
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

1	Introduction	1
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
2	Literature	5
2.1	British Rural Road Network	5
2.1.1	Rural Speed Limits	6
2.1.2	Speed, Road Geometry and Accidents	7
2.1.3	Accessibility	7
2.1.4	Road Maintenance	8
2.2	LiDAR Data Classification	8
2.2.1	Digital Terrain Models	8
2.2.2	Feature Classification	9
2.3	Road Classification	9
2.3.1	LiDAR Road Classification	10
2.3.2	Rural Road Extraction	12
2.4	Supervised Modelling Image Segmentation	13
3	Methodology	14
3.1	Data	14
3.2	LiDAR Preprocessing	14
3.2.1	Last Pulse	15
3.2.2	Normalisation	15
3.2.3	Points Extent	16
3.3	Road Analysis	16
3.3.1	Road Sampling	16
3.4	Aerial Imagery	16
3.4.1	Linear Probability Model	17
3.5	Road Angles	17
3.5.1	Road Node Elevation Change	18
3.5.2	Surface Quality	18
3.5.3	Estimate of True Widths	18

4	Results	19
4.1	Overview of Data	19
4.1.1	LiDAR Point Cloud	19
4.1.2	OS Road Geometry	19
4.2	Data Preparation	19
4.3	Perpendicular Sampling	20
4.4	Linear Probability Model	20
4.5	Noise Filtering	20
4.6	Final Model Analysis	21
4.7	Road Assessment	21
4.8	Widths and Road Quality	21
4.9	Road Analysis	21
5	Discussion	22
A	Code Appendix	25
A.1	Session Information	25
A.2	Functions	26
A.2.1	Formatting	26
A.2.2	Catalog to Dataframe	26
A.2.3	LiDAR Clean	26
A.2.4	Filter Noise	26
A.2.5	Extract Buffer	26
A.2.6	Euclidean Distance	27
A.2.7	Perpendicular Sampling	27
A.2.8	Combine Catalog	28
A.2.9	Compute Samples	28
A.2.10	Greyscale	28
A.2.11	Compute Linear Model	28

List of Figures

1.1	Study area highlighting road locations	3
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	17
3.2	Bearing Angle Between Road Segments	18
3.3	Bearing Angle Between two Road Segments	18
4.1	LiDAR point clouds for one selected road section aggregated into $1m^2$ grids, (A) Base point cloud <i>Z</i> values, (B) Normalised Point cloud <i>Z</i> values for only last returns (<i>lpz</i>) (C) Normalised Point cloud <i>Intensity</i> values for last return, (D) Aerial Data combined to 1 band	19

List of Tables

3.1 LiDAR Point Cloud Summary Data	15
3.2 OS Roads Data Summary	16
4.2 Correlation Matrix	21

1. Introduction

ROAD usage within the United Kingdom has been steadily increasing by year with the highest ever billion vehicle miles (bvm) travelled in 2018 (255 million; Department for Transport 2019). Characterised by tall hedgerows and winding turns, rural roads in the UK are often unsuitable for higher traffic flow due to the obstruction of view due to the protected hedgerows, narrow lanes and often poor condition. Due to 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 (UK Government 2019b). 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 (Highways 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 (UK Government 2011), therefore a road could be considered rural, if either connecting or present within small settlements in the UK. This study will focus particularly on rural connecting roads, outside of rural towns.

A Governmental review of speed policy considered the need for the role of speed and accidents on rural roads to be further considered (?), suggesting a

framework for individual 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 suggested 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, likely due to the costs involved in a blanket change in speed limits. For example, the estimated cost for a complete change in national speed limits from imperial to metric in Ireland cost an estimated €30 million in speed limit signs alone (Noctor 2004). These costs suggest that an alternative to blanket implementation is required.

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.

(UK Government 2019b)

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, bend, 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

The Road Safety Management Capacity Review (Department for Transport 2018) outlines the current limitations with road safety management, with the lack of defined and measureable safety performance framework, noting that such a framework should set out the long term goal of total prevention of road deaths and injuries, achieving this through a reduction 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.

A recent development for guidance in setting local speed limits is the production of the *Speed Limit Appraisal Tool* (Department for Transport 2013b). This tool provides an automated method for the introduction of new speed limits for local councils. This tool

takes observed traffic flow, accidents, speeds, descriptive information regarding the network and current costs, outputting predicted changes in these data for various speed limit changes. While this tool introduces some quantitative method for individual road speed limit assessments, it misses some key features outlined in past government framework proposals (Department for Transport 2018, e.g.), particularly in relation to road geometry.

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 (Harter 2005).

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 *xy* 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 ob-

jects 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.



Fig. 1.1: Study area highlighting road locations

1.2. Objective of this paper

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 probability 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.

Key Aims:

- Produce and assess an automated method using LiDAR, aerial imagery and OS road geometry 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.
- Assess the potential for speed reduction on rural roads, with the assumption that at present the roads chosen are 60mph.

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.

This paper is organised into chapters, first a literature review, outlining the broad implications of speed

limits, rural road networks, and object extraction particularly in relation to LiDAR aerial point clouds. Second, a detailed description of the methodology involved in this paper will outline the techniques used to classify road widths, in addition to the other road geometric information. A results section will primarily assess the method for road extraction, through sensitivity analysis and some qualitative observations, a short section will then explore the findings. Finally a discussion will detail the implications for the findings, and suggest areas for improvement.

// improve this, why am i doing this, e.g. production of method will aid speed limits, emergency response etc.

// Include more infrastructure to support the overarching narrative argument. E.g. The intro. should include a brief overview of the argument (sections) to come [signposting], and there should be more linking sentences between sections to help the reader understand both that the focus is being changed and why.

