

Determining cycle mode choice for commuting using new fine-grained route analysis methods

Richard James Thomas

University of Leeds

August 2014

A dissertation submitted in partial fulfilment of the requirements of the degree of Master of Science in
Geographical Information Systems

Abstract

Previous research has shown that the likelihood of choosing to commute by bicycle can be very dependent on the availability of acceptable cycle routes of reasonable distance, avoiding hills and with a minimum of cycling in motorized traffic. Several papers have also highlighted this is an aspect that has been difficult to measure effectively and thus is often done using indirect methods such as summing quantity of cycle infrastructure within an area. Transport simulation models have historically focused on motorized traffic and thus often do not have enough detail at a small street level to be representative for modeling non-motorized traffic. Taking the UK city of Bristol as a basis for analysis, this study looks at a method of combining route cost functions (derived from third party cycle-specific routing software) for a large number of realistic commuting journeys. To overcome availability limitations on small area census commuting data, an additional method is developed to generate representative synthetic commuting origin-destination flow data at a fine spatial granularity using the census data that is available as constraints.

Acknowledgements

The following people supplied helpful information and advice for which I am grateful. Even where data supplied was beyond the scope of this dissertation, it gave valuable pointers for bounding the limits of the study and highlighting possible directions for further analysis.

- Simon Nuttall (and colleagues) at CycleStreets.net for both writing the routing engine and making the Journey Planner API available for this research,
- Andy Cope and Andy Whitty at Sustrans for supply of information, advice and GIS data for mapping cycle infrastructure,
- James White at West of England Partnership for supply of JLTP3 cycle data and pointers to other sources,
- Dr Caroline Bartle and Dr Kiron Chatterjee at University of West of England (UWE) for supply of information on Bristol cycling-related data sources and contacts,
- GIS.StackExchange.com website for Python shapefile vector processing help,
- Dr Andrew Evans at University of Leeds for helpful advice and an ever-calm manner as my dissertation supervisor,
- Dr Robin Lovelace at University of Leeds for data, advice and boundless enthusiasm on cycling research,
- Dr Paul Norman & Professor Walter Gilks at University of Leeds for statistics advice.

This work used public sector information licensed under the Open Government Licence v2.0 obtained via the NOMISWEB website (ONS, 2014a). GIS boundary data was obtained via UK Data Service Census Support facility (EDINA, 2014a); a service supported by ESRC and JISC. Digital boundary data and elevation data was supplied by Ordnance Survey. These data are Crown copyright and are reproduced with permission of the Office of Public Sector Information (OPSI).

Use was also made of base maps supplied by Google Maps and OpenStreetMap (with Stamen Toner Map Styling) and cycle infrastructure GIS data supplied by Sustrans.

1	Introduction.....	7
1.1	Aims and Objectives	8
1.2	Outline of the Document Structure.....	9
2	Context: Utility Cycling in the UK.....	10
2.1	Cycling in the UK	10
2.2	Cycling in Bristol.....	11
3	Literature Review	13
3.1	Introduction.....	13
3.2	Factors Influencing Cycle Mode Choice.....	13
3.2.1	General Journey Factors.....	13
3.2.2	Cycling Infrastructure	14
3.2.3	Location-based factors	16
3.2.4	Personal factors.....	17
3.3	Research methods	18
3.3.1	Survey Types	18
3.3.2	Data Synthesis and Modelling	19
3.3.3	Route Choice Analysis.....	20
3.3.4	Mode Choice Analysis.....	21
4	Data Sources.....	23
4.1	Relevant Data Sources Available	23
4.2	OpenStreetMap and CycleStreets.net	25
4.3	2011 Census of England and Wales	25
5	Methodology	27
5.1	Tools.....	27
5.2	Defining Area Boundaries and Distance Limits.....	27
5.3	Routing Cost Functions	29
5.4	Synthetic Flow Data Generation (GenSynthFlow)	31
5.5	Route-based Regression (MSOA level)	33
5.6	Area-based Regression (MSOA and OA level)	34
6	Results	35
6.1	Synthetic Flow Data Generation (GenSynthFlow).....	35
6.2	Route-based Regression (MSOA level)	37
6.3	Area-based Regression (MSOA level)	37

6.4	Area-based Regression (OA level) using Synthetic Data.....	39
7	Analysis and Discussion	42
7.1	Synthetic Flow Data Generation (GenSynthFlow)	42
7.2	Cycle Routing of Synthesized (OA level) Flows.....	42
7.3	Cost Functions Aggregated by Origin Area.....	44
7.3.1	Mean Euclidean Distance	44
7.3.2	Travel Time	45
7.3.3	Directness	45
7.3.4	Cycling in Traffic.....	46
7.3.5	Effort to Distance Ratio	47
7.4	Regression Analysis.....	49
8	Conclusions.....	50
9	References	52
	Appendix I: Glossary of Acronyms.....	56
	Appendix II: GenSynthFlow (Java software)	57
	Appendix III: Correlations between ‘Independent’ Variables	59

List of Tables

Table 1: Summary of factors affecting commute mode choice.....	13
Table 2: Selection of large cross-sectional quantitative bicycle mode choice studies since 2000	21
Table 3: Data Sources relevant to Bristol Cycle Commuting Analysis.....	23
Table 4: 2011 Census Variables Used	26
Table 5: Software tools used	27
Table 6: Routing Cost Function Definitions	30
Table 7: Synthetic Flow Data Generation Algorithm	32
Table 8: MSOA PWC Route-based Bivariate Regression against Cycling % of Commuters	37
Table 9: MSOA PWC Route-based Multivariate Regression (Best Fit) against Cycling % of Commuters	37
Table 10: MSOA level Area-based Bivariate Regression against Cycling % of Commutes < 20km	38
Table 11: MSOA Area-based Multivariate Regression (v1) against Cycling % of Commutes < 20km	38
Table 12: MSOA Area-based Multivariate Regression (Best Fit) against Cycling % of Commutes < 20km.....	38
Table 13: OA level Area-based Bivariate Regression against Cycling % of Commutes < 20km	39
Table 14: OA Area-based Multivariate Regression (MSOA comparison) against Cycling % of Commutes < 20km	39
Table 15: OA Area-based Multivariate Regression (Best Fit) against Cycling % of Commutes < 20km	40
Table 16: MSOA Route-based Correlations between Nominally Independent Variables	59
Table 17: Notable MSOA Area-based Correlations between Nominally Independent Variables.....	59
Table 18: Notable OA Area-based Correlations between Nominally Independent Variables	60

List of Figures

Figure 1: Great Britain Bicycle Usage (billion vehicle miles) (Source: Keep, 2013)	10
Figure 2: 2011 Census Bicycle Commuting Flows (greater than 7) between MSOAs around Bristol	12
Figure 3: Cyclist commute distances for Bristol and South Gloucestershire UAs (table DC7701EW)	28
Figure 4: Geographic boundaries used in the analysis	28
Figure 5: Origin-Destination Table Synthesis (and Validation) Flow Diagram (MSOA & OA level).....	31
Figure 6: Route-based Regression Flow Diagram (MSOA level only).....	33
Figure 7: Area-based Regression Flow Diagram (MSOA & OA level).....	34
Figure 8: Distance distribution of synthesized MSOA origin-destination data within the 4 Unitary Authorities	35
Figure 9: Distance distribution of synthesized and actual MSOA flow data within 3km buffer of BUA	35
Figure 10: MSOA level commuter flows larger than 100 within 3km buffer of Bristol BUA.....	36
Figure 11: Distance distribution of synthesized OA-WZ origin-destination data within the 4 Unitary Authorities	36
Figure 12: Actual (Census recorded) Cycle Commute Proportions by area (OA) for Bristol BUA.....	40
Figure 13: Predicted Cycle Commute Proportions by area (OA) for Bristol BUA.....	41
Figure 14: for Cycle Commute Proportion Model by area (OA) for Bristol BUA.....	41
Figure 15: Summed Cycle Routing for Commutes less than 20 km from Bristol BUA	43
Figure 16: Cycle Routing compared to Cycle Infrastructure in Bristol BUA	43
Figure 17: Mean Euclidean Distance by OA for Commutes less than 20 km from Bristol BUA	44
Figure 18: Mean Travel Time by OA for Commutes less than 20 km from Bristol BUA.....	45
Figure 19: Mean Directness by OA for Commutes less than 20 km from Bristol BUA.....	46
Figure 20: Mean Traffic Exposure by OA for Commutes less than 20 km from Bristol BUA	47
Figure 21: Mean Effort to Distance Ratio by OA for Commutes less than 20 km from Bristol BUA	48
Figure 22: Digital Terrain Map (DTM) for Bristol Area	48
Figure 23: UML Class Diagram for GenSynthFlow Java software	58

1 Introduction

Cycling to work in England and Wales is very much a minority pursuit - only 2.8% of the working population recorded cycling as their main commute mode in the 2011 census (ONS, 2014b, p1). So why should we analyse cycle commuting?

"As commuting is non-discretionary and fixed in time and place for most people, it contributes disproportionately to traffic congestion and environmental pollution. Commuting by bicycle can therefore make a greater contribution to reducing congestion than cycling for other purposes."
(Heinen et al, 2010, p60)

Cycling data in many surveys is often limited to commuting, though this is a good proxy for utility cycling (Parkin, 2004); the UK National Travel Survey (NTS) shows a significant correlation, with the ratio of commuting to general utility cycling being 1:0.77 averaged over 2002-2010 (Goodman, 2013, p8). Census figures also generally under-represent cycle commuting as many people cycle (often weather-dependently) only on certain days or times of the year (Schoner and Levinson, 2013, p8), or in conjunction with a train journey as their main mode.

Considering cycle *routing* specifically is important: the location of cycling infrastructure can be key to whether it is effective in encouraging cycle commuting. Although various analyses take into account general measures of infrastructure provision (for example Parkin et al, 2008; Schoner and Levinson, 2013), it has been recommended that specific commuting routes should be analysed as characteristics of routes have a strong influence on cycle commuting uptake (Schoner, 2013, p48). Knowledge of existing routes can inform policy on locating new cycle infrastructure, but it could also highlight areas where infrastructure is adequate but cycle uptake is poor - indicating a policy need to address other determinants.

Doing such an analysis now is timely: much 2011 census data relating to Workplace Zone (WZ) geographies and origin-destination data has just become available in recent months. Because the census was only taken a few years ago, cycle-routing based on today's infrastructure will also still be largely representative of routes at census time.

The city of Bristol in the southwest of England was chosen for this analysis. This was partly because of the author's first-hand knowledge of the city, but also because the recent release of the first draft of a new Bristol Cycling Strategy (Bristol City Council, 2014a) proposed a change in cycle policy to focus on effectiveness of the network for utility cycling (as well as leisure provision).

The scope of this paper is to examine two new methods which in combination allow fine-grained cycle-commuting route analysis to be done between Population Weighted Centroids (PWCs) of the lowest census levels of geography: between Output Areas (OA) origins and Workplace Zone (WZ) destinations. The first method is to use a (third party) cycle-routing engine to generate plausible commuting routes and 'cost

functions' that characterise those routes. These route results will be aggregated based on known commuting origins to give area-based measures which can then be used in more traditional area-based aggregate regression analyses of determinants of cycle mode choice. Unfortunately, even when OA-level commuting origin-destination census flow data does become available at the end of 2014, it will only be accessible within a 'secure' environment, thus restricting its use. The second new method is to synthesize OA-level commuting flows (based on other available census data) which should be representative enough for the intended routing regressions analyses.

In summary, the regression model being built will address the likelihood of whether for each area, the local workers are likely to commute by bicycle, given typical locations commuted to from that area. However, the key aim of this study is not to produce a definitive model, but to determine if these new methods can improve on routing-related aspects of cycle mode commuting analysis and to explore where routing aspects fit into the very broad and complicated picture of what determines whether people choose to commute by bicycle.

1.1 Aims and Objectives

This study has several specific objectives which can be summarised by outlining how the key method steps relate to them:

1. Route analysis of MSOA (Middle layer Super Output Area) level commuting flows: Test correlations between routing cost functions and cycle mode choice

At the 2011 census MSOA-level of aggregation (2000-6000 households), a complete list of all origin-destination commuting flows (by any method) within an area surrounding Bristol will be extracted from census data. For each of these flows, a third party cycle routing engine will generate representative cycle paths and associated cost functions relating to route directness, time, effort and separation from traffic. Correlation and linear regression will highlight how these cost functions relate to actual cycling proportions taken from the census.

2. Residential area analysis at MSOA level: Combine route information with aggregate census data to test effectiveness of cost functions within an area-based regression

For each residential MSOA and each routing cost type, costs from all flows originating from this area will be combined to produce area cost functions. In combination with some additional census variables these will be compared with area-based cycling proportions from the census using linear regression, with the sole aim at this point of just validating the cost functions.

3. Synthesis of commuting flow data at OA (Output Area)/WZ (Workplace Zone) level: Generate representative origin-destination pairs using census data constraints.

Across an area the size of an MSOA, there can be large variations in cycle route quality. Calculating route costs for more finely-grained OA-level (40-250 households) commuting flows should give more representative results. To overcome restrictions on availability and usage of 2011 OA-level commuting flow data, such data will be synthesized based on census constraints data in the form of commuting distance ranges and numbers of cyclists for each OA origin and (separately) for each WZ destination. By additionally considering distances between OA and WZ PWCs, it should be possible to synthesize a set of OA-WZ flows which would be representative of the actual flows. This will be initially attempted at MSOA-level for validation (as actual MSOA flows are known), then repeated at OA-WZ level to provide an input to objective 4. Additionally, plotting synthesized OA-WZ routes will give a picture of the potential cycling network demand across the city which can be compared to the location of existing cycling infrastructure to determine how well it might be suited to commuting.

4. Residential area analysis at OA-WZ level: Repeating and extending the area analysis at the smallest census area level to make full use of fine-grained routing.

Although routing cost functions generated for synthesized OA-WZ flow data cannot be validated against known flows, if converted to area-based cost functions (as in objective 2) then a similar regression can be used to validate them against (known) OA-level cycle commuter counts and to determine how much bearing each cost function has on the desired outcome. Some compensation for other factors will be made by the addition of other OA-aggregate variables identified in the literature as potentially important determinants.

1.2 Outline of the Document Structure

- **Chapter 2:** Context of cycle commuting within the UK
- **Chapter 3:** Literature review for on determinants of cycle mode commuting choice and associated research methods
- **Chapter 4:** Critique of data sources available and details of those used in this study.
- **Chapter 5:** Methodology for creating synthetic flow data, cycle-routing, generation of cost functions and regression.
- **Chapter 6:** Results (with supporting correlation tables in Appendix III)
- **Chapter 7:** Analysis of results
- **Chapter 8:** Conclusions

2 Context: Utility Cycling in the UK

2.1 Cycling in the UK

Since 1949, UK cycling rates have fallen dramatically (Figure 1). Wardman et al (2007) report that even recently (the period 1992-2002) there was a 20% drop in number of bicycle trips per person, with the cycle commuting share falling in successive censuses: 3.8% (1981), 3.0% (1991), 2.9% (2001). However, NTS 2008-2010 shows cycling as a main commuting mode only accounts for 31% of all cycling time (Goodman, 2013), with part-journey or occasional commuting accounting for a further 10% of cycling time. The distribution of cycling is not spread evenly, either geographically or demographically; 2011 census data reveals that age.sex profiles for cycle commuting in London are very distinctive, but have smaller variations elsewhere (ONS, 2014b, p13). Although average cycle commuting mode proportion was unchanged from 2001 to 2011, there was a decline in absolute cycle commuter counts in most local authorities, but also dramatic increases in certain cities: Brighton (+109%), Bristol (+94%), Manchester (+83%), Newcastle (+81%) and Sheffield (+80%). The current National Transport Model predicts that cycling will rise in 2015 then fall for many decades (DfT, 2011).

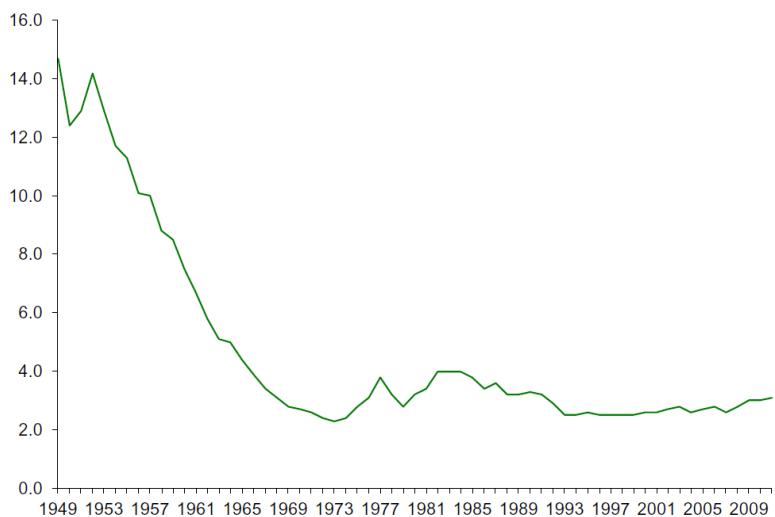


Figure 1: Great Britain Bicycle Usage (billion vehicle miles) (Source: Keep, 2013)

Building cycling-specific infrastructure has been seen by many as a large part of the solution in reversing the decline. In a UK-based correlation model, Parkin et al (2008, p105) predicted that creation of traffic-free radial routes in cities could increase cycling by between 17% and 101%, depending on hilliness. In a separate study, Wardman et al (2007, p14) forecast a best-case scenario (completely traffic-free commuting routes) of an increase of 55% (bringing cycling to a total of 9% of all commuters).

Two non-governmental organisations that have played a major part in promoting cycling are the CTC (national cycling charity) whose current "Space for Cycling" campaign focuses on lobbying local councils for high standards of cycle-friendly planning and design, and the sustainable transport charity Sustrans. Sustrans (which started as a local campaign group in Bristol) aims to promote walking, cycling and public

transport. It has been responsible in conjunction with local councils for the development of the National Cycle Network which was initiated in 1995 with a £42.5 million grant from the National Lottery.

In 1996 the UK government produced a National Cycling Strategy to quadruple the number of cycle trips by 2012 (Golbuff and Aldred, 2011, p15). However, such plans were already falling well short of targets by the time of a revised plan in 2000 (Gatersleben and Appleton, 2007, p303). Investment was made in 6 Cycling Demonstration Towns in 2005, plus a further 11 in 2008 and Bristol as the first Cycling City in 2008. (Golbuff and Aldred, 2011, p29). Government cycle promotion initiatives since then include the 2010 Active Travel Strategy for England and £1 billion funding for local sustainable transport initiatives (Goodman, 2013, p1). In 2013, the All Party Parliamentary Cycle Group (APPCG) produced the “Get Britain Cycling” report (Goodwin, 2013) bringing together an (internationally) wide range of statistics and making policy recommendations. The government’s response to this (the cross-departmental ‘Cycle Delivery Plan’) is due to be published later in 2014.

