.0-84     cluster_2.1.0
[19] rgeos_0.5-1         compiler_3.6.1   httr_1.4.1
[22] backports_1.1.4    assertthat_0.2.1 Matrix_1.2-17
[25] lazyeval_0.2.2      cli_1.1.0       prettyunits_1.0.2
[28] acepack_1.4.1       htmltools_0.3.6   tools_3.6.1
[31] gtable_0.3.0        glue_1.3.1       reshape2_1.4.3
[34] Rcpp_1.0.2           RCSF_1.0.1       cellranger_1.1.0
[37] vctrs_0.2.0          nlme_3.1-140    iterators_1.0.12
[40] timeDate_3043.102   gower_0.2.1     xfun_0.8
[43] globals_0.12.4      rvest_0.3.4     XML_3.98-1.20
```

```
[46] MASS_7.3-51.4      ipred_0.9-9      hms_0.5.0
[49] parallel_3.6.1     RColorBrewer_1.1-2 geosphere_1.5-10
[52] gridExtra_2.3      rpart_4.1-15    latticeExtra_0.6-28
[55] stringi_1.4.3      highr_0.8       foreach_1.4.7
[58] e1071_1.7-2        checkmate_1.9.4 lava_1.6.6
[61] rlang_0.4.0         pkgconfig_2.0.2  evaluate_0.14
[64] labeling_0.3        recipes_0.1.6   htmlwidgets_1.3
[67] tidyselect_0.2.5    plyr_1.8.4      R6_2.4.0
[70] generics_0.0.2      DBI_1.0.0       pillar_1.4.2
[73] haven_2.1.1         foreign_0.8-71  withr_2.1.2
[76] units_0.6-3         nnet_7.3-12    modelr_0.1.5
[79] crayon_1.3.4        KernSmooth_2.23-15 rmarkdown_1.14
[82] progress_1.2.2      grid_3.6.1      readxl_1.3.1
[85] ModelMetrics_1.2.2  digest_0.6.20   classInt_0.4-1
[88] webshot_0.5.1       stats4_3.6.1   munsell_0.5.0
[91] quadprog_1.5-7
```

## A.2. Functions

### A.2.1. Formatting

```
make_table <- function(df, cap, dig = 2, ...) {
  require(kableExtra)
  require(tidyverse)

  options(knitr.kable.NA = "")
  kable(df,
    digits = dig, caption = cap,
    linesep = "",
    longtable = FALSE, booktabs = TRUE,
    format = "latex",
    escape = F
  ) %>%
    kable_styling(font_size = 8)
}
```

### A.2.2. Catalog to Dataframe

```
ctg_to_df <- function(cluster) {
  las <- readLAS(cluster)
  if (is.empty(las)) {
    return(NULL)
  }
  las <- as.spatial(las)
  las <- as.data.frame(las)
  return(las)
}
```

### A.2.3. LiDAR Clean

```
lidr_clean <- function(cluster) {
  las <- readLAS(cluster)
  if (is.empty(las)) {
    return(NULL)
  }
  epsg(las) <- 27700
  # remove all but last return
  las <- lasfilter(las, NumberOfReturns == ReturnNumber)

  # find ground points
  las <- lasground(las, csf())

  ## Create Point DEM
  # interpolate ground points to create raster dtm. Uses Classification = 2
  # very large number of points, therefore idw used as opposed to kriging
  dtm <- grid_terrain(las, 1, knnidw(k = 10, p = 2))
  # normalise heights using dtm
  las <- lasnormalize(las, dtm)
  return(las)
}
```

### A.2.4. Filter Noise

### A.2.5. Extract Buffer

```
extract_buff <- function(cluster, clip_input) {
  las <- readLAS(cluster)

  if (is.empty(las)) {
```

```

        return(NULL)
    }

    if (!is.null(clip_input)) {
        las <- lasclip(las, clip_input)

        if (length(las) > 1) {
            for (i in 1:length(las)) {
                if (!is.empty(las[[i]])) {
                    las <- do.call(rbind, las)
                    return(las)
                }
            }
        }
    }
}

```

#### A.2.6. Euclidean Distance