2. Literature

Typical road extraction techniques have focused purely on urban road networks and involved methods which can be both computationally intensive and time consuming. Given the pressure for full quantitative assessments of the current speed limits in the rural network in the United Kingdom, there is a demand to produce comprehensive methods for rural road feature extraction. This paper primarily focuses on techniques for assessing the road geometry for roads considered to be rural connecting roads in the United Kingdom. This literature review will first outline the current understanding of the rural road network, considering the role of speed and speed limits in accident likelihood, rural accessibility, and a detailed look at current road extraction techniques involving aerial imagery and LiDAR, presenting the key differences and limitations of these studies when considering the rural road network in the UK.

Road Geometry, from documentation of both road geometry and road environment. Road geometry means the parameters used for the geometric design of roads, such as design speed, stopping sight distance, passing sight distance, line of sight, number of lanes, lane width, side foot path for pedestrians, longitudinal and transverse slope, road pavement materials, etc. Recently built roads follow modern geometric design parameters, and most of this information is already available, but for the oldest roads these data, which are required mostly for upgrading or assessing the quality of the roads, are unavailable or unusable. By road environment we refer to the vicinity of the road on both sides of the road, including buildings, trees, vegetation, electricity, traffic signs, traffic light poles and other objects [1]. We expect that the geometry of this kind of road environment can best be obtained either using airborne or vehicle-based laser scanning or using an integrated approach combining

both systems. Supported with imaging sensors, the documentation can be completed with texture information, traffic sign semantics, etc.

2.1. British Rural Road Network

- How many roads are "rural", how much traffic Describe for the UK.

Taylor et al. (2002) conducted a comprehensive study outlining the key features of British rural roads, in an effort to improve the characteristics associated with accident rates, above that which was explored by the MASTER study which primarily consisted of European road data, with limited data for England (Baruya 1998). Taylor et al. (2002) identify key features of 174 selected rural british roads across England which they use to classify roads into certain categories, this data was obtained through drive through video recordings. The following data was obtained and used to categorise the quality of the road;

Discrete data

- Type of junction (if present);
- Number and type of minor junctions;
- Number and type of accesses within the section;
- Number of bends, classified into:

Sharp (bends with a warning sign);

medium;

slight

Semi-continuous data

- Lighting;
- Reflecting road studs;
- Kerbs;

- Number of lanes;
- Road markings;
- Land use.

Continuous data

- Visibility;
- Verge width and type;
- Roadside type;

Notably, the features unique to rural roads were observed to be the land use either side of a road, consisting mainly of residential, farming, wooded, open or industrial. Verges consisting of grass verges, pavement, low banks, ditches or none. Roadside type included the most dominant vertical feature closest to the road, including trees which may overhang the road, hedges, banks, fences and so on. This study also manually measured the road width for each site, and to determine the "hilliness" of roads, the number of 10m contour lines crossed were counted to give the total change in height.

Taylor et al. (2002) found that the average width of British rural roads varied from 2m to 10.2m with a mean width of 6.5m. Most were two lane roads, however a small number were single track. The mean length of roads was 3km, around half the sites had more than 3 bends per km with 1 third showing 1 severe bend per km. Most roads were flat, showing less than a 10m rise per km.

Road quality in this assessment was determined through categorising roads into four key groups:

- Group 1: Roads which are very hilly, with a high bend density and low traffic speed. *These are low quality roads.*
- Group 2: Roads with a high access density, above average bend density and below average traffic speed. *These are lower than average quality roads.*
- Group 3: Roads with a high junction density, but below average bend density and hilliness, and above average traffic speed. *These are higher than average quality roads.*

- Group 4: Roads with a low density of bends, junctions and accesses and a high traffic speed. *These are high quality roads.*

This study therefore attempted to outline a rural road hierarchy in relation to road function, and certain road geometries, which addresses issues outlined in the Government's review of speed policy (?). However, due to the nature of the data collection for this study, time constraints mean that producing a full road hierarchy for all rural roads within England is impractical and requires a review of the methodology. Additionally the speed policy review notes that road hierarchies should inform appropriate speeds for each road.

See (Taylor et al. 2000)

2.1.1. Rural 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* when 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, often considered dangerous driving. A study by the Department for Transport (2013a) assessed the impact of inappropriate speed on rural roads, which contributed to 20% of all crashes on minor rural roads with a 60mph limit, whereas excess speed only accounted for around 16% of collisions. An observation of 270 single carriageway rural roads in England found that the mean distribution of speeds had a wide distribution, often significantly below the 60mph limit (Department for Transport 2006).

The Department for Transport found that rural roads account for around 66% of all road deaths, despite accounting for around 42% of the total distance travelled by all vehicles. Notably 51% of all deaths in Britain in 2011 occurred on rural single carriageway roads, with the national speed limit of 60mph (Department for Transport 2011).

2.1.2. Speed, Road Geometry and Accidents

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) found that a reduction in the speed limit of a road resulted in a mean speed reduction of around one quarter of the difference, noting that drivers will often obey speed limits that they determine to be reasonable. A reduction in average speed subsequently leads to a reduction in road traffic accidents (?). Taylor et al. (2000) produced the EURO model to predict accident frequencies given the proportion of drivers exceeding the speed limit and the average speed, finding that excess speed and a higher speed limit were both associated with a higher accident frequency. Particularly, the risk of death at various speeds has been assessed in various studies,? found that at 60mph the risk of a driver dying in a head on collision between two cars is around 90%, but with a reduction in speed, this drops to around 50% at 48mph.