2.2 Cycling in Bristol

“More people in Bristol commute to work by bicycle or on foot than any other Local Authority in England and Wales” (Bristol City Council, 2014b)

The Bristol “Cycling City” investment programme (2008-2011) was a multi-faceted project: providing new cycle infrastructure, but also taking ‘softer’ approaches to engage with the population such as the “Bike It” schools programme and providing personalised travel planning (Greater Bristol Cycling City, 2011). Though effective in increasing numbers it was criticized by some with local authority representatives reportedly stating that it “targeted middle-class commuters” (Aldred, 2014). The visualisation of 2011 census data cycle commuting flows (Figure 1) gives a snapshot of cycling in Bristol at the end of this period.

Various cycling groups worked with Bristol and South Gloucestershire councils and consultants Arup to produce the “Greater Bristol Cycling Strategy 2011-2026” report (SAP, 2010) as an input to the follow-on Joint Local Transport Plan 3 (JLTP3) (West of England Partnership, 2011), but this was reportedly never adopted (Bristol Cycling Campaign, 2014a). JLTP3 is a joint plan between the 4 adjacent Unitary Authorities (UA) of Bristol (City), Bath and North-East Somerset, South Gloucestershire and North Somerset. £30 million of the funding for JLTP3 was awarded from the (national) Local Sustainable Transport Fund (LSTF). Working within the framework of the JLTP3, the ‘WEST’ Outcome Monitoring project of the University of the West of England (UWE) is not just evaluating the impact of LSTF spending, but also advancing the bringing together of a wide range of existing surveys and data sources and making plans for further data collection (Chatterjee et al, 2013a) – all of which could prove invaluable for future research by others.

The new Bristol Cycling Strategy (Bristol City Council, 2014a) was developed in collaboration with various campaign groups, with the proposed network viewable on the Bristol Cycling Manifesto website (Bristol Cycling Campaign, 2014b).

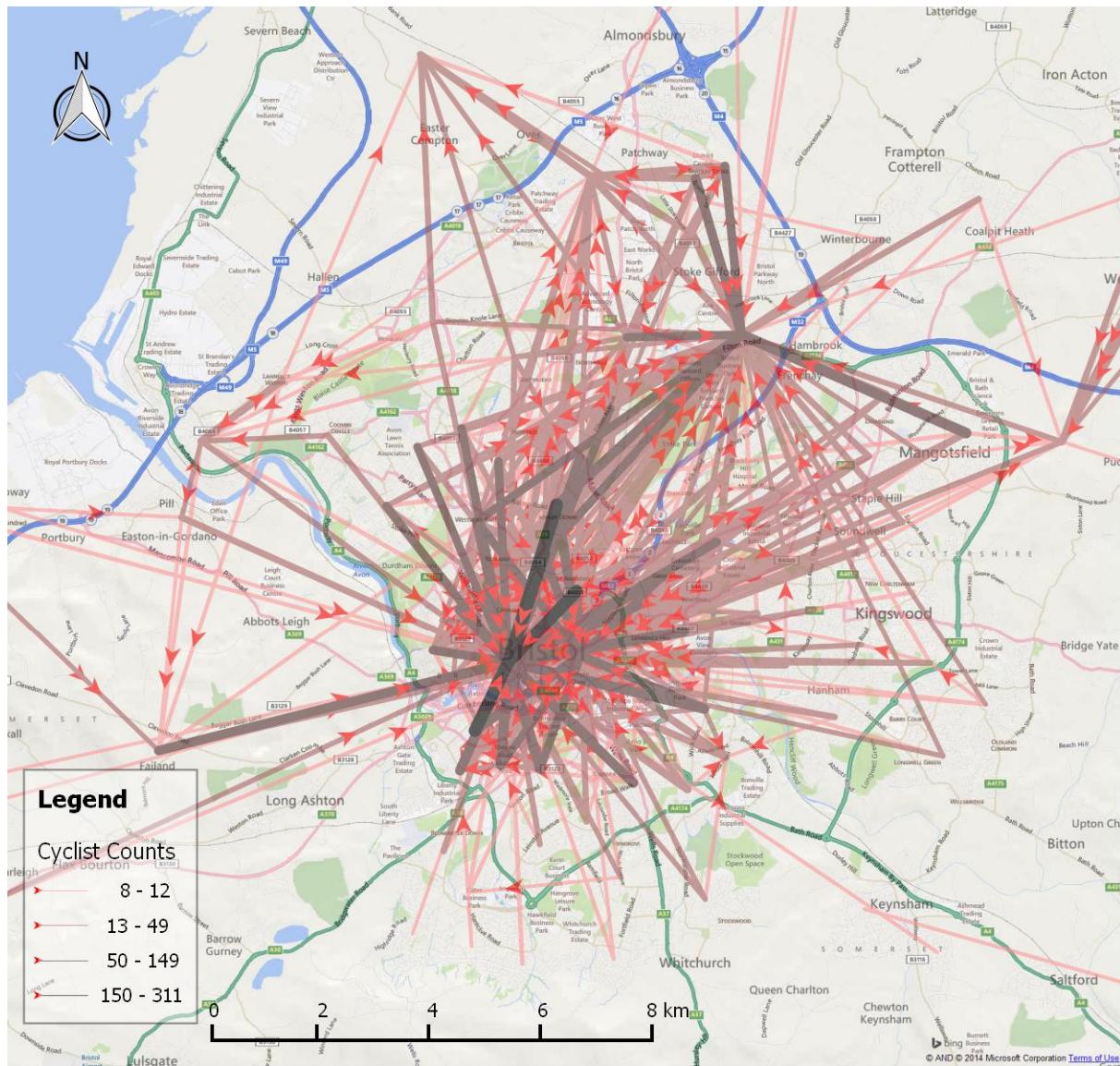


Figure 2: 2011 Census Bicycle Commuting Flows (greater than 7) between MSOAs around Bristol

3 Literature Review

3.1 Introduction

The promotion of cycling has in recent years become a significant policy focus in many countries and there is a wide body of literature (much of it quite recent) considering cycle mode choice for utility journeys. Two recent review papers give very thorough summaries of the many different factors:

- Heinen et al (2010): general focus on a wide range of factors affecting commuting by bicycle
- Pucher et al (2010): more specific focus on types of infrastructure and effects of cycling-promotion government policies

3.2 Factors Influencing Cycle Mode Choice

Table 1 summarizes the factors considered in this section, in the order they are discussed.

Table 1: Summary of factors affecting commute mode choice

General Journey Factors:	Location Factors:
Distance	Residential Density
Time (Duration)	Workplace facilities
Effort Required	Environment (Climate and Weather)
Safety (and Perceived Safety)	
Public Transport Alternatives	
Cycling Infrastructure:	Personal Factors:
Network Integration	Gender and Age (Demographics)
Complexity and continuity	Socio-Economic Factors
Cycle Lanes vs. Traffic-free Paths	Psychology and Attitudes
Segregation vs. Shared Space	

3.2.1 General Journey Factors

Journey Distance, Time: Unsurprisingly, the additional time often required to commute by bicycle rather than motorized means is a major factor that comes up in most studies – one of the biggest dissuading factors alongside safety (Parkin et al, 2007b) and typically the top factor in choosing a commuter cycle route (Stinson and Bhat, 2003). In Britain, only 1% of cycle journeys are greater than 5 miles (Gatersleben and Appleton, 2007, p303). However, where traffic congestion is particularly bad cycling can be quicker – being able to cycle past stationary traffic can be a strong motivating factor for some cyclists (Gatersleben and Appleton, 2007). To relate journey time to distance, it is important to consider achievable cycling speeds and effort required; in considering cycling infrastructure design, Parkin and Rotheram (2010) estimate 22kph (14mph) on the level as a realistic (85th percentile) speed.

Effort Required: This becomes less correlated with distance where there are large variations in hilliness and stop/start event locations such as traffic lights and busy junctions. Considering limits of a typical cyclist's power output and likely sustained effort Graham (1998) made a mathematically rigorous analysis of

variations of cycling effort required due to hills. Philips et al (2013) calculated “pedalling power” and consequent speed based on age, weight, height and fitness. The online CycleStreets.net cycle-routing engine calculates energy used based on equations for the fundamental physics and human physiology of cycling defined by Di Prampero et al (1979). However, though acknowledging tiredness as a significant detractor, Wardman et al (2007) were surprised to not detect hills as having significant impact in their survey even though it was in a relatively flat area. Parkin et al (2008) used a measure for the effect of hills based on the number of 1km squares in an area with gradients above an arbitrary steepness.

Safety (and Perceived Safety): This was typically the biggest issue for new cyclists (Cope et al, 2003, p16). Wardman et al (2007) highlight perceived danger as the principal deterrent. Vandenbulcke et al (2011) noted high traffic volumes as a major detractor alongside actual cycling accidents. However, where there are large numbers of cyclists their safety improves due to becoming more visible to motorized traffic (Pucher et al, 2010, pS121). Actual and perceived safety has been noted as often very different in many studies (Heinen et al, 2010). Care must be taken if cycle accident rates are used as a measure of risk as it can be proportional to the number of cyclists and also unreliable as many accidents go unreported (Parkin, 2004, p147). Perceived cyclist safety can affect motorist behaviour too – in experiments (Parkin and Meyers, 2010), drivers would pass closer to a cyclist wearing a helmet.

3.2.2 Cycling Infrastructure

Varying from (at its most basic) routes excluding certain traffic (such as shared bus lanes) to dedicated cycle-only paths, infrastructure has the potential to improve the appeal of a cycle commute. More detail will be considered here than in other sections because effects of infrastructure as a determinant of cycle mode choice are somewhat contentious in the literature and very dependent on infrastructure type.

3.2.2.1 Network Integration

An “If You Build Them, Commuters Will Use Them” hypothesis was proposed by Nelson and Allen (1997) based on the study of the effect of new cycle infrastructure in 2 US cities. However, this was rebuffed by Cleaveland and Douma (2009) who extended the study to 6 other US cities. They found that for new infrastructure to have an impact it had to be aligned with commuting routes, have good network connectivity and be accompanied by suitable publicity and promotion. Pucher et al (2010) also noted that individual infrastructure interventions had limited effect, but combined with a coordinated package of measures to promote cycling could have a much bigger impact than the sum of its parts. In Europe, the increase in impact of cycle infrastructure integrated into city-wide cycle networks has been shown in Freiburg, Germany (Buehler and Pucher, 2011). In a mathematical graph theory based analysis of the effect of network quality on cycle commuting across 75 US cities, Schoner and Levinson (2013) concluded that network density was the most important factor, followed by directness, then lack of fragmentation. However, where there are large increases in cycling uptake but only limited provision of cycling

infrastructure on a route, its impact will reduce as it approaches its “carrying capacity” (Lovelace, 2011, p2081).

Parkin et al (2007b, p7) notes that Dutch cycle design guidance (CROW, 1993) formed the basis for UK design recommendations. This emphasized the importance of a coherent and comprehensive network serving required origins and destination with the minimum of detours. The notion of route ‘directness’ or ‘circuitry’ (ratio of Euclidean distance to routed distance) was emphasized as an important factor by Philips et al (2013).

3.2.2.2 *Complexity and Continuity*

Quiet roads can form important elements of a cycle network, though in urban areas this can lead to complicated routes through residential streets. The appeal of a cycle network is reduced if it cannot be easily followed (particularly when a cyclist first tries to establish a commuting route) – this requires coherent mapping and signing (Cope et al, 2003) and the avoidance of excessive turn frequency (Broach et al, 2012). The ultimate benefit of using cycle infrastructure on a journey is not dependent on the number of bits of infrastructure that can be strung together, but on actual time spent using the infrastructure itself (Wardman et al, 2007).

3.2.2.3 *Cycle Lanes vs. Traffic-free Paths*

Although on-road cycle lanes (or shared lanes) encourage separation of much motorized traffic from bicycles, Parkin et al (2007a) noted that the perception of risk was only significantly reduced on traffic-free paths. Indeed, in experiments (Parkin et al, 2007b) found that interviewees shown roundabouts with marked cycle lanes often counter-intuitively perceived them as more dangerous than those without; he concluded that it was perceived that there was more need for cycle markings to cope with added danger. However, traffic-free routes were often disliked due to street furniture obstacles, poor surfaces, inconvenient routes and pedestrian conflict on shared-use paths (an issue even acknowledged by Sustrans (Cope et al, 2003)). Cleaveland and Douma (2009) found that off-street routes did not seem any better than on-street routes in increasing cycle commuting take-up. They reasoned this was primarily due to not being on commuting routes – often being hidden on old railway or river paths.

Schoner and Levinson (2013) noted that although in stated preference surveys cycle commuters said they would happily detour to use such routes, analysis shows that they are actually very sensitive to route length. In an analysis of 90 US cities particularly focused on the effects of bike paths and lanes on commuting (but controlling for a very wide range of factors), Buehler and Pucher (2012) found no significant difference in their effect on cycle commuting. Gatersleben and Appleton (2007, p310) even noted that cycle paths are sometimes avoided despite safety issues. In a study to monitor routes of 79 cyclists using GPS tracking devices, Yeboah and Alvanides (2013) noted that over half did prefer to cycle on the cycle path, though females expressed a stronger preference than males.

However, although traffic-free routes are often not ideal for commuting, they can play a major role in encouraging leisure and less time-pressured utility cycling. In 2003, two-thirds of cyclists surveyed on the UK National Cycling Network (NCN) were leisure cyclists (Cope et al, 2003, p14) and indeed the “primary aim” of the NCN had been to encourage new people to start cycling. Despite government surveys (limited to road routes) showing little change in cycling numbers for the period 1998-2004, broader survey work by Sustrans showed an increase of over 40% in (largely traffic-free) NCN use (Cope et al, 2007).

3.2.2.4 Segregation vs. Shared Space

In a (UK) analysis of urban traffic-free cycle paths Jones (2012) concluded that for an effective increase in utility cycling a combination of “segregated cycle facilities on major roads” and cycle promotion was needed. Segregation aims to combine the directness of following roads with guaranteed space and the risk reduction of traffic-free off-road routes. Although in the UK, this is typically just a white line to mark separation from pedestrians on a pavement, potential pedestrian conflict is reduced when there are physical barriers (such as kerbs or bollards) placed between the cycle lane and both pedestrians and cars. In a recent analysis of such “Protected Bike Lanes” in 5 cities in the US (Monsere et al, 2014), increases in cycling on these roads varied from +21% to +171%, with 10% of users switching from other modes of transport.

There has recently been much debate about the merits of an alternative “shared space” approach that removes most of the road markings and street furniture – “all street users moving and interacting in their use of space on the basis of informal social protocols and negotiation” (Hamilton-Baillie, 2008, p166). Although some of these schemes are quite radical in their proposals to transform places like busy junctions, they have been implemented in a less controversial way in many residential areas as ‘home zones’. To be successful “shared space” depends on the more vulnerable pedestrians (and cyclists) feeling confident to move around in the space and for drivers to have a suitably raised alertness to changes in priority (Kaparias et al, 2012). However, based on further study of one of the official UK shared space test sites in Ashford, Moody and Melia (2013) contested many of the claims for shared space over segregation, noting particular difficulties in high traffic flow situations.

3.2.3 Location-based factors

Residential Density: Higher housing density often correlates with increased cycle commuting - possibly due to lack of available car parking (though bike storage can be an issue for small properties) and reduced car speeds (Parkin et al, 2008). Such areas are often close to urban centres (Heinen et al, 2010, p62).

Workplace Facilities: Facilities such as showers and secure cycle parking (or lack of car parking) can strongly affect cycle commuting (Heinen et al, 2013; Parkin et al, 2007b), with social attitudes also being important. This is helped by government incentive schemes like the UK ‘cycle to work’ bike-funding scheme and

employer-based promotional programmes (Pucher et al, 2010, pS113). However, some have measured this as only having a minor impact on cycling (Wardman et al, 2007).

Environment: Non-cyclists trying cycle commuting in a UK trial (between February and April) reported their main issues as darkness and poor weather (Gatersleben and Appleton, 2007, p307). Although climate, weather and seasonal variations in temperature and daylight can have a statistically significant impact (Parkin et al, 2008; Dill and Carr, 2003) such factors can be considered invariant across the proposed area in this analysis.

Public Transport Alternatives: given suitable availability, cost and convenience these can be a direct competitor to cycling (Rietveld and Daniel, 2004, p535)

3.2.4 Personal factors

3.2.4.1 *Gender and Age (Demographics)*

Although the 2001 and 2011 censuses of England and Wales showed distinct peaks in cycling around the 30-34 age group in London, the variation was much reduced in the rest of the country (ONS, 2014b); in 2011 there was a slight dip in numbers in the 30-39 age group for those commuting less than 5km, with the peak being in the 45-49 age group. However, different age groups are sensitive to quite different factors (Cope et al, 2003) with young people most concerned about saving money and journey efficiency, whereas those age 35-44 concerned more about safety and personal fitness. For the 2001 census, Parkin et al (2007b) found higher numbers in the under-35 age group and considered this likely to be due to more central (urban) living and less likelihood to own a car. Generally the probability of cycle commuting reduces with older groups (Wardman et al, 2007). Heinen et al (2010, p69) noted that a large majority of studies found cycling was more popular with men than women. Gatersleben and Appleton (2007) concluded that this imbalance was largely an issue requiring culture change.

3.2.4.2 *Socio-Economic Factors*

Cycling propensity has been seen to reduce with increasing affluence (Schoner and Levinson, 2013) and with lower social groups (Badland and Schofield, 2006). However, this is not clear cut: in an analysis of UK national surveys, Goodman (2013) concluded that though historically cycling was less prevalent with increased affluence, by 2011 it was largely equal across groups and the trend suggests that in future it will increase with affluence. He noted that “higher social grade” increased likelihood of having cycled in the last week. Across a wide range of papers, Heinen et al (2010) concluded that the effect of income is ambiguous, though car ownership has a strong negative effect. However, Parkin (2004) found that although households with multiple car ownership were less likely to cycle commute, those with one car were actually more likely to cycle than those with none. Family or household responsibilities such as caring for children apply practical constraints on opportunities for cycling (Pooley et al, 2011).

3.2.4.3 Psychology and Attitudes

Different groups of people have different reasons for not cycling (Cope et al, 2003). This is often dependent on cycling experience and frequency (Gatersleben and Appleton, 2007); they proposed use of a “transactional model” to identify factors that cause changes between states – noting in their experiments how non-cycling commuters’ perceptions were changed by actually trying cycle commuting. The UK Department for Transport (DfT) guidance since 2008 attempts to do something similar, considering 5 classes of cyclist (Parkin and Rotheram, 2010). Lovelace (2011) noted that in the last 10 years that the DfT emphasis for promoting cycling has changed from just infrastructure building to more ‘soft’ approaches to improve people’s understanding and perceptions. Correlation between cycle commuting and location does not necessarily indicate an area encourages more people to cycle – there is often a process of self-selection where those already prone to cycling move into areas better for cycling (Schoner, 2013; Handy and Xing, 2011, p109).