```

# Function to calculate Euclidean distance between 2 points
euclidean_distance <- function(p1, p2) {
    return(sqrt((p2[1] - p1[1])**2 + (p2[2] - p1[2])**2))
}

```

#### A.2.7. Perpendicular Sampling

```

# Function to calculate 2 points on a line perpendicular to another defined by 2 points p1,p2
# For point at interval, which can be a proportion of the segment length, or a constant
# At distance n from the source line
calc_perp <- function(p1, p2, n, interval = 0.5, proportion = TRUE) {
    # Calculate x and y distances
    x_len <- p2[1] - p1[1]
    y_len <- p2[2] - p1[2]

    # If proportion calculate reference point from tot_length
    if (proportion) {
        point <- c(p1[1] + x_len * interval, p1[2] + y_len * interval)
    }
    # Else use the constant value
    else {
        tot_len <- euclidean_distance(p1, p2)
        point <- c(
            p1[1] + x_len / tot_len * interval,
            p1[2] + y_len / tot_len * interval
        )
    }

    # Calculate the x and y distances from reference point
    # to point on line n distance away
    ref_len <- euclidean_distance(point, p2)
    xn_len <- (n / ref_len) * (p2[1] - point[1])
    yn_len <- (n / ref_len) * (p2[2] - point[2])

    # Invert the x and y lengths and add/subtract from the reference point
    ref_points <- rbind(
        point,
        c(point[1] + yn_len, point[2] - xn_len),
        c(point[1] - yn_len, point[2] + xn_len)
    )

    # Return the reference points
    return(ref_points)
}

```

### A.2.8. Combine Catalog

```
comb_ctg <- function(x) {
  las <- readLAS(x)
  if (is.empty(las)) {
    return(NULL)
  }
  return(las)
}
```

### A.2.9. Compute Samples

```
sample_lines <- c()
compute_samples <- function(x) {
  if (nrow(x) > 1) {
    road_node <- st_coordinates(x)
    tot_len <- 0
    len_inc <- 10
    len_ofs <- len_inc
    for (i in 2:nrow(road_node) - 1) {
      n1 <- road_node[i, ]
      n2 <- road_node[i + 1, ]

      len_seg <- euclidean_distance(n1, n2)
      len_ofs <- len_ofs + len_inc

      while (len_ofs <= tot_len + len_seg) {
        len_ofs <- len_ofs + len_inc

        # Add results to output vector
        perp_segments <- calc_perp(
          n1, n2, 30,
          len_ofs - tot_len,
          proportion = FALSE
        )

        multipoints <- st_multipoint(matrix(perp_segments, ncol = 2))
        pts <- st_cast(st_geometry(multipoints), "POINT")
        n <- length(pts)

        pair <- st_combine(c(pts[1], pts[2], pts[3]))
        linestring <- st_cast(pair, "LINESTRING") %>%
          st_buffer(2) %>%
          st_sf() %>%
          mutate(road_id = as.character(unique(x$road_id)))
        sample_lines <- rbind(sample_lines, linestring)
      }
      tot_len <- tot_len + len_seg
    }
  }
  return(sample_lines)
}
```

### A.2.10. Greyscale

```
greyscale <- function(x) {
  x <- (x[[1]] + x[[2]] + x[[3]]) / 3
}
```

### A.2.11. Compute Linear Model

```
lm_compute <- function(x, f) tryCatch({  
  m <- lm(formula = f, data = x)  
  
  p <- m %>%  
    tidy() %>%  
    dplyr::select(p = p.value)  
  
  pred_m <- predict(m, x, type = "response")  
  
  if (sum(p) / nrow(p) < 0.05) {  
    x$lm <- pred_m  
  } else {  
    x$lm <- NA  
  }  
  
  x$lm_individual <- ifelse(x$lm > quantile(x$lm, .95), 1, 0)  
  x$pred_lm_individual <- ifelse(x$lm > quantile(x$lm, .95), 1, 0)  
  x$p_val <- sum(p)  
  
  return(x)  
}, error = function(e) NULL)
```