Taylor et al. (2000) demonstrated that traffic flow, link length, and the number of minor junctions all directly increased the number of accidents, while wider roads were associated with a reduction in the number of accidents. The *MASTER: Speed-accident relationship on European roads* (Baruya 1998), assessed road geometry and other features of roads, however road data for the United Kingdom was limited to a small area in the South East, suggesting that a comprehensive methodology for the extraction of UK rural road geometry is required for a more comprehensive study.

Developed speed tool: (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.

2.1.3. Accessibility

Rural accessibility can be defined in terms of economic and social opportunity as the proximity or ability for spatial interaction (Gutiérrez 2009). In relation to transport specifically, accessibility may be defined as the ability or opportunities by which basic services can be reached by either public or private transportation (Gutiérrez 2009).

Journey time on rural roads is often a primary concern when considering a reduction in speed limits, often higher speed is perceived to bring with it shorter travel times, and greater accessibility for people and goods (Department for Transport 2013a). Despite this, there is evidence to support that traffic travelling at constant, and lower speeds may result in overall more reliable journey times, and the time saved by travelling at faster speed is often overestimated (Stradling et al. 2008).

Despite this, 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 (Mungall 2005). 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 (Nicholl et al. 2007). 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.

Velaga et al. (2012) conducted a study outlining the key transport and technological challenges that limit accessibility within rural areas, demonstrating that the quality of rural infrastructure, and access to services are key factors driving social exclusion and limited access to goods and services. Velaga et al. (2012) suggest that technological innovations targeting rural areas may alleviate this problem through the production of user generated data, they note that certain challenges will come with this technology, notably their first point;

Understanding basic technological infrastructure requirements in rural areas.

Suggesting a greater understanding of the existing rural road infrastructure is required before the introduction of existing rural technologies may improve access to rural areas.

More on emergency response.

// expand all, summarise the mentioned papers.

2.1.4. Road Maintenance

(Yadav et al. 2018): road surface three dimensional information important for pavement condition assessment, road maintenance citep Mcelhinney2010; yang2013

(?) information can be obtained with airborne laser scanning at low altitudes. When the road has been constructed, documentation of road information is needed for an increasing number of other applications, such as noise modelling, road safety, road maintenance, location-based services, and car and pedestrian navigation. Documentation of the road environment includes documentation of both road geometry and road environment. Road geometry means the parameters used for the geometric design of roads, such as design speed,

2.2. LiDAR Data Classification

Aerial LiDAR classification typically follows two objectives, the classification of ground and non-ground points, and the classification of surface objects, including buildings, trees or roads (Charaniya et al. 2004). Classification takes two forms, *supervised* and *unsupervised*, supervised classification taking a *training* dataset and using it to estimate the parameters associated with the outcome hoping to be classified. These parameters are then used on unknown data, with a similar distribution to the training set, and used to classify features (Charaniya et al. 2004).

Most past work on the classification of LiDAR data has been produced through unsupervised classification, where no training dataset is provided, often producing a coarse classification than could be obtained through supervised classification.

from (?):

2.2.1. Digital Terrain Models

Early LiDAR classification primarily focused on the production of digital terrain models (DTM), Kraus & Pfeifer (1998) used an iterative linear prediction algorithm which used residuals to compute weights for each point. Ground points produced residuals with negative weighting, and vegetation was more likely to produce residuals with higher weighting. From this 47.8% of points were classified as vegetation, from this they note that advances in the technology, such as the inclusion of multiple laser pulses enable for a higher quality DTM.

Vosselman (2000), proposed a laser pulse filtering technique for distinguishing laser points that were reflected from buildings and vegetation, than those from ground, weights were assigned to points, determined by an assigned acceptable height difference between two neighbouring points. Limitations at the time of this study meant that processing requires points to be aggregated into a grid for processing, resulting in significant information loss. Results were generally good for the simplicity of the method, giving a RMS error of 20-30cm, however Vosselman (2000) note that this error likely does not reflect the true precision of the DEM, which may be slightly worse.

The observation of height textures are another method for DEM extraction enabled through the use of LiDAR data. ? used height variation between neighbouring LiDAR points to classify buildings, trees, and flat terrain, with enough detail to determine whether a building had a flat or sloped roof. Similarly Elberink & Maas (2000) noted that man made structures often have smooth, regular height textures, with small variations in height, while trees and other vegetation give an irregular height pattern. Empirical assessment of the accuracy of their technique resulted in a 98

Zhang et al. (2003) used morphological filtering of points assigned into a regular grid by inputting a minimum elevation for each cell in the grid, and interpolating the elevation of cells not containing any points. Accuracy of the results were assessed both quantitatively and qualitatively, giving 3% error from misclassified points, however observation of the DEM showed artefacts primarily from larger buildings.

This paper will utilise a recent method for DEM

production in order to classify ground and non ground points for subsequent road classification. The method chosen was proposed by Zhang et al. (2016) uses *cloth simulation* to generate a DTM from LiDAR data. This algorithm, unlike other filtering algorithms, allows for a simplistic input, without the need for numerous parameters to ensure an accurate DTM. This method consists of four main steps:

- *Initial State* A cloth is placed above the inverted LiDAR measurements.
- The displacement of each LiDAR point is calculated under the influence of gravity, meaning some points appear below ground measurements.
- *Intersection check*. For any points detected as being under the ground, they are moved to ground level and set to be unmovable.
- *Considering internal forces*. Movable points are moved according to neighbouring points.

Quantitative accuracy of this methodology gave results similar to top existing DTM production algorithms, but with a far more simplistic implementation.