The Lancaster University study “Understanding Walking and Cycling” (Pooley et al, 2011) outline many psychological and cultural considerations both at home and in the workplace that affect propensity to cycle, including whether cycling is considered the norm for an individual’s circumstance. Parkin et al (2007b, p9) emphasize the need for better modelling of choice mechanisms based on such factors – including “life stage”. Chatterjee et al (2013b) propose that for significant increases in cycling a big change in psychology and attitudes is required. Pucher et al (2010, pS113) review a wide range of papers on the ‘soft’ aspects of promoting cycling including working with schools and employers, offering training, increasing access to bicycles as well as various marketing and awareness programmes. In their later paper on policies to promote cycling, Pooley et al (2013) focus on what it would take to transform social attitudes, economic and spatial environment so that “walking or cycling for short trips in urban areas is perceived as the logical and normal means of travel and using the car is viewed as exceptional.”

3.3 Research methods

3.3.1 Survey Types

Preference surveys: these can be helpful in understanding reasons people choose to cycle and can be subdivided into ‘stated preference’ (SP) and ‘revealed preference’ (RP) surveys. RP survey responses are more valuable as they show actual choices made by individuals, but are more difficult and time consuming to carry out (Stinson and Bhat, 2003). RP surveys can often be very practical in nature, with some using GPS tracking to reveal cyclists’ actual behaviour (Yeboah and Alvanides, 2013; Dill and Gliebe, 2008). Cope et al (2007, p2) highlights the (SP) National Travel Survey (NTS) as a key national transport data source.

Quantitative measures: In a review of the literature, Philips et al (2013) emphasized the strong preference for quantitative measures in policy making both in the UK and internationally. Though there are many studies exploring the subtleties of certain aspects (such as perceptions of risk), these are often based on

relatively small surveys or are more qualitative in nature. Pucher et al (2010) looked at forty studies comparing effects of cycle lanes and cycle paths, noting the rarity of quantitative results.

Aggregate/Non-aggregate data: Data from surveys with very large sample sizes (such as censuses) are often only available in aggregate format. Total counts for various characteristics are known but very little cross-tabulation can be done. Wardman et al (2007) stated that though cycling models based on such aggregate data are common they are “not well suited to the analysis of cycling attributes in detail.” However, Parkin et al (2008) noted the fact that aggregate census data being geographically specific allowed additional incorporation of data such as hilliness and thus was “a significant advance over disaggregate modelling where the hilliness effect has not been modelled successfully (e.g., Wardman et al. 2007).” – but additionally highlighted the wide range of non-aggregate modelling techniques applied to cycle mode choice in the literature. One of the risks of non-aggregate surveys is in how they select participants; Schoner (2013, p10) criticized the conclusions of Wardman et al (2007) for not reasonably compensating for exclusion of those not contemplating cycling.

3.3.2 Data Synthesis and Modelling

Transport modelling: There are a wide range of modelling and simulation methodologies in use. The Transport Analysis Guidance (Webtag) website (Dft, 2014a) provides recommendations on many modelling techniques for evaluating transport project proposals. However, much traditional transport modelling has too coarse a spatial resolution for consideration of walking or cycling (Iacono et al, 2010, p134), with the zone size of (vehicular) flow models causing many cycling trips to not leave the origin zone. Consequently they proposed a better approach for evaluating walking and cycling would be to use GIS network analysis. To try to overcome such issues, Eash (1999) built a model of cycling trips at small area level which was effectively an early Agent Based Model (ABM), with random trip data generated based on probabilities.

Microsimulation: This is a method of creating synthetic populations from multiple data sets, with techniques such as simulated annealing, deterministic reweighting and Monte Carlo simulations being used to iterate the data for a better model fit. Harland et al (2012) outlined these techniques and considered the effects at different geographical scales. Lovelace et al (2014) in applying microsimulation to commuter behaviour noted that although microsimulation data is not real, its power lies in providing a good representation of small area variations. They also highlighted the batch-processing potential of new routing tools to better investigate the impact of the travel network.

Origin-Destination (OD) Synthesis: In order to predict road network load estimation, sophisticated OD synthesis has been used, based on factors such as measured traffic volumes, historic records and predicted commuting times (Sherali et al, 2004), which can be split into matrix-estimation and (gravity model-based) parameter-calibration methods. Ding et al (2007) outline a matrix-estimation method addressing some of the problems of the data being underspecified using Bayesian models and iterative procedures. Barbour

and Fricker (1994) outlined a much simpler OD synthesis algorithm (SHAPE-2) that was not dependent on having historic OD data, but (similarly to the constraints available in this study) used flow counts at origins and destinations and a gravity model for routing between them.

Simulated Origin-Destination Routing: In a methodology designed to identify suitable locations for new cycle infrastructure, Larsen et al (2013) used known OD data for other transport modes as a basis for “potential” cycle trips. Although the routing method was quite a crude shortest-distance one, the results when aggregated to 300m grid cells were considered informative enough to identify appropriate cycle transport corridors.

3.3.3 Route Choice Analysis

Compared to mode or destination choice, route choice is a complex problem with a much larger range of possibilities (Prato, 2009, p67) although for regular commuting this is reduced and made more predictable as a cyclist would have acquired a good overview of the network. This can be broken down from route level to ‘link’ level analysis (Stinson and Bhat, 2003), where each ‘link’ (path between intersections) is often assessed by qualitative weighting of various attributes. Ehrgott et al (2012) argued that information can be gained by keeping separate the cost functions used in choosing routes as they might apply differently to different types of cyclists.

Although most studies looking at detailed cycle routing focus on how cyclists choose routes rather than whether people choose to cycle or not, route cost functions based on actual routes that these studies found gave best fit in a model should also give a good indication of appropriate cost functions for mode choice. In an analysis of 2435 GPS-tracked utility cyclists in Zurich, Menghini et al (2010) found that routed distance dominated route choice, followed by the steepness of maximum gradients (average slope was found insignificant). Start/stop events due to traffic controlled junctions were found to be a big detractor. Traffic-free bike paths were the only other factor to make a contribution, but this was less significant than stops. In a smaller study of 164 GPS-tracked cyclists in Portland Oregon, Broach et al (2012) found directness was a major factor that also scaled linearly with distance. Stop/start events were also generally a detractor, but where cyclists had to make turns in the presence of heavy traffic, traffic-controlled junctions were a positive factor. For hills, proportions of routes at various slope gradients proved the most useful variables, with cyclists willing to cycle 70% further on the flat to avoid an incline of 2-4%. Although traffic-free paths were preferred, cycle lanes on arterial roads proved no more appealing than parallel quieter (typically residential) roads.

Various studies have been done using estimated routing based on commuters’ origins/destinations: Larsen and El-Geneidy (2011) used an online survey of 2917 cyclists to model how cyclists trade directness against cycle infrastructure. However, this used a fairly crude measure of shortest network paths to link origins/destinations to the cycle infrastructure specified as used by the respondents.

3.3.4 Mode Choice Analysis

There are many quantitative longitudinal studies typically considering the effect of changes in infrastructure on cycle commuter count: Pucher et al (2010, pS118) provide a comprehensive review of these since 1990. However as this study considers mode choice in absolute terms, recent major cross-sectional quantitative studies are of more interest (see Table 2) – in particular which (non-routing) variables are significant factors that could help in validating a model against mode choice outcomes.

Table 2: Selection of large cross-sectional quantitative bicycle mode choice studies since 2000

Reference	Locations	Methodology	Datasets	Notes
Dill and Carr, 2003	35 large US Cities, expanding an earlier study (Nelson and Allen, 1997)	Aggregate data regression	2000 Census	Level of bike infrastructure significantly correlates. Cars per household strong negative correlation.
Rietveld and Daniel, 2004	103 municipalities in Holland	Aggregate data regression	National statistics, plus Fietsersbond (Dutch Cyclists Union) collected data	Policy focused, but finds major factors: altitude, city size, age, ethnicity, car park costs, safety, travel time (few stops)
Wardman et al, 2007	Secondary data for Great Britain, plus primary data from cities of Leicester, Norwich, York and Hull	Non-aggregate data. Hierarchical “Joint RP-SP Multinomial Logit Mode Choice Model”	NTS 1985-1997 (23926 records, RP). New RP survey of 969 people. Two new SP surveys of 2115 & 3106 people.	Claims “most comprehensive and largest model”. +55% cycling increase possible with full segregation
Parkin et al, 2008	All of England and Wales	Aggregate data binary logit regression model at ward level. “has contributed to official government [Webtag (DfT. 2014a)] guidance on estimating changes in levels of cycling use”	2001 Census, Income Deprivation (ID), new perceived risks survey, quantity of cycle infrastructure, traffic estimates, road condition, Met Office weather records.	Model predicts 81%. “Saturation point” estimated at 43% of commuter trips. Major factors: hills, distance rain, temperature, population density.
Lawson et al, 2013	Ireland	Aggregate. multiple logistic regression (MLR)	Irish 2006 census	Gender, car ownership and journey distance strongest factors
Schoner, 2013	Twin Cities (Minneapolis, St Paul) in US	Non-aggregate data. Negative Binomial regression	New postal survey of 1303 people.	Recommended improvement is to analyze actual commuting routes
Heinen et al, 2013	4 municipalities in Netherlands	Non-aggregate data. 2 binary logit regression models	New internet survey of 4299 people	Big focus on facilities at work

In the conclusions to some of these analyses, methodology recommendations were made for better route and infrastructure analysis: Parkin et al (2008) could not detect significant (expected) perceived risk effects associated with variations in cycle infrastructure, but noted they only had infrastructure mapping for 24% of the population. Schoner and Levinson (2013) found that in their model network analysis alone explained 50% of the outcome and recommended analysis of actual commuting routes and population adjacent to networks.

4 Data Sources

4.1 Relevant Data Sources Available

There are a large number of data sources available relevant to the analysis of cycle mode choice—Table 3 outlines those that were considered for Bristol; much survey data was either limited in detail to Local Authority level or had sample sizes too small to guarantee reliability at smaller geographies.

Traffic counts: It was hoped to use traffic count data for two purposes: motorized traffic for giving a measure of busyness on routes and cycle traffic data as a means of validating the final cycle commuting model. Unfortunately, motorized traffic counts are only made in a limited number of locations (mainly on major roads) which a cycle routing engine would generally try to avoid. In order for cycling route cost functions to be developed that would rate volume of motor traffic along the route, it would be quite complicated to make use of this sparse dataset. However, if access to route data from a transport network model calibrated by such traffic counts was available this could be invaluable. For validating with cycle commuting validation, the data is problematic in that it is for complete days (rather than just commuting times), so will include much non-commuting cycle traffic. In the JLTP3 data, different locations were counted on different days and at different times of the year, so although they might be good indicators of changes in traffic flow, they are less suited for comparing with each other given that cycle mode choice can be quite temperature/weather dependent. For this reason, it was decided to use only 2011 census data for cycle counts (see section 4.3).

Cycle Infrastructure Mapping: Although OpenStreetMap data was chosen for routing analysis (see next section), the Sustrans data proved invaluable for visualisation.

Characterizing Population: LSOA-level MOSAIC and Indices of Deprivation data could be mapped down to the intended final OA-level of analysis to provide useful income data missing from the census. However, it was decided that as the main focus of the study was on the contribution of routing measures, for simplicity population data would be limited to that available directly at OA-level from the census.

Table 3: Data Sources relevant to Bristol Cycle Commuting Analysis

Dataset (source)	Relevant Data	Comments
2001 Census of England and Wales (ONS, 2014a)	Aggregate population data and commuting flow data	Unrivalled sample size, but some problems with errors introduced to prevent disclosure of individuals.
2011 Census of England and Wales (ONS, 2014a)	Aggregate population data and commuting flow data	Some data only just becoming available.
Experian MOSAIC (UK Data Service, 2014)	Demographic data at lowest (LSOA) level	Available for periods 2004-2005 and 2008-2011.
Indices of Deprivation (ID) 2010 (ONS, 2014e)	Income Deprivation Domain at lowest (LSOA) level	From metadata: “some of the indices used to produce the index are

		derived from the 2001 Census and in other places it has not been possible to update the indicators since ID 2007”.
National Travel Survey (NTS) (UK Government, 2014)	How, why, when and where people travel	Annual: 20,000 people in 8,000 households. Acknowledged volatility in cycling statistics due to small numbers.
Active People Survey (DfT, 2014b)	Measures of utility cycling by frequency (e.g. 3 times a week)	National telephone survey of over 160,000 but only as little as 500 persons per local authority
Annual Population Survey (ONS, 2014b)	Information about commuter flows	(Annual): 300,000 people. Data aggregated to local authority level.
Bristol Quality of Life Survey (Bristol City Council, 2014c)	Non-aggregate cycling data tied in with socio-demographics	(Annual): 3500-5000 people. Lowest level of detail for cycling statistics is ward level.
DfT Traffic Counts (DfT, 2014c)	Motorized (and some cycle) traffic counts on major roads (101 in Bristol) and a small number on minor roads	Limited sampling locations, with none on traffic-free National Cycling Network (Cope et al, 2007).
JLTP3 Cycle Matrix (West of England Partnership)	Cycle Count Data across the 4 Unitary Authorities surrounding Bristol. Includes counts on many of the major traffic-free cycle paths.	(Available by permission only) Large number of cycle traffic counts over several years, but in limited number of locations.
Boundary Data (ONS, 2014d; EDINA, 2014a)	Census boundaries and Population Weighted Centroids (PWC), Built Up Areas	(EDINA limited to academic access) Formed the basis for all spatial data calculations in the analysis.
Ordnance Survey Data (EDINA, 2014b)	OS Terrain 50 DTM (Digital Terrain Map)	(EDINA limited to academic access) Useful as a second source for assessing hilliness of routes.
OpenStreetMap (OpenStreetMap, 2014)	Location of road and cycle network with details of cycle-specific infrastructure	Crowd-sourced, but remarkably detailed, though inconsistent in classification in places.
Sustrans cycle network GIS database	Detailed layout and thorough classification of different types of cycle infrastructure	(Available by permission only) Appears more accurate and consistent than OpenStreetMap.
Ordnance Survey “Urban Paths” GIS data	Routes not available to motor traffic but available to bicycles	(Available by permission only) Generally much less comprehensive than OpenStreetMap or Sustrans data.

4.2 OpenStreetMap and CycleStreets.net

The crowd-sourced OpenStreetMap project provides extensive details of cycle routes and existing cycle infrastructure types even if it appears not quite as comprehensive, consistent or up to date as Sustrans internal data. However, the data is publicly accessible and even downloadable in its entirety for offline processing. Several publicly available routing engines can be applied to OpenStreetMap data (pgRouting, Routino, Osmar, Osmosis, Graphhopper). CycleStreets.net was chosen as a source for this study as it is now a reasonably mature product, comes recommended by Sustrans and its “Journey Planner” API (CycleStreets.net, 2014) provides additional data that can form the basis (with minor modifications) for suitable route cost functions.

4.3 2011 Census of England and Wales

This provides the basis for most of the numerical analysis within this study and quantifies the key outcome variable: cycling as a proportion of commuter’s main travel method. Although (as discussed in the literature review) this is a good proxy for both cycle commuting and utility cycling in general, its limitations should be considered. A big shortcoming is that it does not include any aspect of frequency of cycle commute or partial commute (such as to a train station). It also does not take account of seasonal variations – the census is recorded in the month of March which can be cold, dark and wet. Thus it would be expected that the census would underreport the average proportion of cycle commuting trips.

A key problem with aggregate data is that characteristics that might define minority groups (such as cycle commuters) may be hidden by the dominance of others. However, as there is a tendency for similar people to group together (Tobler, 1970, p236), this is not as significant problem as if a random sample of population was made. Selecting data for the smallest available census geographic areas (OAs) also reduces the aggregation impact, but potentially at a cost of errors intentionally introduced to prevent individual disclosure, particularly affecting small counts. In the 2001 census, the procedure of Small Cell Adjustment (SCA) randomly changed all counts of 1 or 2 to either 0 or 3. Additionally random record swapping was applied to move individual’s details between areas. Fortunately, the processing applied to the 2011 census data is much less damaging and was limited to “smart” record swapping - only done where small counts risk identifying individuals. Additionally, record swapping was done to as nearby geographies as possible (ONS, 2014c). Because SCA was not used in 2011 data, there is no increase in errors associated with OA data removing one of the previous benefits of using larger areas.

Thus although OA-level flow data was only available for the 2001 census, all data for this study was taken from the 2011 census to avoid those errors as well as to provide more timely data given the large changes in cycle infrastructure and major employers in the Bristol area since 2001. Additionally, the 2011 census introduced the very useful new ‘Workplace Zone’ (WZ) geography which is scaled to employment rather than residential populations and thus more numerous at commuting ‘destinations’. Table 4 outlines the tables and variables used from the 2011 census.

Table 4: 2011 Census Variables Used

Table	Title	Geography	Notes
WU03EW	Location of usual residence and place of work by method of travel to work	MSOA	Origin-Destination Flow Data
QS701EW	Method of travel to work, Usual Resident	MSOA, OA	Only OA-level table that can identify cyclists. Defines commute origins
DC7701EW	Method of travel to work by distance travelled to work	LA	Indicates distribution of cyclists by commute distance (most detailed geography available)
QS702EW	Distance travelled to work	MSOA, OA	Provides cycle commute distances of interest in fixed intervals (0-2km, 2-5km, 5-10km, 10-20km)
WP702EW	Distance travelled to work (Workplace population)	MSOA, WZ	Equivalent to QS702EW, but for destinations rather than origins
QS102EW	Population Density	MSOA, OA	
LC6107EW	Economic Activity by sex by age	OA	Can extract age/sex for just those in employment
LC4609EW	Car or van availability by economic activity	OA	Number of cars per household
LC6201EW	Economic activity by ethnic group by age	OA	Can extract ethnicity for just those in employment

5 Methodology

5.1 Tools

Table 5 outlines the software tools used for this analysis. As the focus of this study is on demonstrating the potential contribution of new methods rather than producing the perfect model for cycle mode choice, analysis has been limited to simple linear regression. However, for a more optimal model a range of more sophisticated statistical methods should be considered as indicated by the varied selection in the cross-sectional studies outlined earlier in Table 2.

Table 5: Software tools used

Software Tool	How used
GenSynthFlow (Java application)	Newly authored software specifically for this project to synthesize origin-destination data
IBM SPSS	Linear regression and bivariate correlation
Python	Newly authored scripts to handle API calls, cost function generation from JSON data, histogram generation / binning data, flow count vector summation, shapefile generation, general data manipulation
QGIS	Spatial query selection, database ‘join’ functions as well as general data manipulation and map generation / data visualisation.
ArcGIS	Early exploratory work

5.2 Defining Area Boundaries and Distance Limits

The study effectively aims to address the question:

Of the people within Bristol BUA travelling within a 20km Euclidean distance, can the new cost functions give:

- (a) a good indication of likelihood to commute by bicycle,*
- (b) better information than existing data such as census distance to work tables (QS702) ?*

The range limit of 20km was chosen as very few cyclists commute further than this by means of bicycle alone (see Figure 3). The anomaly shown of a rapid increase in cycling above 60km is considered by the ONS (2014b) to be due to people cycling from second homes, though it could also be people who combine cycling with train travel. By assuming that total cyclist counts are limited to 20km this anomaly and the “Other” count will be factored into the analysis which should give a truer representation of cycling numbers. Also, by limiting total commuter count to the 20km range the statistical problems of handling small numbers for cycling proportions will be reduced.

Methodology implications of this 20km limit are:

- Proportion measures of census “distance to work” counts for ranges (0-2, 2-5, 5-10, 10-20km) should use a denominator of the 0-20km total, not total commutes.

- Only origin-destination flows less than 20km should be used.
- Assumption that cyclist count totals equal totals for less than 20km.
- Assumption that census characteristics of people commuting less than 20km are the same as for all commuters.

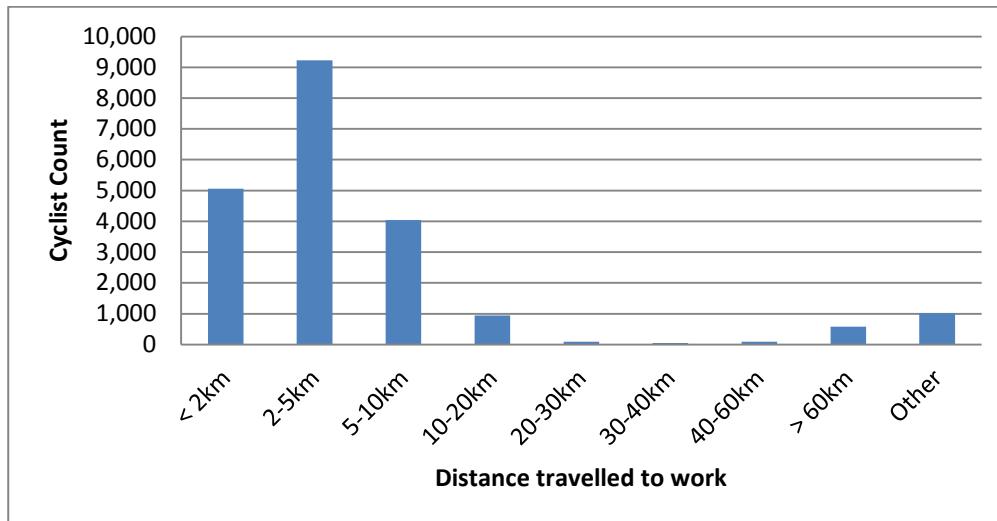


Figure 3: Cyclist commute distances for Bristol and South Gloucestershire UAs (table DC7701EW)

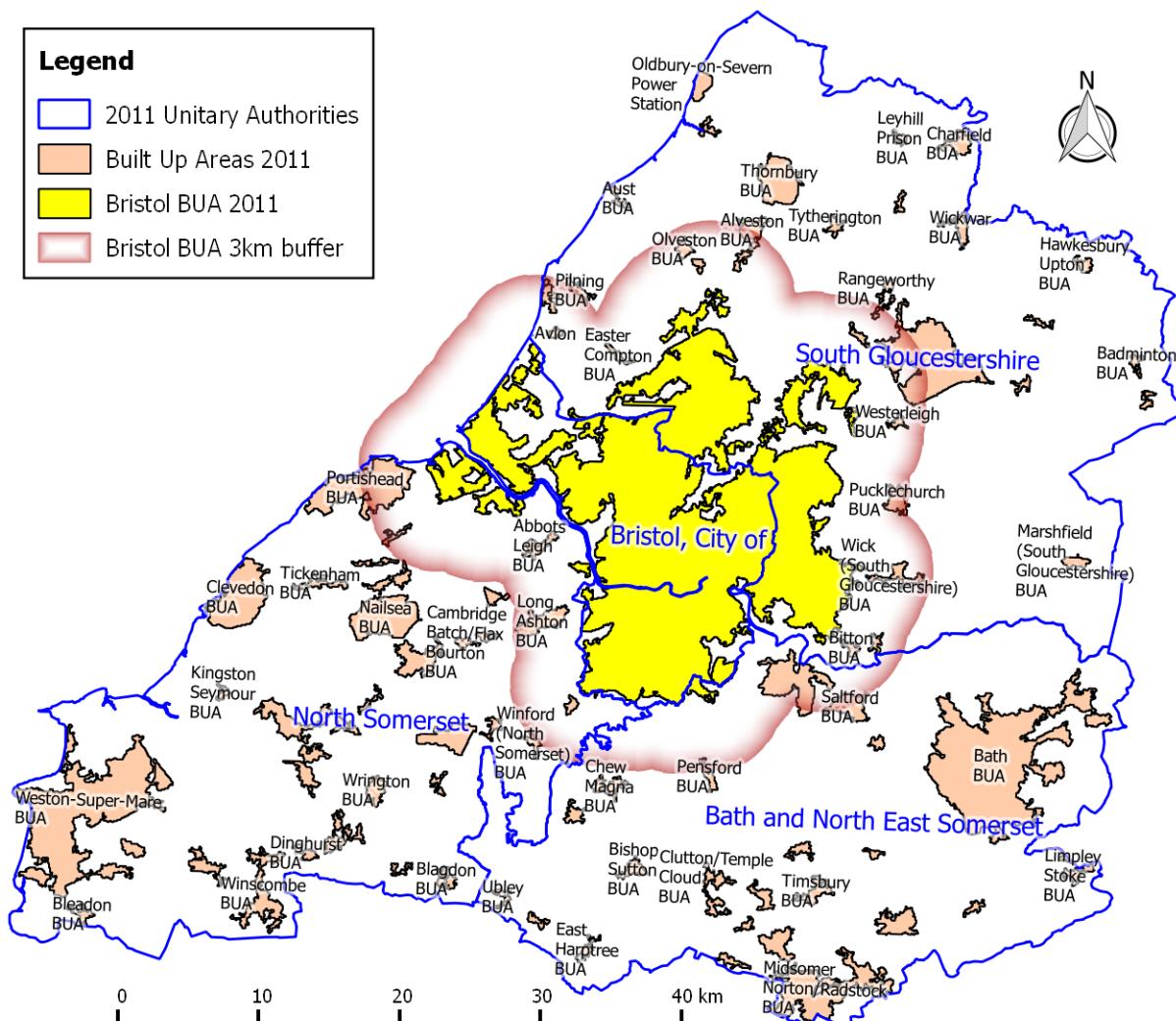


Figure 4: Geographic boundaries used in the analysis

Figure 4 illustrates various boundaries used within the analysis. Residential area classification and regressions were limited to areas falling within Bristol “Built Up Area” (BUA) as defined by ONS (2014d). The selection of census areas (MSOA, OA, WZ) considered to lie within the BUA and the 3km buffer around the BUA was made by considering whether their Population Weighted Centroids (PWC) fell within the boundary. This gives 73 MSOAs and 1951 OAs within Bristol BUA, with 86 MSOAs and 736 WZs within the 3km buffer.

Origin-destination data synthesis was performed right across the 4 Unitary Authorities (UA) shown to avoid geographically overly-constraining the algorithm. Although distances of up to 20km were considered for cycle commuting, cycle destinations sent to the routing engine were limited to within the 3km buffer to limit the number of routes to a more manageable number whilst analysing the area of most interest. However, calculations of mean and median Euclidean commute distance measures are not dependent on routing and so were calculated for all origin-destination pairs from Bristol BUA to within the 4 UA. This includes all the major population centres within 20km of any part of Bristol BUA (notably Bath and Weston-Super-Mare).

5.3 Routing Cost Functions

As discussed in section 4.2, the OpenStreetMap-based CycleStreets.net routing engine was chosen to generate routes and associated data for provided origin-destination commuting flows via the “Journey Planner” API (CycleStreets.net, 2014). This returned detailed route coordinates (in WGS1984 latitude/longitude format) along with a multitude of descriptive data, packaged into a JSON format. Python scripts were written to automate API calls and post-process results to produce cost functions – an approach alluded to by Lovelace et al (2014, p290). Additionally, coordinate files suitable for direct route plotting in ArcGIS or QGIS were generated.

Given the number of OAs within the BUA, the number of commute routes from each and that return routes from WZs will double routing required (unlike with MSOAs), it was decided to limit routing API requests to the four most popular commutes (by any means) for each origin area. Limiting routes chosen to only those with destinations within a 3km buffer around the BUA should avoid excluding popular cycling routes on the fringes of the BUA but ensure that cost functions largely reflect characteristics of the BUA. A minimum route distance of 1 km was used to exclude routes more likely to be walked.

Table 6 details the cost functions and how they were calculated. A key aim in cost function design was avoiding collinearity where possible and in particular avoiding functions that would correlate directly with distance as these functions are only being evaluated within the 3km buffer. Note that separate non-routed distance functions derived directly from origin-destination tables were also used (discussed later). A direct measure of ‘hilliness’ could have been generated from the step-by-step elevation data returned by the API.

However, it was hoped that this would be more usefully reflected in the ‘Calories consumed’ effort estimation which additionally takes into account likely stops on routes.

Table 6: Routing Cost Function Definitions

Cost Function	Calculation	Notes
Directness (%)	$(\text{Crow-fly Distance in metres}) / (\text{Routed Distance in metres}) * 100$	Also known as ‘circuitry’, this indicates how out of the way a cyclist reasonably needs to travel to avoid the busiest roads or take advantage of nearby cycle infrastructure (the CycleStreets ‘balanced route’ option was selected).
Traffic (%)	$100 - (\text{“Quietness” measure})$	An indication of how much cycling in traffic would need to be done; busy roads would be 100%, with cycle lanes or quiet residential streets scoring lower and traffic-free cycle paths scoring 0%.
EffortRatio	$(\text{Estimated Calories Consumed}) / (\text{Routed Distance in metres}) * 50$	An indication of how hard cycling on the route is, arbitrarily scaled to give values close to 1.0. Refer to section 3.2.1 in the literature review for the derivation of Calories.
SpeedKmh	$(\text{Routed Distance in kilometres}) / (\text{time in hours})$	Speed is clearly dependent on individual cyclists, but using default settings this should give a good indication of relative route speeds.
TimeMin	$(\text{Journey Duration in minutes})$	Note that this is highly correlated with distance and only useful for individual route analysis (rather than averaging) as routing is limited to the 3km buffer - not all commutes under 20km are being considered.

5.4 Synthetic Flow Data Generation (GenSynthFlow)

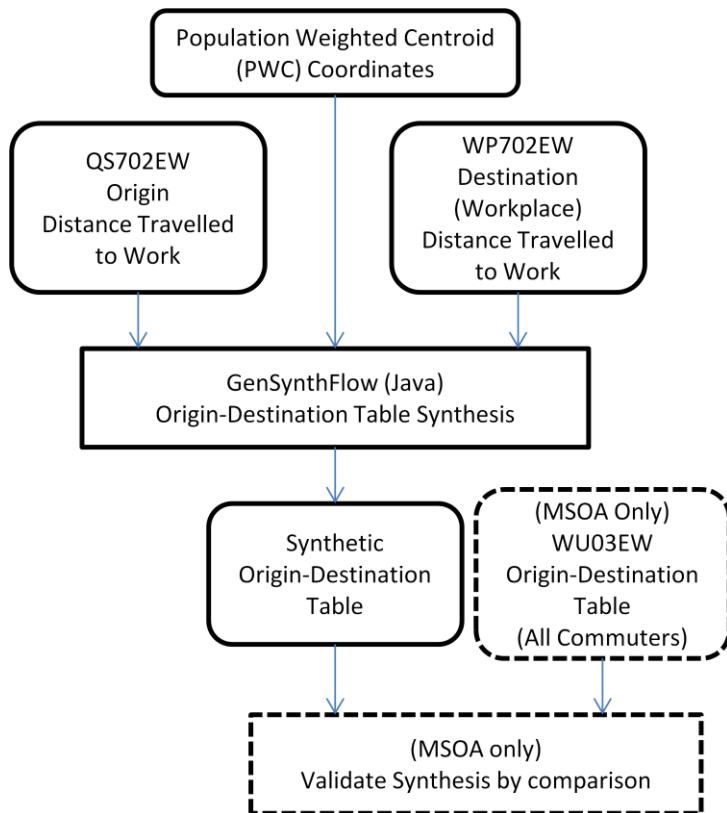


Figure 5: Origin-Destination Table Synthesis (and Validation) Flow Diagram (MSOA & OA level)

In order to avoid the coarse granularity transport modelling problems outlined in section 3.3.2 of the literature review, it was decided to model origin/destination locations at OA/WZ PWC resolution.

Fortunately because it is not required to model dynamic network loading in response to traffic counts, a much simpler origin-destination synthesis than most of those discussed could be used, though the SHAPE-2 gravity model of Barbour and Fricker (1994) formed a good basis. Although the characteristics to be matched were entirely spatial (location and distance) rather than the (typically) socio-economic ones used in microsimulation, some of the ideas of allowing randomised iteration not always restricted to lower cost solutions were adopted from that field.

Before describing the steps in the algorithm used, some of the assumptions and limitations will be addressed. Ideally, an origin-destination table for just cyclists rather than all commuters would be synthesized. However, this is not possible because below Local Authority (LA) level, census data is not available for cyclists combined with distance to work categories. The input constraints data is shown in the synthesis flow diagram (Figure 5) – see earlier Table 4 for variable descriptions. Because commuting would occur across the boundaries of any area modelled (unless all of England and Wales was modelled), it would not be expected to get a perfect fit to all the commuter counts. However, by modelling a larger area (4 complete UAs, covering most of the 20km range required) than that of interest (Bristol BUA), these errors should be minimized in the final analysis.

Note that some intra-MSOA flows (or OA flows to extremely near WZs) were generated; this is seen in real flow data and is not a problem, though such routes would be excluded from routing given the 1km minimum distance already discussed. In line with the census constraints data, synthesis was done for one ‘distance to work’ range at a time (e.g. 2-5km). Given that commuters in an area are not all located at the PWC, it might seem better to allow a tolerance beyond that range (e.g. 1-6km). However, this would result in unwanted boosts in commuter counts in the overlapping regions so it was avoided. The gravity model used assumes that the number of people travelling to a workplace will reduce with increased distance; how accurately this is modelled will become apparent in the results when origin-destination synthesized data for adjacent distance intervals (e.g. 0-2km, 2-5km) are combined.

Table 7 shows the algorithm for processing a single distance interval (actual Java software implementation details given in Appendix II).

Table 7: Synthetic Flow Data Generation Algorithm

Algorithm Step	Description
Create Synthetic Population	Population sized to equal totals of all workplace destination counts from table WP702EW. Each commuter can thus be assigned a destination.
Calculate ‘candidate’ origins	For each destination, a list of all “candidate origins” within current ‘distance to work’ interval range is calculated.
Assign origin lists to each commuter	Each commuter assigned a (newly randomly shuffled) list of “candidate origins” associated with their destination.
Assign initial origin to each commuter	Each commuter assigned first origin from its list of candidates.
Count current commuters per origin	Sum origin counts across initial population.
Loop for each commuter (at last commuter wrap to first commuter):	
Calculate origin count ‘proportional’ mismatch	Find difference in number of commuters assigned to commuter’s current origin with required number (specified in residential origin table QS702EW) as a proportion of required number.
Try changing commuter’s origin	If next origin in commuter’s candidate list gives a smaller mismatch or reduces distance of commute without increasing mismatch beyond a certain threshold, then switch origins and update counts.
Reduce mismatch threshold	Threshold starts off quite large (favouring the distance gravity cost function), but progressively reduces to zero (favouring mismatch reduction).
(Repeat loop until no commuters are moved over a full sweep of all their candidate origins)	

Running this synthesis for MSOA data allowed for validation of this method against known origin-destination data (dashed boxes in Figure 5). Synthesis was then repeated using OA origins and WZ destinations to generate the data required for small area routing. To test sensitivity to initial population, a loop was created to run the whole algorithm many times, with the final summed mismatches after each run compared.

5.5 Route-based Regression (MSOA level)

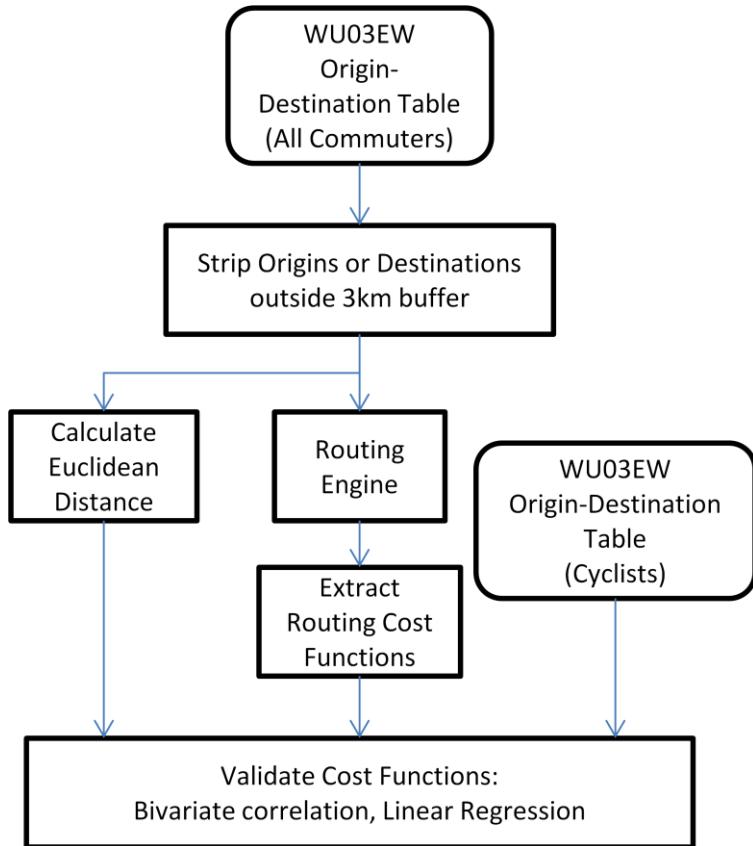


Figure 6: Route-based Regression Flow Diagram (MSOA level only)

As shown in Figure 6, validation of cost functions was attempted by performing bivariate then multivariate linear regression on a route-by-route basis against an outcome variable of known cyclist flows. This was done by running all (6887) non-zero commuter flows within the 3km buffer through the routing engine in both directions. Prior to the multivariate regression, cost functions were tested for collinearity and general usefulness using Pearson correlation. Multivariate regression was done by selecting variables that as well as increasing the Adjusted R² fit measure also did not have significant collinearity and seemed like plausible relationships. Although this method was intended to provide some validation for the cost functions, there are a number of issues that could cause problems:

- Routing from MSOA PWCs is quite a coarse spatial resolution which might produce unrepresentative results.
- Small census flow counts for cyclists may have had errors intentionally introduced to prevent disclosure which could ill-condition the regression.
- Because the total commuters on each flow varied from 1300 down to just 1, for very small flows a small error would cause a large swing in cyclist proportion. For this reason 'case weighting' by number of commuters was applied within SPSS to reduce the effect, even though this can give unrepresentative reports of significance (as they are not truly independent measures).
- No compensation was made for other determinants that might have a strong influence on the outcome variable. (This will be addressed in the next step).