2.2.2. Feature Classification

Developments in LiDAR enabled the possibility of classification by using laser intensity information and multiple returns, features of more advanced LiDAR systems. The TopEye system used by ? allowed for classification of buildings and electrical power-lines using reflectance ($1/intensity$) to obtain radiometric information about the area and note that this can be used to separate paved area from grassland. Power lines in particular benefited from the multiple returns produced by the LiDAR system used as they often gave one return from the power line, and one from ground.

// if i want more,... Maas [12] height texture for segmentation of lidar. Filin [5] surface clustering technique for identifying regions in lidar exhibiting homogeneity in certain feature space consisting of position, tangent plane, and relative height difference attributes for every point. the surface are categorised as hiehg vegetation, (height jumps), low vegetation, smooth surfaces and planar surfaces // note that smooth surfaces often man made (find ref). song

[17] separation of materials, trees grass asphalt, roads, and roofs on intensity data, interpolated using three trechniques, idw, median filtering and kriging. Hebert [7] presented outling or some classification. Approaches using mixture models [14,10,4,3] for parametric classification. Maceda [9] ground based laster discriminting grass and rocks. Inclusion of aerial imagery, improvement of classification [2], or others [13,15].

Ground based lidar: (Yadav et al. 2018)

2.3. Road Classification

// small history of road extraction here // list varying resolution increases per papers? // timeline of aerial image quality/lidar

short intro:

Much tougher than object extraction from surface, height differences less prominent, less relevant for road mapping. vosselman2009a

Many proposed road classification methods have used purely unsupervised...

Road classification is essentially a data clustering method to categorise data into road and non-road points, enabled through discovering patterns and relationship between variables and validation of findings (Saeedi et al. 2009) Clustering may be achieved through various algorithms, categorised generally into partitioning methods, hierarchical methods, density-based methods, grid-based methods, and model-based methods (Saeedi et al. 2009). Yadav et al. (2018) 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. Due to the heterogeneous nature of certain roads types, the road environment is often complex, meaning collection and accurate processing of road features is challenging(Yadav et al. 2018).

Road classification methodologies have historically used purely aerial imagery, providing only road pixels and 2D location information (Yadav et al. 2018). Ferchichi & Shengrui Wang (2005) used high resolution satellite imagery for road centreline extraction from a suburban area based on cluster coverage. They used image segmentation through maximum likelihood to assign pixels to road and non-road

classes, making the assumption that both classes gave a Gaussian distribution. Features for classification include the angular second-movement, contrast, and entropy of the image, based off image texture analysis by [Dubes & Ohanian \(1992\)](#). Results of the classification gave significant noise, noted as a possibility due to non conformance with a Gaussian distribution, however, centreline extraction from the resultant classification removed any pixels with low density and ran a *K*-means clustering algorithm iteratively to determine road clusters. [Wan et al. \(2007\)](#) produced an automated method for mapping urban and suburban roads using high resolution satellite imagery, using spectral, context, shape and the structural features or roads. Using fuzzy segmentation, classification of buildings, parking areas, and road clusters were obtain, but with large levels of noise. Angular texture analysis was able to reduce this noise, however large areas of incorrectly classified roads appear present results. Additionally, while not mentioned in this study, the method proposed fails to address the differentiation between road and pavement, a key issue with the use of 2D imagery.

2.3.1. LiDAR Road Classification

Several methods for road classification from LiDAR follow similar patterns, first identification of ground points through height data, and DTM extraction through interpolation, and subsequent classification based on the attributes of the point cloud [Vosselman \(2009\)](#). Due to the 3D information provided by LiDAR data, kerbs can be segmented from roads, achieving a more accurate road edge location when kerbs are present. However, due to the often small jump between kerb and road, this methodology is limited to new terrestrial mobile LiDAR data collection.[?](#) produced a methodology using mobile LiDAR data to classify road edges by segmenting kerbstones based on the height gradient along the scanned profile, with a final completeness of 73.9%, and correctness of 85.6%. Additionally, road markings were segmented from the road surface using intensity information. The mobile data collection for this study obtained a point resolution of between 10cm² and 100cm², with driving speeds of 20km/h to 50km/h. [Jaakkola et al. \(2008\)](#) note that due to the size of the data produced through

this method, computation times are lengthy and algorithms have to ensure efficiency. [Yoon & Crane \(2009\)](#) similarly developed an approach for urban drivable terrain detection for automated vehicles through mobile LiDAR data, road edges were obtained through the slope and standard deviation of the height of points, finding the standard deviation to be far more efficient for edge detection. They note that while results were promising, it would take a very long time to collect data for a large number of roads, and at higher vehicle speeds, LiDAR points density is reduced below a usable threshold.

[Vosselman & Zhou \(2009\)](#) used low flying airborne LiDAR point cloud data with a very high density of 10-20m²

Road Edge Detection:

[Vosselman2009](#), developed a method for detecting kerbstones from airborne LiDAR data, based on the detection of small height jumps caused by the kerbstones in the LiDAR point cloud data. Cars parked occluded kerbstones. [Smadja2010 \(Smadja et al. 2010\)](#), algorithm for extracting roads from Lidar data based on the detection of slope break points coupled with ransac algorithm, road boundaries further processed to compute road curvature and road width information. [zhang2010](#) detecting road edges in urban environment using terrestrial lidar data, edge points identified based on elevation information. identified 3d road edges points projected on a ground plane to estimate road kerbs. [McElhinney2010](#), extracting road edges from terrestrial model lidar data, first set of lines fitted to the road cross sections based on the navigation data and then lidar points within the vicinity of the lines were determined. second stage, the points were analysed based on information such as slope, intensity, pulse width and proximity to vehicle information in order to extract the road edges. [Ibrahim2012](#), street kerb and surface from terrestrial mobile lidar point cloud data, segmented to ground and non ground based on point density, further refined analysing the morphological characteristic of the neighbourhood of each point in the segment, kerb edges finally extracted from the refined ground segment by applying the derivative of the gaussian function to 3d points. [Clode et al. \(2004\)](#) used a Hough transformation to classify roads from aerial LiDAR data. Data was first aggregated into Road pixel must

lie on or near DTM, and have a certain intensity and normalised local point density. **My paper will extend this, inclusion of road shapefiles to remove known non roads (supervised), 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.** 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.