5.6 Area-based Regression (MSOA and OA level)

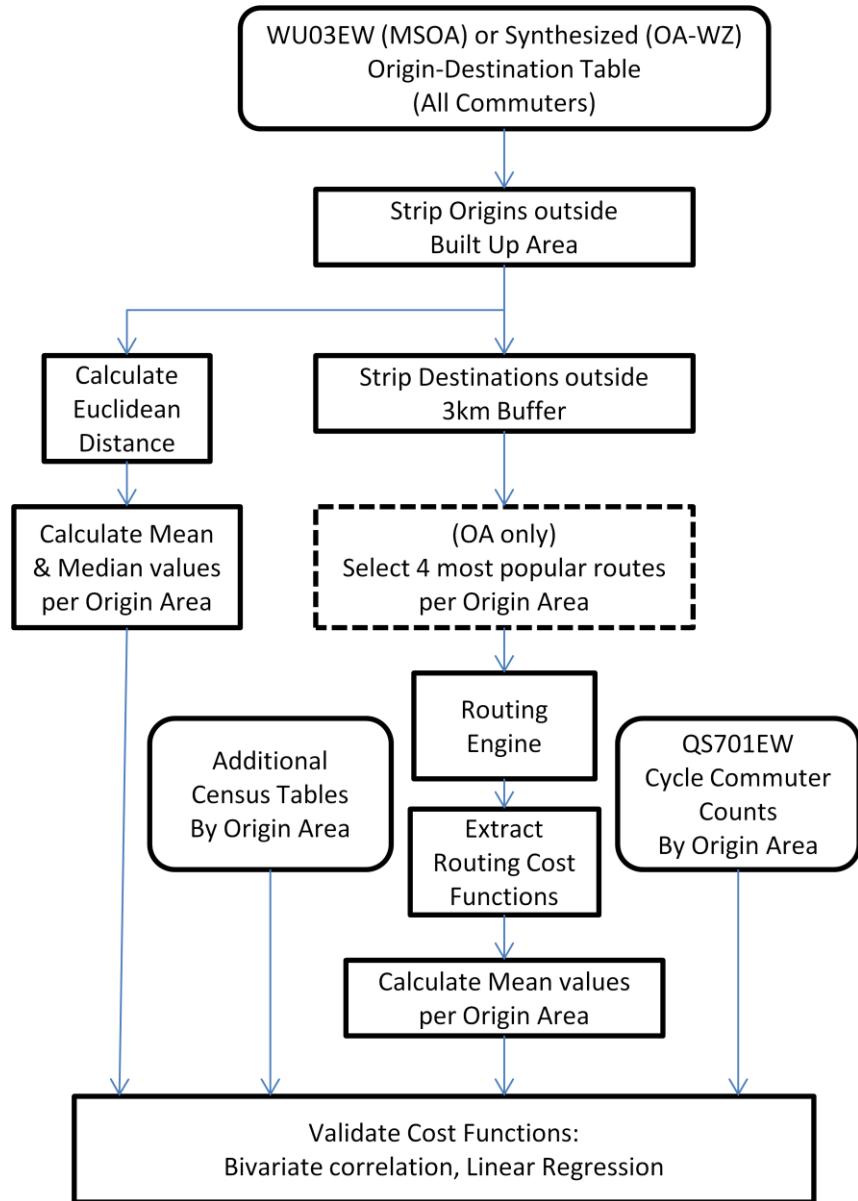


Figure 7: Area-based Regression Flow Diagram (MSOA & OA level)

Figure 7 shows how cost functions for multiple commute flows were combined for each origin area (OA or MSOA). MSOA analysis using census origin-destination (rather than synthetic) data as its input was used to further validate the routing cost functions and to act as a basis for comparison with equivalent OA results based on synthesized origin-destination data. All MSOA flows within the 3km buffer were routed, but only the most popular 4 flows for each OA due to there being 1951 OAs. Note that median and mean Euclidean distances were calculated separately as additional cost functions for *all* flows across the 4 UA, not just those routed. Finally, linear regression based models were developed using area proportions of cycling commuters from table QS701EW as the outcome variables. Various additional census variables indicated as likely to be important cycle mode determinants from the literature (population density, age, sex, car ownership, ethnicity) will also be incorporated. Additionally, the census QS702EW “Distance to Work” binned measures (0-2km, 2-5km etc) were tried in the regression, particularly for comparison with the new Euclidean distance average cost functions.

6 Results

6.1 Synthetic Flow Data Generation (GenSynthFlow)

When synthesizing origin-destination tables, if there were two adjacent origins (or destinations) that had large numbers of commuters for any specified distance range then there would be a high likelihood that commute paths might be swapped over between them. Although when routed this would still give a good representation of typical journey characteristics, it would make it difficult to directly compare the synthesized origin-destination data against the equivalent WU03EW census table. However, to evaluate the distance gravity modelling aspect of the synthesis it can be helpful to look at the distribution of all Euclidean distances produced for each distance interval (Figure 8), particularly when stripped down to the area of interest (Figure 9). To compare actual flows, a visualisation (Figure 10) of synthesized flows (red) overlaying census flows (blue) is quite informative; where flows are common, the colours blend to purple. (Note that the presence of only synthesized (red) flows at the east and southeast edges is due to census flows there being below the threshold being displayed).

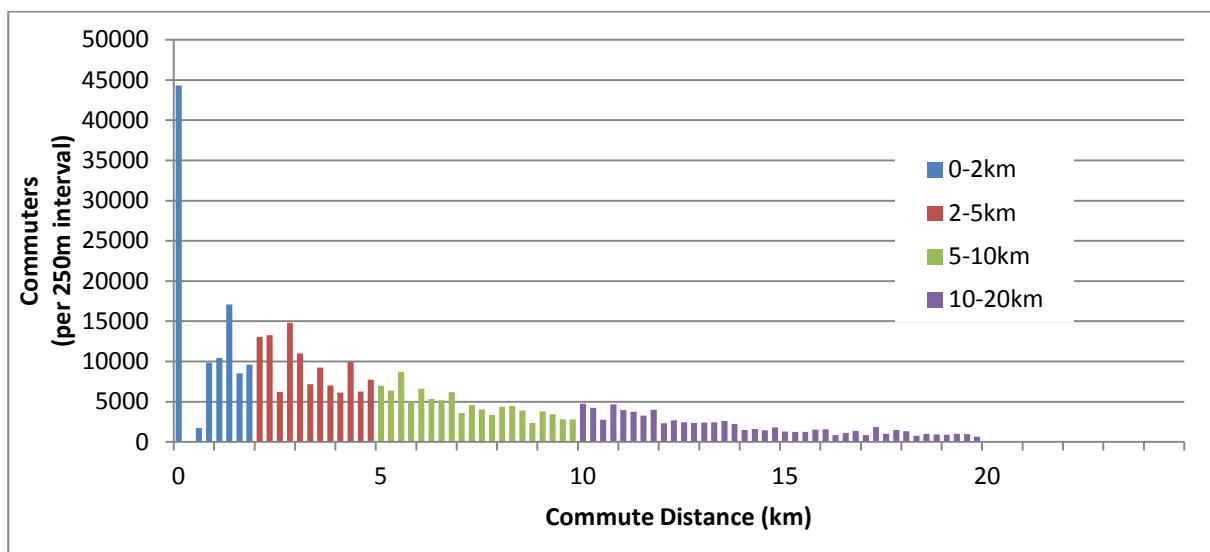


Figure 8: Distance distribution of synthesized MSOA origin-destination data within the 4 Unitary Authorities

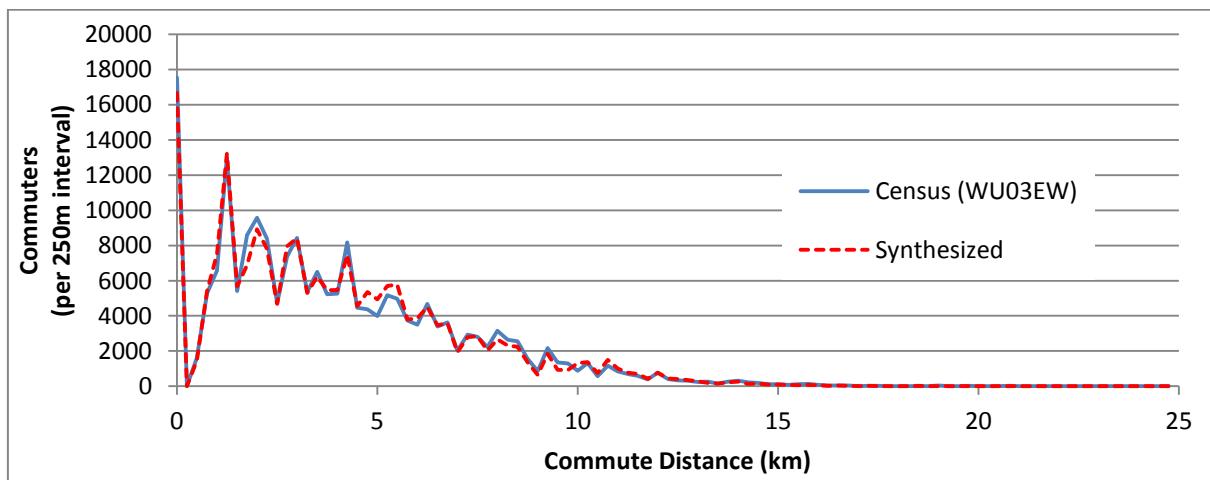


Figure 9: Distance distribution of synthesized and actual MSOA flow data within 3km buffer of BUA

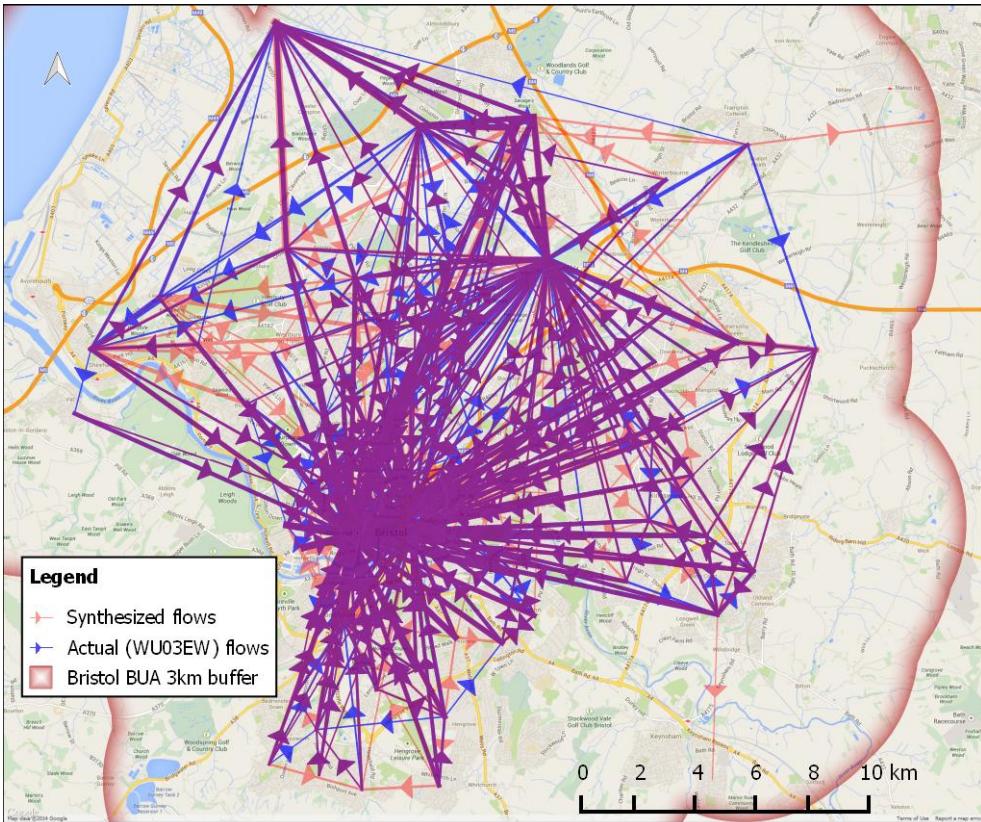


Figure 10: MSOA level commuter flows larger than 100 within 3km buffer of Bristol BUA

Figure 11 shows the results of OA-WZ synthesis with the subset of flows originating in the Bristol BUA in black. There are clear steps occurring at distance interval boundaries suggesting a decay curve with distance that is slightly too aggressive. Within the Bristol BUA, actual total flows per OA varied from 15 ... 270, with the synthesized flows matching this within an error range -1 ... +8. Towards the periphery of the full 4 UA region, actual flow assignment became more difficult as many of the real origins/destinations were missing: error range accordingly increased to -23 ... +34.

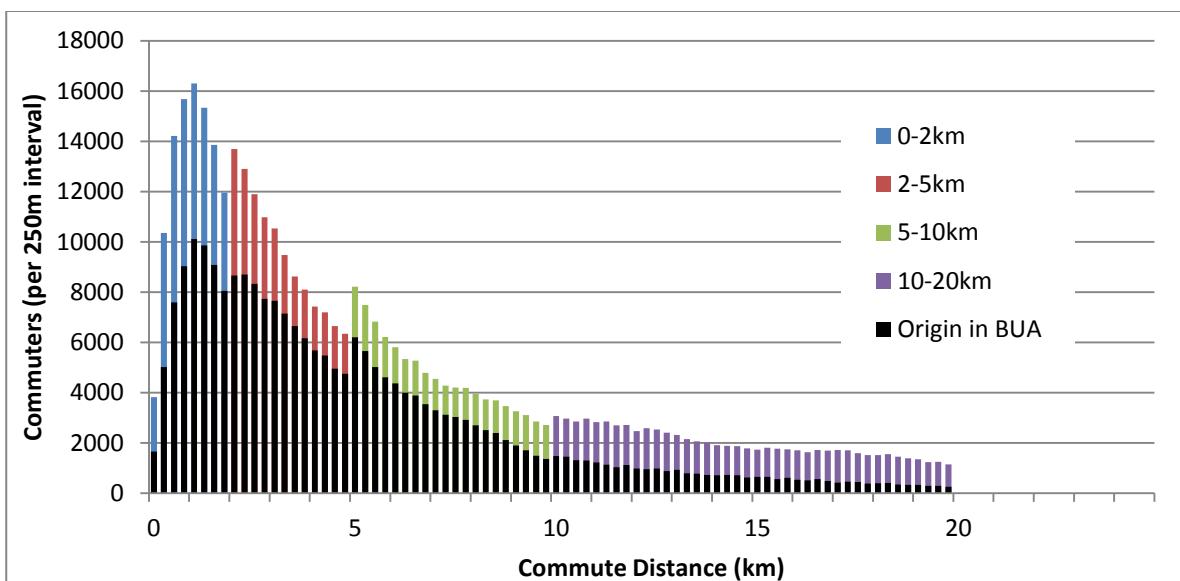


Figure 11: Distance distribution of synthesized OA-WZ origin-destination data within the 4 Unitary Authorities

6.2 Route-based Regression (MSOA level)

The adjusted R² values in Table 8 show how good a model fit each cost function on its own is for predicting proportion of cyclists on each origin-destination flow path. The direction of correlation of EffortRatio and SpeedKmh were the opposite of what might be expected. Checks for variable collinearity using matrix correlation of all variables were made (see Appendix III for full details). These indicated (as might be expected) that EuclidDistKm, RoutedDistKm and TimeMin were all highly correlated, so only one of these could be used in a multivariate regression model. The only other major correlation noted (-0.903) was SpeedKmh and EffortRatio. The somewhat disappointing (Adjusted R² = 0.146) best fit multivariate model that could be produced – even with case weighting by number of commuters per flow (as discussed earlier) – is specified in Table 9 .

Table 8: MSOA PWC Route-based Bivariate Regression against Cycling % of Commuters

Variable (6887 Routes)	Pearson Correlation	Bivariate Linear Regression		
		Adjusted R ²	Unstandardized Coefficient B	Significance p-value
EuclidDistKm	-0.347	0.120	-0.755	0.000
RoutedDistKm	-0.362	0.131	-0.577	0.000
TimeMin	-0.354	0.125	-0.354	0.000
SpeedKmh	-0.076	0.006	-0.259	0.000
EffortRatio	+0.016	0.000	+0.453	0.000
Traffic (%)	+0.036	0.001	+0.024	0.000
Directness (%)	+0.117	0.014	+0.106	0.000

Table 9: MSOA PWC Route-based Multivariate Regression (Best Fit) against Cycling % of Commuters

Variable (6887 Routes)	Multivariate Linear Regression (Adjusted R ² = 0.146)		
	Standardized Coefficient	Unstandardized Coefficient B	Significance p-value
(Constant)	-	1.395	0.000
EuclidDistKm	-0.366	-0.797	0.000
Directness (%)	+0.161	+0.146	0.000

6.3 Area-based Regression (MSOA level)

The mean value of cost functions (prefixed “Av”) from all routed commuter flows weighted by flow count were calculated for each MSOA area. Table 9 shows that the Euclidean distance variables and some of the routing cost variables showed better bivariate model fits than before, though AvEffortRatio and AvTraffic correlations were not statistically significant (p-value > 0.05). Census QS702 distance to work measures were also introduced for comparison (and combination) with the new variables. When testing for possible collinearity between variables, perhaps unexpectedly AvTraffic, AvEffortRatio and AvSpeed were found to be all quite closely correlated (see Appendix III for full correlation details).

Table 11 shows the best fit multivariate model that could be produced using the new variables – none of the routing cost functions were significant when added to this model. Table 12 shows the best fit model using none of the new variables, which was actually better. As an alternative to averaging routed distance, an experiment with splitting it into 3 or 4 quantile bins (similar to census data) was tried, but this gave no improvement.

Table 10: MSOA level Area-based Bivariate Regression against Cycling % of Commutes < 20km

Variable (for 73 MSOAs)	Pearson Correlation	Bivariate Linear Regression		
		Adjusted R ²	Unstandardized Coefficient B	Significance p-value
MeanEuclidDistKm	-0.698	0.480	-3.180	0.000
MedianEuclidDistKm	-0.607	0.369	-2.527	0.000
AvRoutedDistKm	-0.693	0.444	-1.927	0.000
AvTimeMin	-0.641	0.403	-0.464	0.000
AvSpeedKmh	-0.258	0.053	-1.150	0.028
AvEffortRatio	+0.183	(0.020)	(+5.250)	0.122
AvTraffic (%)	+0.194	(0.024)	(+0.176)	0.099
AvDirectness (%)	+0.282	0.067	+0.409	0.016
QS702 Distance 0-2km (%)	+0.385	0.136	+0.146	0.001
QS702 Distance 2-5km (%)	+0.382	0.134	+0.149	0.001
QS702 Distance 5-10km (%)	-0.571	0.317	-0.233	0.000
QS702 Distance 10-20km (%)	-0.614	0.369	-0.634	0.000

Table 11: MSOA Area-based Multivariate Regression (v1) against Cycling % of Commutes < 20km

Variable (for 73 MSOAs)	Multivariate Linear Regression (Adjusted R ² = 0.539)			
	Standardized Coefficient	Unstandardized Coefficient B	Significance p-value	95% Confidence Interval for B
(Constant)	-	11.0	0.024	1.5 – 20.5
QS102 Population Density (/ha)	+0.370	+0.070	0.003	0.025 – 0.115
MeanEuclidDistKm	-0.352	-1.61	0.009	-2.79 – -0.42
QS702 Distance 2-5km (%)	+0.197	+0.077	0.005	0.005 – 0.148

Table 12: MSOA Area-based Multivariate Regression (Best Fit) against Cycling % of Commutes < 20km

Variable (for 73 MSOAs)	Multivariate Linear Regression (Adjusted R ² = 0.579)			
	Standardized Coefficient	Unstandardized Coefficient B	Significance p-value	95% Confidence Interval for B
(Constant)	-	12.758	0.000	8.82 – 16.7
QS702 Distance 10-20km (%)	-0.368	-0.380	0.000	-0.599 – -0.202
QS102 Population Density (/ha)	+0.340	+0.064	0.001	0.029 – 0.100
QS702 Distance 2-5km (%)	-0.265	-0.108	0.005	-0.182 – -0.035

6.4 Area-based Regression (OA level) using Synthetic Data

Using bivariate correlation, the only unexpected correlation detected (full details in Appendix III) was between AvSpeedKmh and AvEffortRatio (-0.882). Table 13 shows bivariate regression results for all new measures and census distance measures of which only the 2-5km measure had a correlation of less than 0.7 with MeanEuclidDistKm. For comparison with MSOA level a multivariate regression with the same variables was done (Table 14) showing a fall in Adjusted R² from 0.539 to 0.292. At OA level this could be improved to 0.312 if AvDirectness and AvTraffic were also included, but once again fit was improved slightly (and collinearity risk reduced) if all new distance measures were abandoned in favour of the census distance measures.

Table 13: OA level Area-based Bivariate Regression against Cycling % of Commutes < 20km

Variable (for 1951 OAs)	Pearson Correlation	Bivariate Linear Regression		
		Adjusted R²	Unstandardized Coefficient B	Significance p-value
MeanEuclidDistKm	-0.488	0.238	-2.498	0.000
MedianEuclidDistKm	-0.485	0.235	-1.590	0.000
AvRoutedDistKm	-0.296	0.087	-0.994	0.000
AvTimeMin	-0.262	0.068	-0.218	0.000
AvSpeedKmh	-0.246	0.060	-1.211	0.000
AvEffortRatio	+0.253	0.063	+10.36	0.000
AvTraffic (%)	+0.117	0.013	+0.105	0.000
AvDirectness (%)	+0.200	0.042	+0.200	0.000
QS702 Distance 0-2km (%)	+0.252	0.063	+0.103	0.000
QS702 Distance 2-5km (%)	+0.310	0.096	+0.130	0.000
QS702 Distance 5-10km (%)	-0.438	0.191	-0.204	0.000
QS702 Distance 10-20km (%)	-0.406	0.165	-0.393	0.000

Table 14: OA Area-based Multivariate Regression (MSOA comparison) against Cycling % of Commutes < 20km

Variable (for 1951 OAs)	Multivariate Linear Regression (Adjusted R² = 0.292)			
	Standardized Coefficient	Unstandardized Coefficient B	Significance p-value	95% Confidence Interval for B
(Constant)	-	+13.7	0.000	12.2 – 15.3
MeanEuclidDistKm	-0.373	-1.91	0.000	-2.22 – -1.69
QS702 Distance 2-5km (%)	+0.221	+0.093	0.000	0.076 – 0.109
QS102 Population Density (/ha)	+0.146	+0.016	0.000	0.012 – 0.021

Table 15 describes the best fit multivariate linear regression model found when additional census variables were added based on experimentation with recommendations from the literature. It was noted that although cycling showed a strong negative correlation with two or more cars per household, it showed a positive correlation with owning one car. Adding a variable for gender was not significant within the

regression, even if it was added in the cross tabulated form of males 25-34 or 35-39 where it made less contribution than all people in those age groups. White ethnicity was also not significant. The actual cycle commute proportions are mapped in Figure 12 with the model predictions in Figure 13 and the model errors in Figure 14 (note these are absolute errors in the % cycling, not % errors in the % cycling).

Table 15: OA Area-based Multivariate Regression (Best Fit) against Cycling % of Commutes < 20km

Variable (for 1951 OAs)	Multivariate Linear Regression (Adjusted R ² = 0.357)			
	Standardized Coefficient	Unstandardized Coefficient B	Significance p-value	95% Confidence Interval for B
(Constant)	-	-30.7	0.000	-35.1 – -26.4
QS702 Distance 2-5km (%)	+0.719	+0.302	0.000	0.264 – 0.339
QS702 Distance 0-2km (%)	+0.716	+0.291	0.000	0.251 – 0.331
QS702 Distance 5-10km (%)	+0.325	+0.151	0.000	0.108 – 0.195
LC6107 Age25-34(%)	+0.226	+0.117	0.000	0.091 – 0.142
LC6107 Age35-49 (%)	+0.185	+0.134	0.000	0.102 – 0.167
AvDirectness	+0.134	+0.130	0.000	0.095 – 0.166
LC4609 No Cars in Household (%)	-0.105	-0.050	0.000	-0.074 – -0.026
QS102 Population Density (/ha)	+0.097	+0.011	0.000	0.006 – 0.015

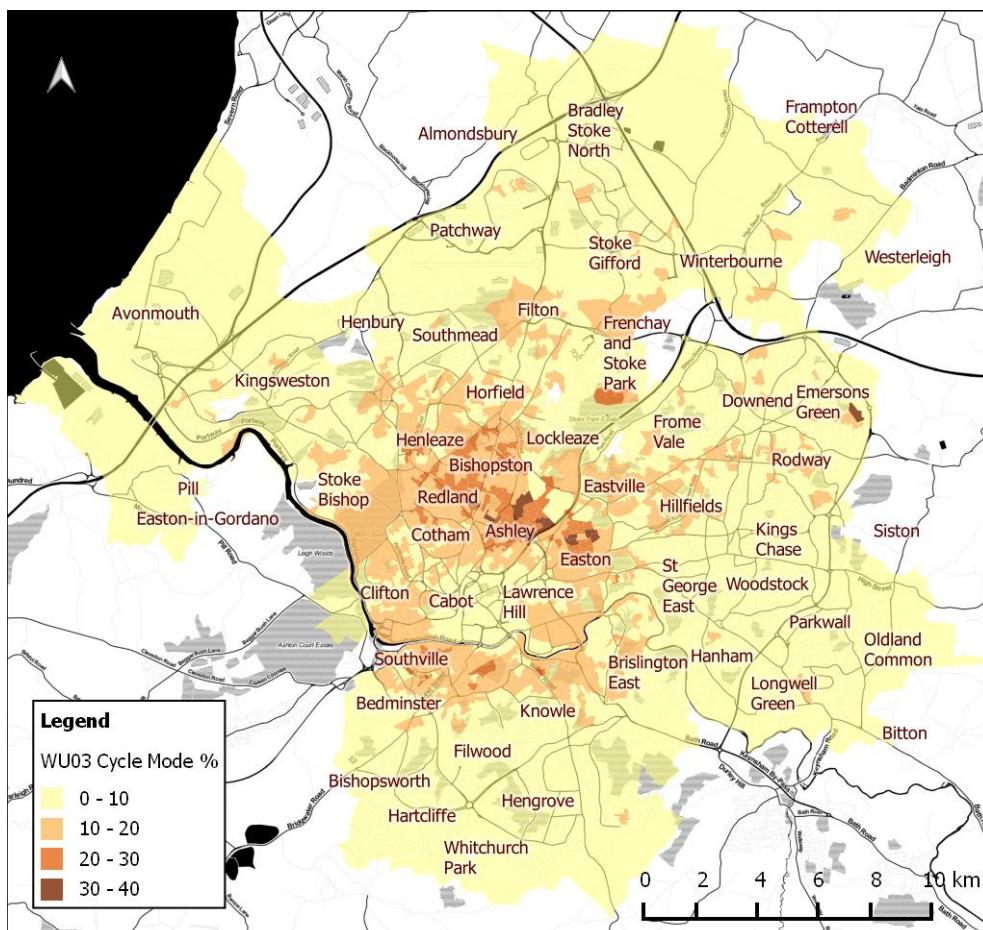


Figure 12: Actual (Census recorded) Cycle Commute Proportions by area (OA) for Bristol BUA

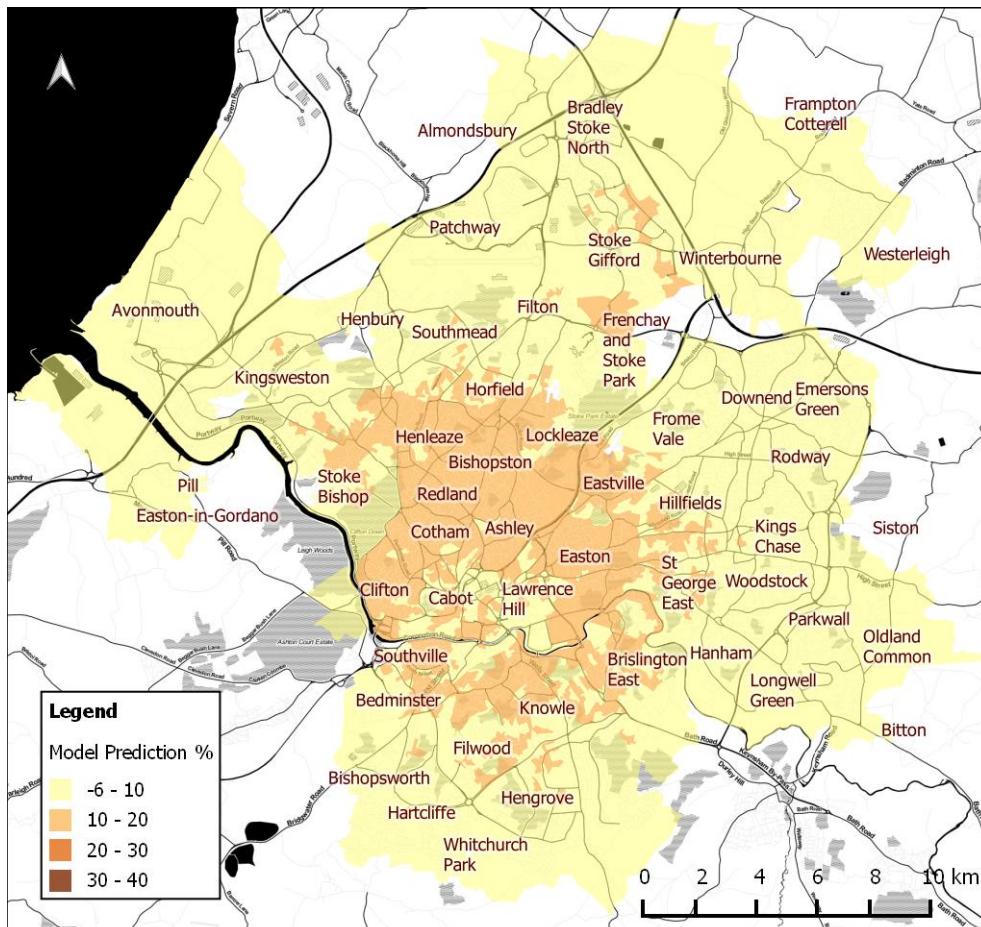


Figure 13: Predicted Cycle Commute Proportions by area (OA) for Bristol BUA

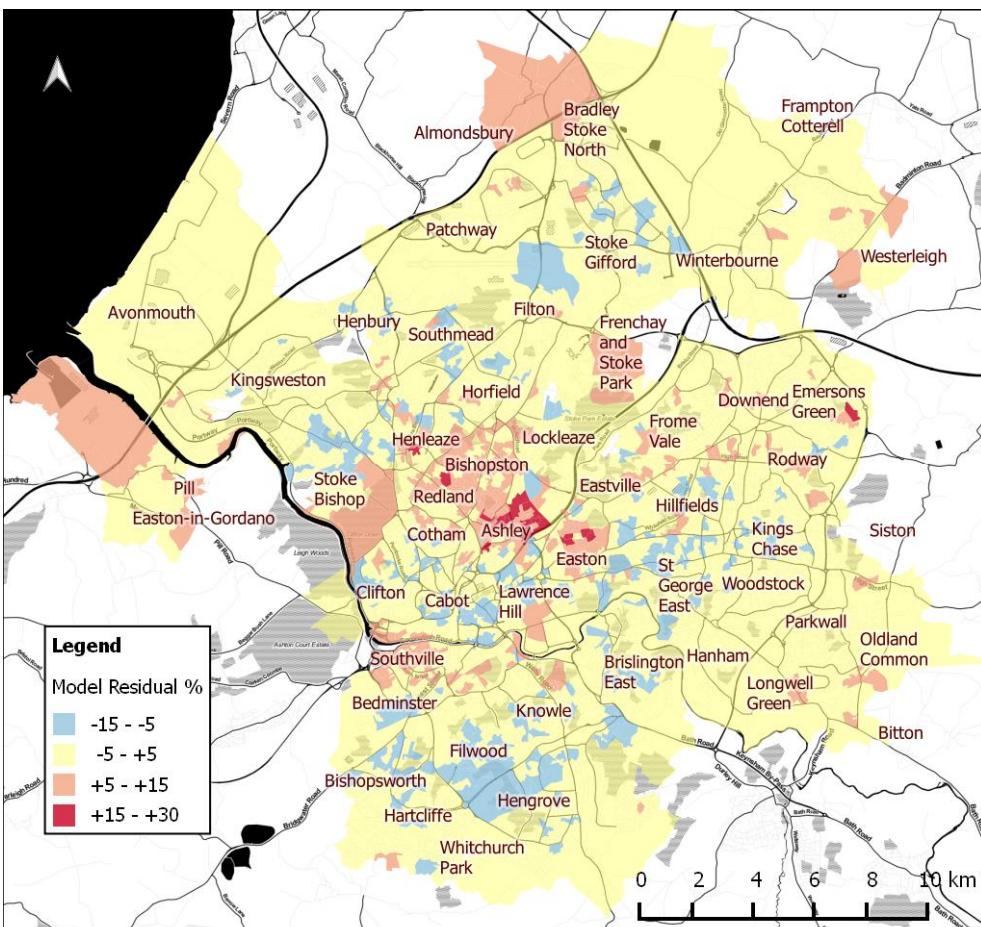


Figure 14: for Cycle Commute Proportion Model by area (OA) for Bristol BUA

7 Analysis and Discussion

7.1 Synthetic Flow Data Generation (GenSynthFlow)

Although it proved difficult to find a suitable statistical measure for overall validation of the synthesis against census MSOA origin-destination data, the measures and visualisations shown in the results suggest that the synthesis should be reasonably representative for the routing requirements of this study.

Gravity model-based optimisation: The issue highlighted of overly-steep distance decay curves could be fixed by modifying the rate of reduction of the ‘mismatch threshold’, but a less arbitrary approach would be better. One approach to improving adaptation towards an optimal solution could be to consider swapping over two flow origins in one step, rather than just moving a single flow to a different origin.

Initial population: In tests where the synthesis was repeated with many different initial populations, it was found that after each iteration sequence had stabilised there was very little difference between final largest mismatches as a proportion of total commuters for each flow – suggesting insensitivity to initial population. For this specific synthesis task, a more representative result might be generated by starting with an initial population based on 2001 census flow data, albeit with the reservations stated earlier about problems with this data. However, such an approach is made more complicated by the change in destination geographies to WZs in the 2011 census.

7.2 Cycle Routing of Synthesized (OA level) Flows

Before aggregating results of the cycle routing engine by each journey origin area, it can be informative to examine estimated routes themselves. Figure 15 illustrates the counts for each route if all commuters were to cycle. Where more than one route overlaps, counts were summed to give an indication of potential total traffic flow. Large summed flows can imply several things: high demand, desirable routes (such as traffic-free cycle paths) or lack of available alternatives (network impermeability). To see how well existing cycling infrastructure caters for potential commuting routes, it is helpful to overlay both on the same map (Figure 16); for clarity, routed flows have been reduced to a single colour and the base map has been removed. It should be noted that the ‘balanced’ routing algorithm used will try to take advantage of cycle infrastructure, but only where it does not introduce too much of a detour.