Create DTM from last pulse DSM Intensity identifies the type of road material (bitumen in their paper) This issue may be alleviated with the use of LiDAR data, where the majority of road extraction methodologies use height variation to delineate roads from kerbs.

// manmade smooth? can't find ref now...

See (Kumar et al. 2013):

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: algorithm using fuzzy clusing 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 correspondng road polygons using a LiDAR based seed-growing algorithm then mapped polygons lam2010, terrestrial mobile lidar data random sample consensus (RANSAC) planes to small sections then interconnected these fitterd planes using Kalman filtering. Further filtering powerlines etc.

Majority attempt to delineate roads by distinguishing them from non-road objects but do not attempt to extract the road edges. Most for urban roads, rely on sufficient height or slope difference between the road and kerb points for ectecting road sedges. Little or no research been carried out to extract rural roads, wheer

non-road surface comprises grass-soil and edges not as easily defined by slope changes alone. Intensity and pulse width attributes from liDAR data can be a useful source of intermation for extracting these roads edges.

An overview of road extraction techniques from aerial LiDAR data are described in Hatger (2005), the detection of straight lines within data, to derived feature outlines using a Hough transformation. Region growing algorithms, which take small regions identified as being true parts of the road surface, then growing these areas by adding neighbouring points with similar properties, and hierarchical classification, which splits the data into partitions that follow a common predicate, removing data sequentially. // reword this.

From: (Hatger 2005) - 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 Shapira, 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 arass to the left and right of the structure (brugelmann,2000, wild and krzystek 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 General Planar Region growing Segmentation Iterating following: 1. Find the best seed region which fulfils 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 fulfil 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 $y = y_0$ fulful the equatin $z = as + by_0 + d$ 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

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

- 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 distribute irregularly with non uniformity density. the point density in the overlapping area between flight strips is greater than non-overlapping regions. There is no point in areas occluded by tall objects. can be solved 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 asphalt/soil and asphalt vegetation and asphalt/grassy bank. Exact edge location difficult if point density less and sufficient elevation 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 difference between road surface and shoulder not obvious.

Clode 2004:

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

- from (Vosselman & Zhou 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 are known from a road database. Including height longitudinal and transversal slope, curvature, and width. Height profiles across road split into straight segments. A RANSAC procedure on the end points of the segments with low slope values is used to fit straight lines to the road sides. (?)

2.3.2. Rural Road Extraction

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.

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) (Wu et al. 2013)
- Junction Sharpness (Aarts & van Schagen 2006)

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

2.4. Supervised Modelling Image Segmentation

3. Methodology

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^AT_EX** combined with the **.rnroweb** file type (Ihaka 2011), for *Literate Programming*¹. 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^AT_EX**. **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

¹See Knuth (1984)

- Commercial and Non-commercial use of the information

LiDAR data used in this paper is available [HERE](#) under this licence (UK Government 2019a). 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 500m²), 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 3.1. 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. \(2005\)](#) and subsection 2.2.1. 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), Δt 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

	Example
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 len_n x value from p_k , while adding the len_n y value, and vice versa. These perpendicular sample lines were buffered to a width of 1m, giving a total area of 20m² 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.

$$luminescence = (Band_1 + Band_2 + Band_3) \div 3$$

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;

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

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 ??;

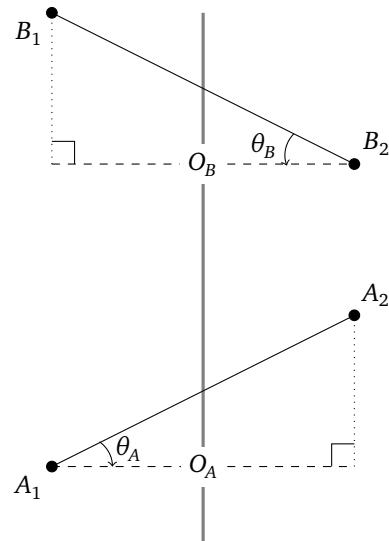


Fig. 3.1: Road LiDAR points at maximum distance apart for each sample location. Showing two example sample locations (A and B), road centreline represented by the thick grey line. True road width is indicated by the dashed lines O_A and O_B .

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* θ , 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;

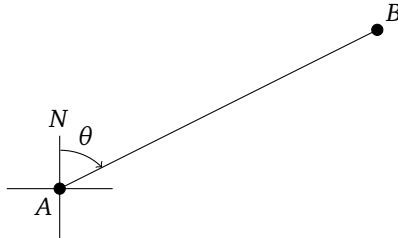


Fig. 3.2: Bearing Angle Between Road Segments

To find the angle θ , 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 θ gives;

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

This equation can be rewritten to calculate the value of θ 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;

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

??

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 $\theta_k = \theta_{N(N+1)} - \theta_{(N-1)N}$.

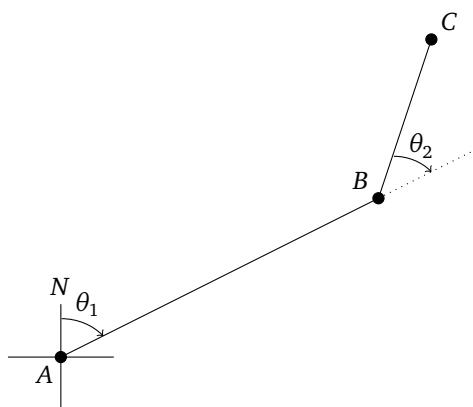


Fig. 3.3: Bearing Angle Between two Road Segments

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 geometric node within a $1m^2$ area. Nodes were obtained from the OS road data, as point coordinates obtained when the linestrings were split into points. LiDAR points were then filtered by those only classified as ground, and with only a single return, to reduce the likelihood of inaccurate z values from canopy or other vegetation and vehicles. The mean z value of points were 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 for sample lines not obstructed by canopy, indicated by single returns.

3.5.3. Estimate of True Widths

QGIS used to measure roughly the width at various points along each road using 25cm resolution aerial imagery.

4. Results

4.1. Overview of Data

4.1.1. LiDAR Point Cloud

Table 3.1 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.

4.1.2. OS Road Geometry

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?

4.2. 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
```

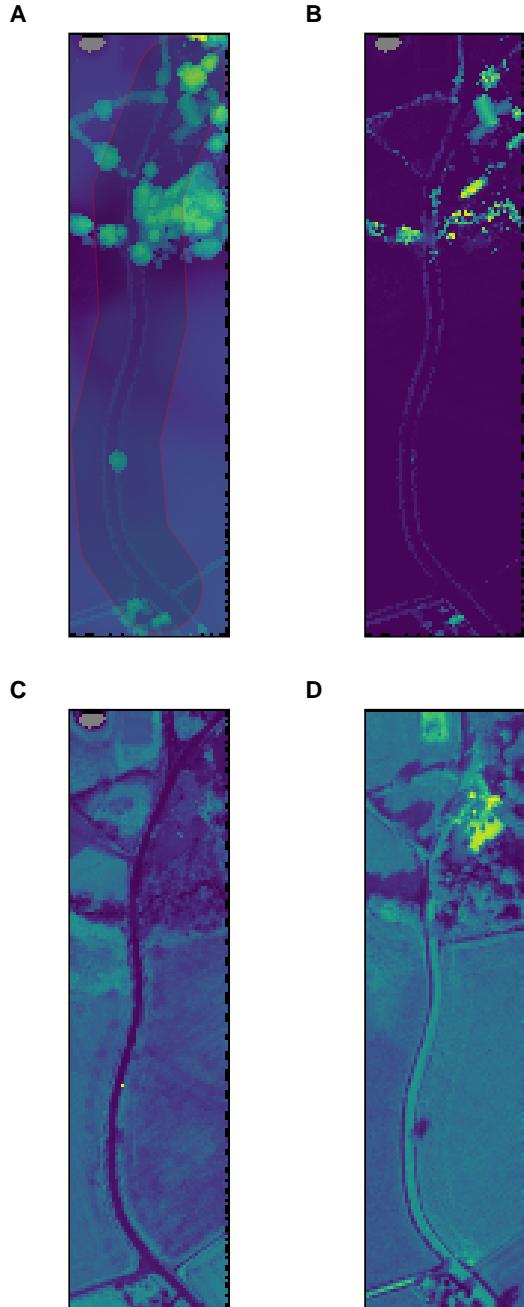
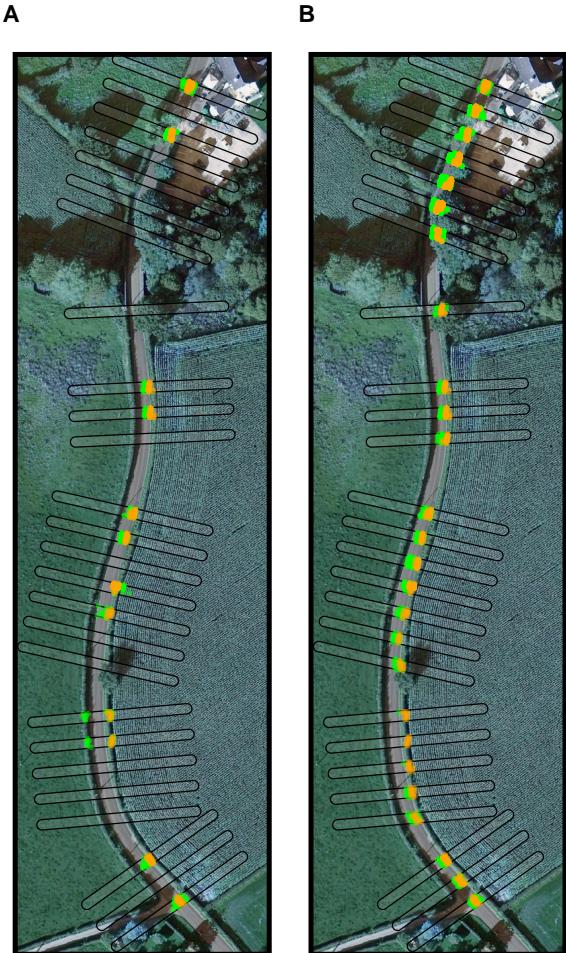


Fig. 4.1: LiDAR point clouds for one selected road section aggregated into $1m^2$ grids, (A) Base point cloud Z values, (B) Normalised Point cloud Z values for only last returns (lpz) (C) Normalised Point cloud Intensity values for last return, (D) Aerial Data combined to 1 band

4.3. Perpendicular Sampling

Covers 01

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



4.4. Linear Probability Model

Covers 02

test.