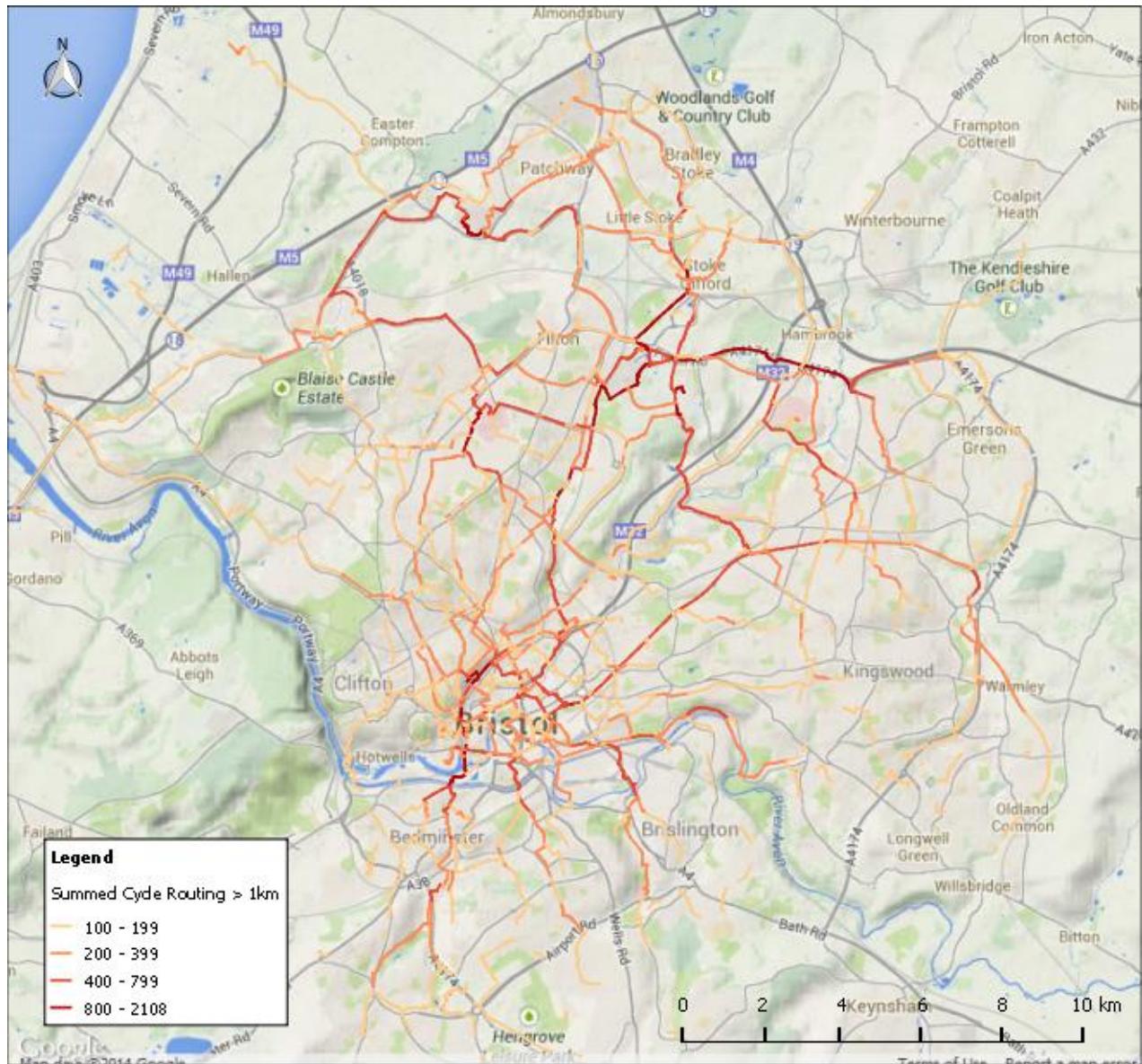


Figure 15: Summed Cycle Routing for Commutes less than 20 km from Bristol BUA



Figure 16: Cycle Routing compared to Cycle Infrastructure in Bristol BUA

7.3 Cost Functions Aggregated by Origin Area

Before combining cost functions in the final area-based regression analysis, each are considered on their own to investigate what exactly they reveal and how this might contribute as a determinant of commuter cycle mode choice.

7.3.1 Mean Euclidean Distance

Euclidean (“as the crow flies”) distance was calculated for all flows less than 20km from the Bristol BUA with destinations within the 4 UA region as (unlike the other cost functions) it was not dependent on routing. The results aggregated by OA are shown in Figure 17 together with proportionally-sized indicators of WZ destinations that have more than 1000 workers. Although WZ areas are sized to try to have similar number of workers in each, particularly large counts indicate a single big employment site (such as a single employer, retail block or business park). The largest of these represents the MOD Abbey Wood site in north Bristol which employed over 10,000 people by 2012. Although many people in Bristol commute to other cities, because range is being limited to less than 20km, this map shows mainly those commuting within Bristol, although for those living on the south-east side, the city of Bath is within range. The results suggest that most people were commuting into the centre or north of the city. The exception is the north-west periphery of the city, probably due to the large dockside industrial estates in the Avonmouth area.

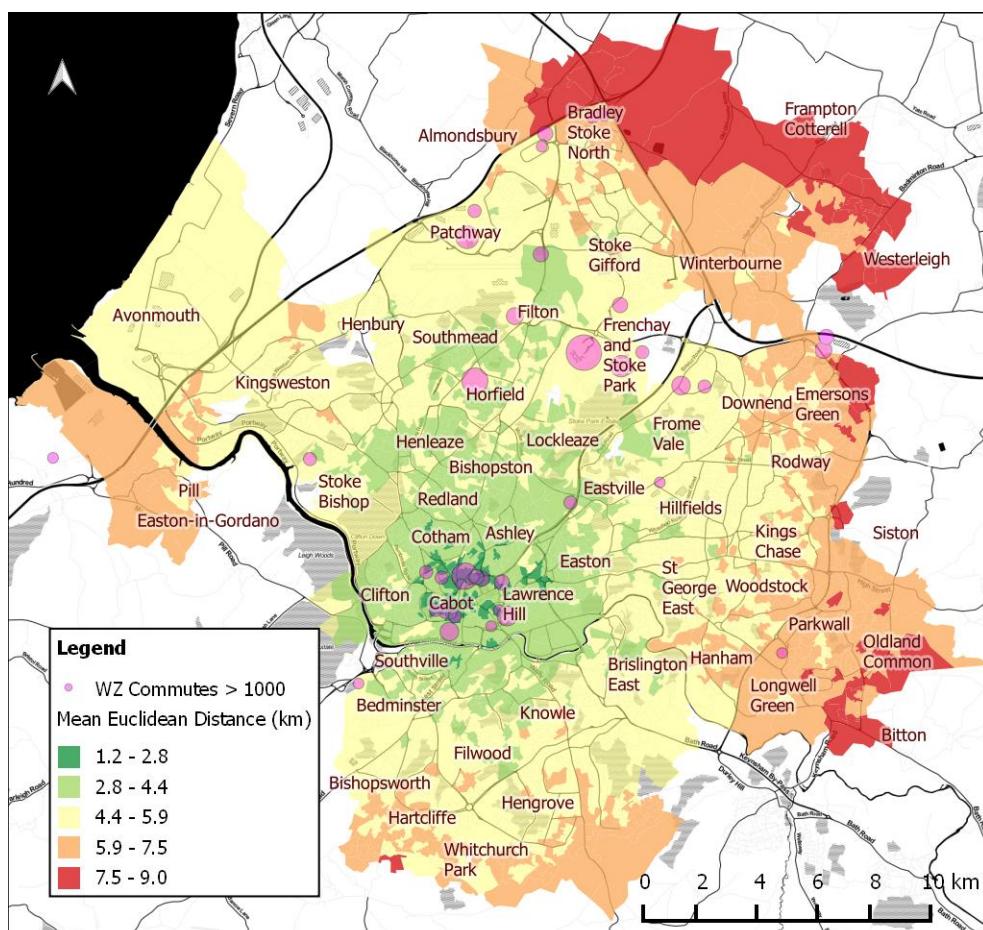


Figure 17: Mean Euclidean Distance by OA for Commutes less than 20 km from Bristol BUA

7.3.2 Travel Time

In order to calculate travel time (and effort required), it was necessary to perform routing in both directions of each commute and average them. Although travel time was too highly correlated with distance to be additionally used in a regression model, when mapped (Figure 18) it is interesting to note that the average cycle commute time for most people living the Bristol BUA (and commuting less than 20km) would be less than 25 minutes. (The large red area in the north is a single very sparsely populated rural OA).

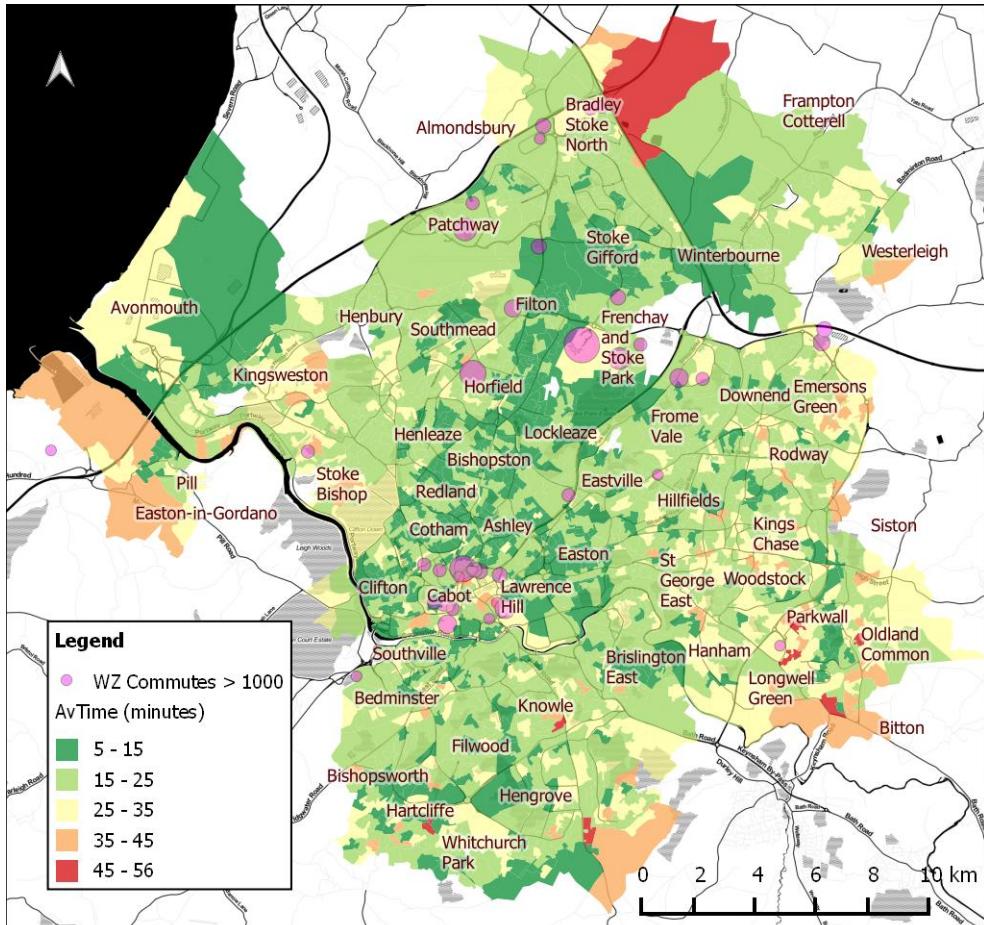


Figure 18: Mean Travel Time by OA for Commutes less than 20 km from Bristol BUA

7.3.3 Directness

This measure of directness (Euclidean distance / routed distance) can indicate how well connected an area is; surrounding the larger red area in Figure 19 is an airport and a golf course. If (effective) connectedness was the same for all forms of transport then it should have little bearing on cycle mode choice. However, because cycle routing was done using the ‘balanced’ mode rather than ‘fastest’, then directness should also indicate where cycle routes were diverted away from direct paths, perhaps to either avoid major roads or to take advantage of nearby cycle infrastructure. A more useful measure might have been to compare distance between ‘balanced’ and ‘fastest’ routing algorithms. The ‘cycle routing’ shown on the map is all routes with summed commuter counts greater than 50 and helps explain the directness measure for larger areas. However, the large variation in directness at small area resolution could be a useful insight into how

difficulty in finding acceptable routes close to home can dissuade people from considering cycle commuting.

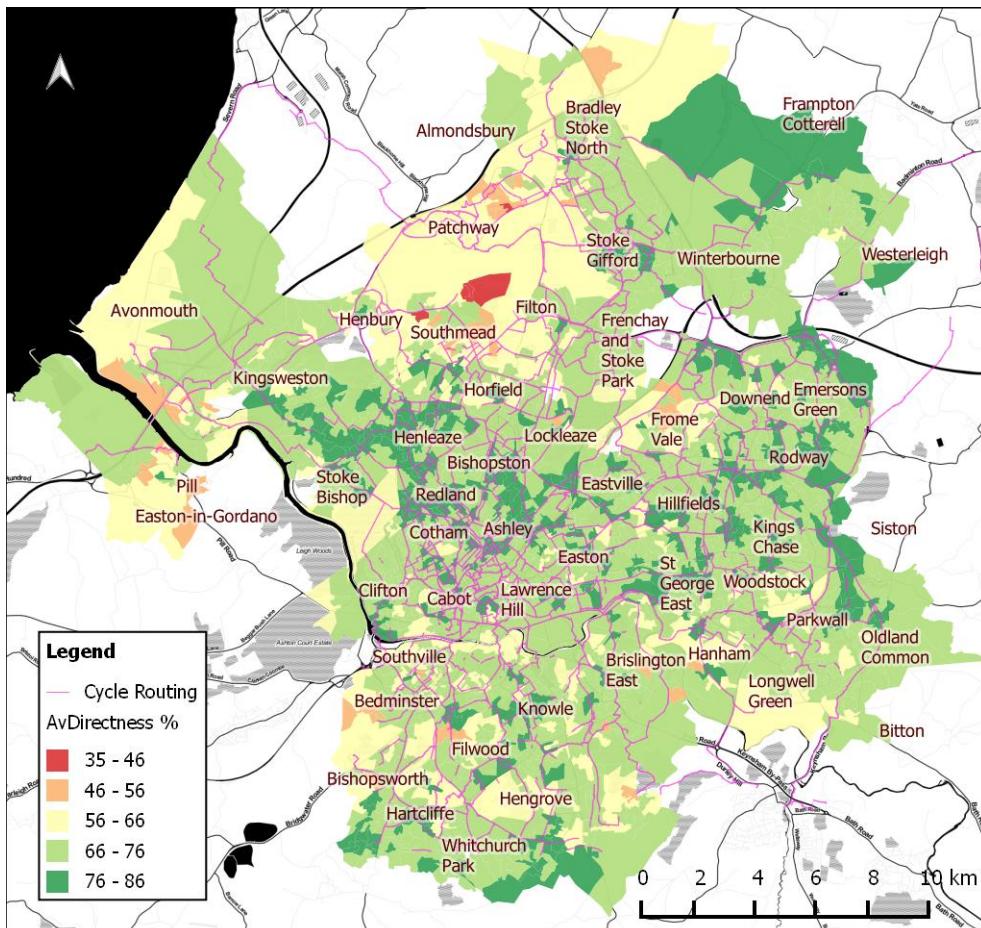


Figure 19: Mean Directness by OA for Commutes less than 20 km from Bristol BUA

7.3.4 Cycling in Traffic

The AvTraffic cost function (Figure 20) is intended to indicate how likely commute routes from an area would involve cycling in traffic. By plotting locations of cycle infrastructure on the map it can be seen that in most areas, the presence of local traffic-free cycle paths tallies with high values of this measure, suggesting the infrastructure is located along or close to actual commuting routes. However there are notable exceptions: the areas marked red in the centre and south of the city imply that in these places either infrastructure is not on commuting routes and/or that the destinations commuted to by people in this area involve much longer sections of high traffic volumes. Either way, these exceptions (and the comparison between routes and infrastructure made earlier in Figure 16) suggest that this route-based measure should be a more representative determinant of the effect of cycle infrastructure on cycle mode choice than area-based averages of cycle infrastructure density such as those used by Parkin et al (2008).

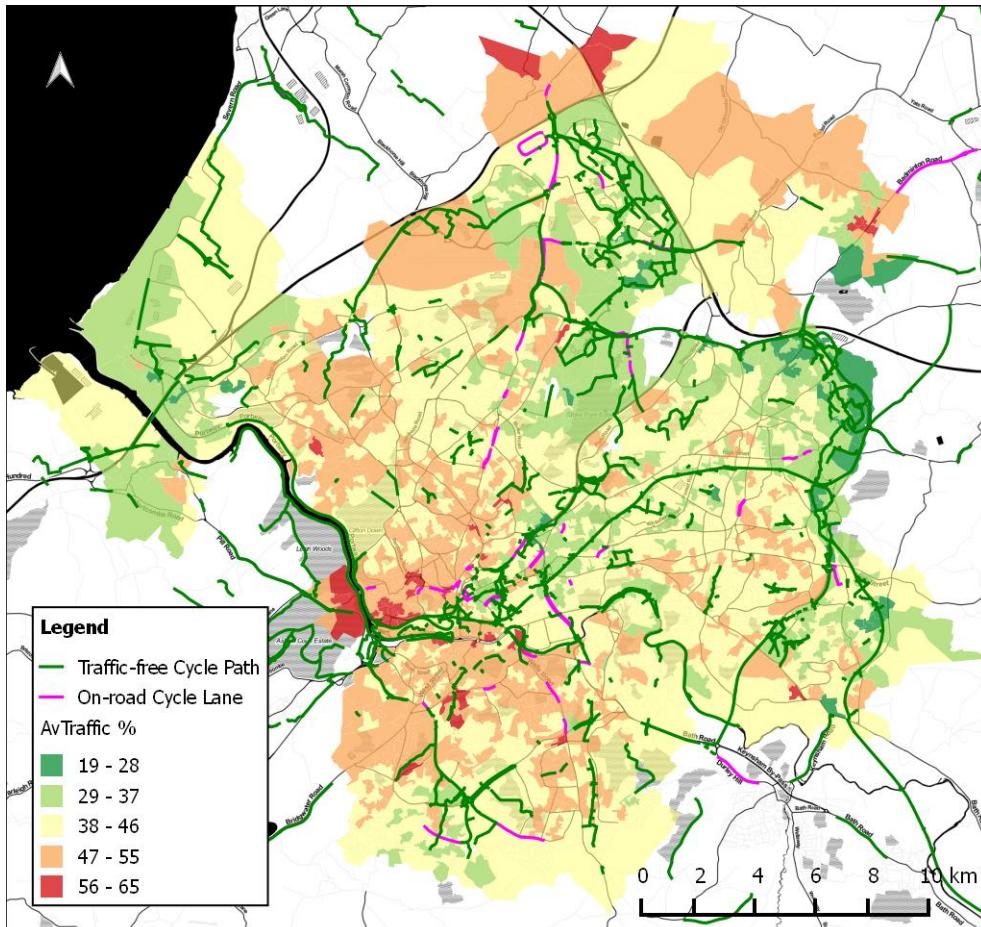


Figure 20: Mean Traffic Exposure by OA for Commutes less than 20 km from Bristol BUA

7.3.5 Effort to Distance Ratio

This measure was intended to indicate how much effort would be required de-correlated from journey length. It was anticipated that this would strongly reflect the impact of hills on commute routes (noted as a strong cycle mode determinant in the literature). However comparing the AvEffortRatio measure (Figure 21) with terrain (Figure 22) this does not appear to be so – particularly for those living in areas at the top of hills. Thus the measure is probably dominated by the estimation of the amount of stopping and starting required – which would tally with it generally being worse in the more congested urban areas with many traffic control measures (such as traffic lights) closer to the centre of the city. It may also explain why as well as sharing a strong negative correlation with AvSpeedKmh, the variable shows a *positive* correlation with cycle mode choice: stops and starts can be more of an impediment to journey time for cars than bicycles, particularly where bicycles can cut through congested traffic.