```
[1] 0.9529064
[1] 0.9514512
[1] 0.9589439
[1] 0.9591193
[1] 0.9580487
```

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

First model (maximal):

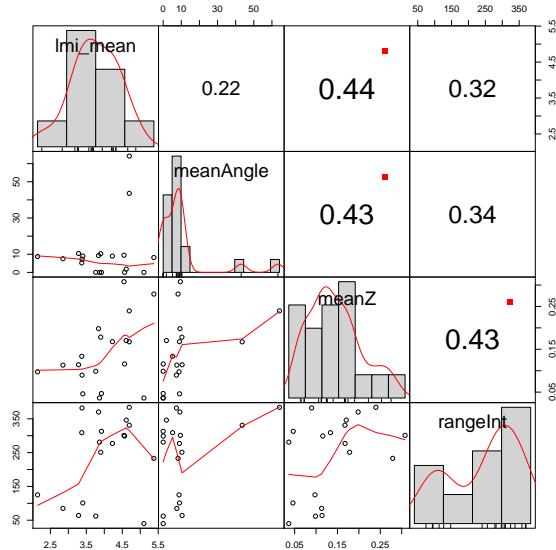
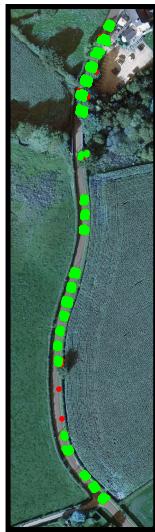
//for this section see 450 assess2, details on validation etc.

//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.

4.5. Noise Filtering

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



// 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.

4.6. Final Model Analysis

Very rough estimate of widths

4.7. Road Assessment

Table 4.2: Correlation Matrix

Variable	Width	Max Angle	Max Z	Int Range
Width		0.22	0.44	0.32
Max Angle	0.22		0.43	0.34
Max Z	0.44	0.43		0.43
Int Range	0.32	0.34	0.43	

*** Significant at the 0.001 level

** Significant at the 0.005 level

* Significant at the 0.01 level

4.8. Widths and Road Quality

4.9. Road 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 stwwpwnts etc)

5. Discussion

==== notes ==== (Häger 2005) 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.

Assessment strategy: assessing the results of an image processing work flow usually a textual description. but results can't be compared easily. Heipke1997 suggest a comparison between methods numerically, ratios derived from statistical test theory:

be identified. True positive (TP) – A phenomenon that is present within the input data and that has successfully been identified within the output data. False positive (FP) – A phenomenon that is not present

within the input data but that has falsely been found to be a phenomenon by the algorithm. Thus it is written to output data. False negative (FN) – A phenomenon that is present within the input data but that has not been identified by the algorithm and therefore has been omitted from the output data. Then we define Completeness by Completeness = $\frac{TP}{TP + FN}$ (2.1) This is obvious and can directly be taken from definition of TP and FN. Furthermore we define Correctness by Correctness = $\frac{TP}{TP + FP}$ (2.2) This is easy to see since we expect results not to contain features which are erroneous in a way that they contain features which are not present within input data. However, it is implied within this statement that we neglect those cases where features are present within the input data but are missing in the output. Finally Quality is defined by Quality = $\frac{TP}{TP + FP + FN}$ (2.3) This can be considered as a kind of global measure for overall assessment of obtained results. It comprises effects caused by both false positives and false negatives. Some but not all of the algorithms investigated in this paper use an assessment that is based on the above mentioned measures. Their results are reproduced where available. Otherwise, assessment is done by means of a textual description.

// 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 (2006), 'Speed Assessment Framework', <http://www2.dft.gov.uk/>.
- Department for Transport (2011), 'Strategic framework for road safety'. ZSCC: 0000006.
- 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].
- Dubes, R. & Ohanian, P. (1992), 'Performance evaluation for four classes of textural features', *Pattern Recognition* **25**, 819–833. ZSCC: 0000550.
- Elberink, S. O. & Maas, H.-G. (2000), 'THE USE OF ANISOTROPIC HEIGHT TEXTURE MEASURES FOR THE SEGMENTATION OF AIRBORNE LASER SCANNER DATA', p. 8.
- 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.
- Ferchichi, S. & Shengrui Wang (2005), Optimization of cluster coverage for road centre-line extraction in high resolution satellite images, in 'IEEE International Conference on Image Processing 2005', Vol. 2, pp. II–201.
- 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.
- 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.
- Hatger, C. (2005), 'Road extraction by use of airborne laser scanner data', p. 19. ZSCC: 0000001.
- Highways England (2019), 'Network management'. ZSCC: 0000000.
- Ihaka, R. (2011), 'Rnoweb: Literate Programming with and for R', p. 5. ZSCC: 0000000.
- Jaakkola, A., Hyppä, J., Hyppä, H. & Kukko, A. (2008), 'Retrieval Algorithms for Road Surface Modelling Using Laser-Based Mobile Mapping', *Sensors* **8**(9), 5238–5249. ZSCC: 0000162.
- Knuth, D. E. (1984), 'Literate programming', *The Computer Journal* **27**(2), 97–111. ZSCC: 0002155.
- Kraus, K. & Pfeifer, N. (1998), 'Determination of terrain models in wooded areas with airborne laser scanner data', *ISPRS Journal of Photogrammetry and Remote Sensing* **53**(4), 193–203. ZSCC: 0001405.
- 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.
- Mungall, I. (2005), 'Trend towards centralisation of hospital services.', 00000.

- Nicholl, J., West, J., Goodacre, S. & Turner, J. (2007), 'The relationship between distance to hospital and patient mortality in emergencies: An observational study', *Emergency Medicine Journal* **24**(9), 665–668. ZSCC: 00162.
- Noctor, I. (2004), 'Change to kph limit to cost €30m', <https://www.irishtimes.com/life-and-style/motors/change-to-kph-limit-to-cost-30m-1.1133469>. ZSCC: NoCitationData[s0].
- 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].
- Saeedi, S., Samadzadegan, F. & El-Sheimy, N. (2009), 'OBJECT EXTRACTION FROM LiDAR DATA USING AN ARTIFICIAL SWARM BEE COLONY CLUSTERING ALGORITHM', p. 6. ZSCC: 0000018.
- Smadja, L., Ninot, J. & Gavrilovic, T. (2010), 'ROAD EXTRACTION AND ENVIRONMENT INTERPRETATION FROM LiDAR SENSORS', p. 6. ZSCC: 0000040.
- Stradling, S. G., Great Britain & Department for Transport (2008), *Understanding Inappropriate High Speed: A Quantitative Analysis*, Dept. for Transport, London. ZSCC: NoCitationData[s0] OCLC: 277067852.
- Taylor, M. C., Baruya, A. & Kennedy, J. V. (2002), 'The relationship between speed and accidents on rural single-carriageway roads', p. 32. ZSCC: 0000125.
- Taylor, M. C., Lynam, D. A. & Baruya, A. (2000), 'The effects of drivers' speed on the frequency of road accidents', p. 56. ZSCC: 0000002.
- The University of Edinborough (2019), 'Aerial Digimap Educational User Licence', <https://digimap.edina.ac.uk/>. ZSCC: 0000000[s0].
- UK Government (2011), 'Rural Urban Classification', <https://www.gov.uk/government/collections/rural-urban-classification>. ZSCC: 0000347[s0].
- UK Government (2019a), 'Find open data - data.gov.uk', <https://data.gov.uk/>.
- UK Government (2019b), 'Speed limits', <https://www.gov.uk/speed-limits>. ZSCC: 0000773[s0].
- 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. (2000), 'SLOPE BASED FILTERING OF LASER ALTIMETRY DATA', p. 9. ZSCC: 0000833.
- Vosselman, G. (2009), 'Advanced Point Cloud Processing', p. 10. ZSCC: 0000074.
- Vosselman, G. & Zhou, L. (2009), 'Detection of curbstones in airborne laser scanning data', *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* **38** (Part 3/W8), 111–116. ZSCC: 0000030.
- Wan, Y., Shen, S., Song, Y. & Liu, S. (2007), A Road Extraction Approach Based on Fuzzy Logic for High-Resolution Multispectral Data, in 'Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007)', Vol. 2, pp. 203–207. ZSCC: 0000014.
- 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.
- Wu, K.-F., Donnell, E. T., Himes, S. C. & Sasidharan, L. (2013), 'Exploring the Association between Traffic Safety and Geometric Design Consistency Based on Vehicle Speed Metrics', *Journal of Transportation Engineering* **139**(7), 738–748. ZSCC: 0000027.
- 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.
- Yoon, J. & Crane, C. D. (2009), Evaluation of terrain using LADAR data in urban environment for autonomous vehicles and its application in the DARPA urban challenge, in '2009 ICCAS-SICE', pp. 641–646. ZSCC: 0000011.
- Zhang, Shu-Ching Chen, Whitman, D., Mei-Ling Shyu, Jianhua Yan & Chengcui Zhang (2003), 'A progressive morphological filter for removing nonground measurements from airborne LiDAR data', *IEEE Transactions on Geoscience and Remote Sensing* **41**(4), 872–882. ZSCC: NoCitationData[s0].
- 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: 8149

A. Code Appendix

A.1. Session Information

```
Machine:  
[1] "AMD Ryzen 5 2600 Six-Core Processor"  
Num cores:  
[1] 12  
Num threads:  
[1] 12  
RAM:  
33.7 GB
```

```
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  
  
attached base packages:  
[1] parallel stats      graphics grDevices utils      datasets methods  
[8] base  
  
other attached packages:  
[1] devtools_2.1.0          usethis_1.5.1  
[3] benchmarkme_1.0.0       doParallel_1.0.15  
[5] iterators_1.0.12        foreach_1.4.7  
[7] Rtsne_0.15              pbapply_1.4-1  
[9] rgdal_1.4-4             future_1.14.0  
[11] varhandle_2.0.3        forcats_0.4.0  
[13] stringr_1.4.0           dplyr_0.8.3  
[15] purrr_0.3.2            readr_1.3.1  
[17] tidyverse_1.2.1          tibble_2.1.3  
[19] tidyverse_1.2.1          lidR_2.1.2  
[21] raster_2.9-23           sp_1.3-1  
[23] scales_1.0.0            kableExtra_1.1.0  
[25] data.table_1.12.2        sf_0.7-7  
[27] ggpibr_0.2.2            magrittr_1.5  
[29] cowplot_1.0.0            viridis_0.5.1  
[31] viridisLite_0.3.0        broom_0.5.2  
[33] ggthemes_4.2.0           RStoolbox_0.2.6  
[35] PerformanceAnalytics_1.5.3 xts_0.11-2  
[37] zoo_1.8-6                Hmisc_4.2-0  
[39] ggplot2_3.2.0            Formula_1.2-3  
[41] survival_2.44-1.1        lattice_0.20-38  
[43] pacman_0.5.1             knitr_1.24  
[45] nvimcom_0.9-83           colorout_1.2-1
```

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$lmI_dum <- ifelse(x$lm > quantile(x$lm, .95), 1, 0)  
  x$lmI_dum <- ifelse(x$lm > quantile(x$lm, .95), 1, 0)  
  x$p_val <- sum(p)  
  
  return(x)  
}, error = function(e) NULL)
```