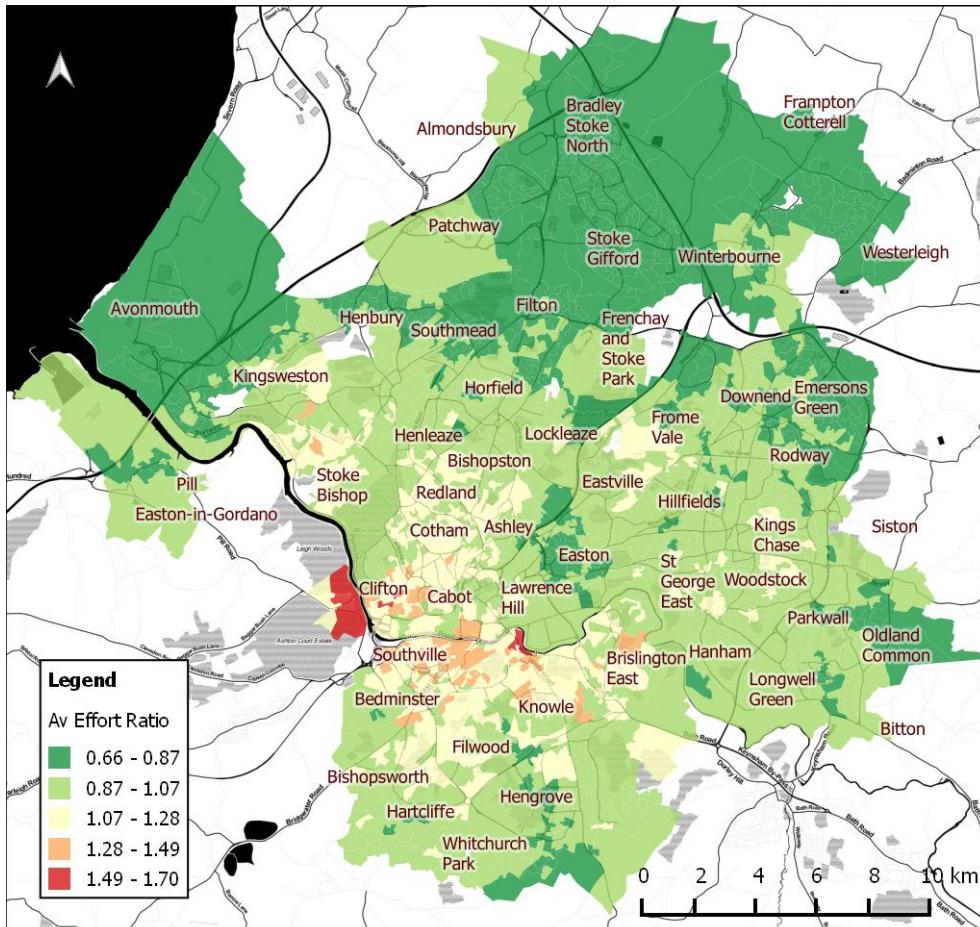


Figure 21: Mean Effort to Distance Ratio by OA for Commutes less than 20 km from Bristol BUA

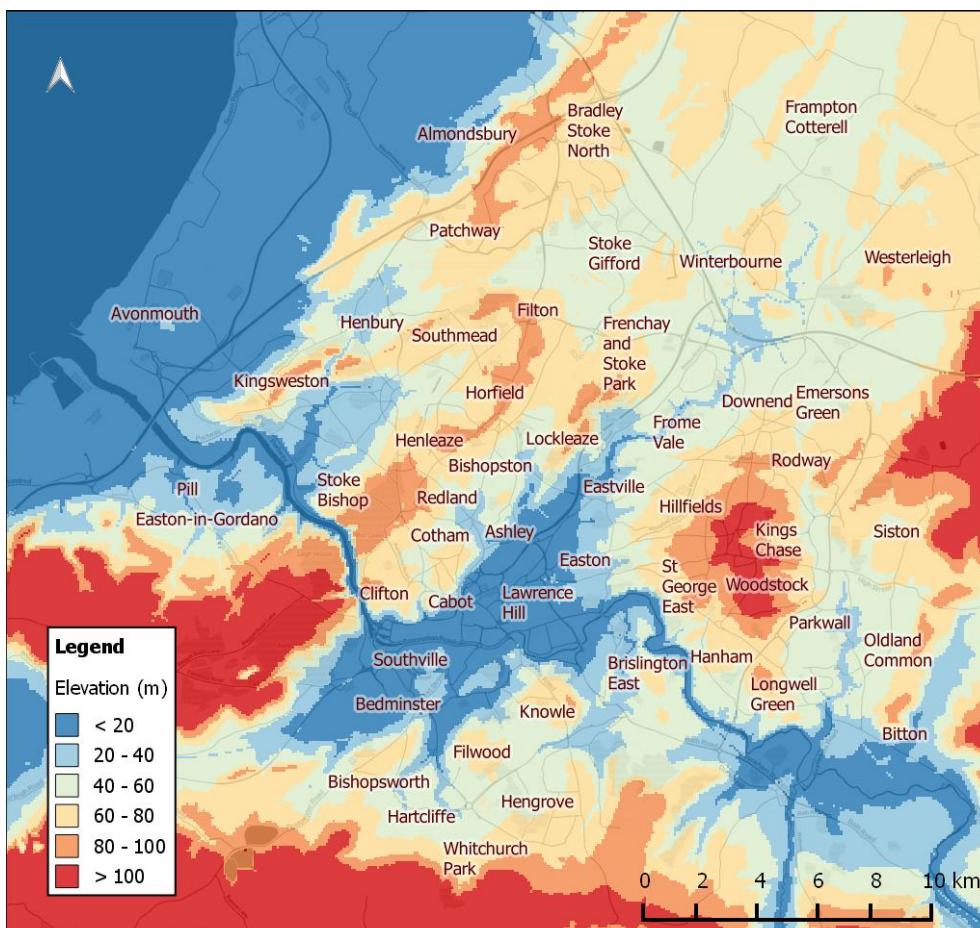


Figure 22: Digital Terrain Map (DTM) for Bristol Area

7.4 Regression Analysis

MSOA Route-based Analysis: This gave disappointingly poor results, with only the Directness measure adding a useful contribution to the regression model. This could reflect that MSOAs are quite large areas so their centres (PWCs) might be a long way from true commute origins and destinations, thus making routing unrepresentative; in previous maps of OA level routing costs, there is typically large variation in values across each MSOA-sized area.

MSOA Area-based Analysis: both AvTraffic and AvEffortRatio were not statistically significant, perhaps reflecting their lack of strong influence and the relatively small sample size (73 MSOAs). AvDirectness did measure as significant, but made no useful contribution to the model which was dominated by distance and population density measures. This suggests that these routing-based cost functions are not useful at MSOA level. The overall fit of the model ($\text{Adjusted } R^2 = 0.579$) using only 3 variables is reasonable (it explains nearly 60% of cases) though less than the 0.81 achieved in a model at similar area granularity (Wards) by Parkin et al (2008). However, that model used many more variables and a much larger sample size.

OA Area-based Analysis: Although all new cost functions were statistically significant, only AvDirectness made a useful contribution to the fit of the final model – one that was greater than number of cars or population density (both of which are noted in literature as useful determinants). The overall fit of the OA-level model ($\text{Adjusted } R^2 = 0.292$) is much worse than at MSOA level using exactly the same variables. This is not entirely surprising as when data is aggregated to larger areas, detail is lost and areas appear more similar to each other – thus correlations would naturally increase. Visually, the model (Figure 13) appears a reasonable fit to the actual (Figure 12) cycle mode proportions, albeit underestimating the larger values (just north of the centre). However, the Adjusted R^2 measure for OA fit is poor enough to suggest that regressions done at such a small area level might not be useful.

Other variables: In the map of OA regression model errors (Figure 14), there is some apparent spatial clustering which suggests that key variables are missing from the model. The under-prediction (red) in areas just north of the centre coincides with quite affluent areas of the city and large student populations, neither of which was represented in the model.

8 Conclusions

This study has investigated two new methods which when combined aim to improve determination of commuting cycle mode choice. The synthetic flow data generation method proved difficult to thoroughly validate quantitatively, but indicator measurements were promising and when mapped it visually gave a representatively close match with 2011 census MSOA-level flow data for the city of Bristol. When applied at OA level, error indicator results were promising, though small issues were noted with decay curves of commuter count against distance travelled. For this specific analysis, results might be further improved if an initial population was created based on available earlier census flow data from 2001.

The second method involved creation of cost functions for commuting routes based on a third party cycle routing engine and aggregation of these by origin (residential) area. The initial validation attempt involving route-based regression at MSOA level failed, though this was considered at least partly due to fine-grain routing adding little extra information given the large uncertainty in origin/destination location in areas the size of MSOAs. A more revealing analysis might be to repeat this step at OA level when (and if) 2011 census OA cycle flows become available in a form suitable to allow this.

The routing cost functions chosen varied in their usefulness:

- ‘**Directness**’ (ratio of Euclidean to Routed commute distance) made a notable contribution to the final regression model and could possibly be further improved by replacing the Euclidean distance measure with a ‘fastest’ routed distance.
- ‘**EffortRatio**’ (estimated energy divided by routed distance) intuitively should be a better determinant of cycle mode choice than just a measure of hilliness (a documented cycle commuting detractor) as it also incorporates energy expended at likely stop/start locations. However, such stop/start behaviour may be more of a hindrance to car drivers (certainly in its impact on speed), so this measure is not ideal for regression analysis. However, as the routing engine used directly returned elevation measures along the route, a direct measure of hilliness could be derived.
- ‘**Traffic**’ (likelihood of cycling in motor traffic) proved representative of the level of cycle infrastructure, but was not helpful in cycle mode choice regression. This could indicate that infrastructure has little impact on cycle mode choice, but the measure gives little indication of how much the infrastructure is actually required at each location. Although it includes a rating for general types of road, the measure could be improved by combining it with estimated motor traffic counts from each road taken from established transport models based on actual traffic count data.

With area-based regression at MSOA level, none of the routing-derived functions made a useful contribution to the (quite good) overall model fit, further confirming the suggestion that MSOAs are too coarse a level for routing to be representative. However, better results might have been achieved with a sample size larger than 73 MSOAs. At OA level, ‘Directness’ (only) did make a useful contribution to the

cycle mode choice regression model but overall model fit was poor. Due to resource limitations, only the four most popular commute routes from each OA had been modelled; more representative measures based on a larger number of routes might have improved regression results. It is thus recommended that a hybrid approach should be tried: calculating routing costs at OA level (using improvements to cost functions outlined above) perhaps on a large number of routes, then aggregating these results up to a higher (LSOA/MSOA) level in order to take advantage of the better overall model fit achievable with the other data.

Within the current study some shortcomings of aggregate data analysis were noted: there was a lack of statistical significance for (univariate) gender measures in the regression model (despite widely documented male cycle commute preference). This was probably due to gender balance being very similar across areas. Cross-tabulation with other variables would help, but there was little opportunity for this at OA level. Thus the potential power of routing-based cost functions might lie in non-aggregate modelling where each individual's commuting (or general utility) cycling origin-destination locations are known.

An additional potentially useful output from this study is the generation of maps of summed routes of simulated cycle commuter flows (fig 15). Even when based on synthesized flow data these should give a good representation of potential cycle network hotspots where it would be worthwhile to target provision or upgrading of suitable cycle infrastructure. To relate such a function to the city of the analysis: in the recently published draft Bristol Cycle Strategy (Bristol City Council, 2014a) there is a strong emphasis on the provision of a coherent network of cycle 'freeways' specifically aimed at efficient commuting and utility cycling. It is clear from this study that consideration of the effects of fine-grain route details tallied with realistic commuting flows is important. Knowing potential hotspots for cycle commuting based on likely commuting flows could thus be helpful in designing and staging the implementation of such a network.

9 References

- Aldred, R. 2014. *A Matter of Utility? Rationalising cycling, cycling rationalities*. [Online]. [Accessed 25 August 2014]. Available from:
<http://rachelaldred.org/category/writing/>
- Badland, H. and Schofield, G. 2006. Perceptions of replacing car journeys with non-motorized travel: exploring relationships in a cross-sectional adult population sample. *Preventive Medicine*. **43**(3), pp.222-225.
- Barbour, R. and Fricker, J.D. 1994. Estimating an origin-destination table using a method based on shortest augmenting paths. *Transportation Research Part B: Methodological*. **28**(2), pp.77-89.
- Bristol City Council. 2014a. *Bristol Cycling Strategy (Draft July 2014)*. [Online]. [Accessed 25 August 2014]. Available from:
<https://www.citizenspace.com/bristol/city-transport/cycle-strategy>
- Bristol City Council. 2014b. *2011 Census Topic Report: Who cycles to work?* [Online]. [Accessed 25 August 2014]. Available from:
<http://www.bristol.gov.uk/>
- Bristol City Council. 2014c. *Quality of Life Survey*. [Online]. [Accessed 25 August 2014]. Available from:
<http://www.bristol.gov.uk/page/council-and-democracy/quality-life-bristol>
- Bristol Cycling Campaign. 2014a. *Bristol Cycling Manifesto: Evidence*. [Online]. [Accessed 25 August 2014]. Available from:
<http://bristolcyclingmanifesto.org.uk/evidence/>
- Bristol Cycling Campaign. 2014b. *Bristol Cycling Campaign Manifesto*. [Online]. [Accessed 25 August 2014]. Available from:
<http://bristolcyclingcampaign.org.uk/campaign/manifesto>
- Broach, J.Dill, J. and Gliebe, J. 2012. Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice*. **46**(10), pp.1730-1740.
- Buehler, R. and Pucher, J. 2011. Sustainable transport in Freiburg: lessons from Germany's environmental capital. *International Journal of Sustainable Transportation*. **5**(1), pp.43-70.
- Buehler, R. and Pucher, J. 2012. Cycling to work in 90 large American cities: new evidence on the role of bike paths and lanes. *Transportation*. **39**(2), pp.409-432.
- Chatterjee, K. Ricci, M. Clayton, B. Bartle, C. and Parkin, J. 2013a. *West of England Sustainable Travel (WEST) Baseline and Year One (2012/13) Annual Outcomes Monitoring Report* [Online.] UWE Centre for Transport & Society. [Accessed 25 August 2014]. Available from:
<http://www1.uwe.ac.uk/et/research/cts/researchprojectsbytheme/sustainablemobilitystrategy/evaluationwest.aspx>
- Chatterjee, K.Sherwin, H. and Jain, J. 2013b. Triggers for changes in cycling: the role of life events and modifications to the external environment. *Journal of Transport Geography*. **30**, pp.183-193.
- Cleaveland, F. and Douma, F. 2009. The impact of bicycling facilities on commute mode share. In: *88th Annual Meeting of the Transportation Research Board, Washington, DC*.
- Cope, A. Cairns, S. Fox, K. Lawlor, D. Lockie, M. Lumsdon, L. Riddoch, C. And Rosen, P. 2003. The UK National Cycle Network: an assessment of the benefits of a sustainable transport infrastructure. *World Transport Policy & Practice*. **9**(1), pp.6-17.
- Cope, A.Abbess, C. and Parkin, J. 2007. Improving the empirical basis for cycle planning. In: *4th IMA International Conference on Mathematics in Transport*.
- CROW. 1993. *Sign up for the Bike. Design Manual for a Cycle Friendly Infrastructure*. Record 10. Centre for Research and Contract Standardisation in Civil and Traffic Engineering. The Netherlands.
- CycleStreets.net. 2014. *Journey Planner API (Application Programming Interface)*. [Online]. [Accessed 25 August 2014]. Available from:
<http://www.cyclestreets.net/api/v1/journey/>
- DfT. 2011. *Road Transport Forecasts 2011*. [Online]. Department for Transport. [Accessed 25 August 2014]. Available from:
<https://www.gov.uk/>
- DfT. 2014a. *Transport analysis guidance: WebTAG*. [Online]. Department for Transport. [Accessed 25 August 2014]. Available from:
<https://www.gov.uk/transport-analysis-guidance-webtag>

- DfT. 2014b. *Walking and Cycling Statistics (Active People Survey)*. [Online]. Department for Transport. [Accessed 25 August 2014]. Available from: <https://www.gov.uk/government/collections/walking-and-cycling-statistics>
- DfT. 2014c. *Traffic Counts*. [Online]. Department for Transport. [Accessed 25 August 2014]. Available from: <http://www.dft.gov.uk/traffic-counts/>
- Dill, J. and Carr, T. 2003. Bicycle commuting and facilities in major US cities: if you build them, commuters will use them. *Transportation Research Record: Journal of the Transportation Research Board*. **1828**(1), pp.116-123.
- Dill, J. and Glieme, J. 2008. *Understanding and measuring bicycling behavior: A focus on travel time and route choice*. [Online]. Oregon Transportation Research and Education Consortium. OTREC-RR-08-03 [Accessed 25 August 2014]. Available from: <http://www.otrec.us/>
- Ding, Y. Mirchandani, P.B. and Nobe, S.A. 1997. Prediction of network loads based on origin-destination synthesis from observed link volumes. *Transportation Research Record: Journal of the Transportation Research Board*. **1607**(1), pp.95-104.
- Di Prampero, P. Cortili, G. Mognoni, P. and Saibene, F. 1979. Equation of motion of a cyclist. *J Appl Physiol*. **47**(1), pp.201-206.
- EDINA. 2014a. *UK Data Service Census Support*. [Online]. [Accessed 9 March 2014]. Available from: <http://edina.ac.uk/census/>
- EDINA. 2014b. *EDINA Digimap Ordnance Survey Service*. [Online]. [Accessed 25 August 2014]. Available from: <http://digimap.edina.ac.uk/>
- Eash, R. 1999. Destination and mode choice models for nonmotorized travel. *Transportation Research Record: Journal of the Transportation Research Board*. **1674**(1), pp.1-8.
- Ehrgott, M. Wang, J.Y. Raith, A. and Van Houtte, C. 2012. A bi-objective cyclist route choice model. *Transportation research part A: policy and practice*. **46**(4), pp.652-663.
- Gatersleben, B. and Appleton, K.M. 2007. Contemplating cycling to work: Attitudes and perceptions in different stages of change. *Transportation Research Part A: Policy and Practice*. **41**(4), pp.302-312.
- Golbuff, L. and Aldred, R. 2011. Cycling policy in the UK: a historical and thematic overview. [Online]. University of East London Sustainable Mobilities Research Group, London. [Accessed 25 August 2014]. Available from: <http://rachelaldred.org/>
- Goodman, A. 2013. Walking, cycling and driving to work in the English and Welsh 2011 census: trends, socio-economic patterning and relevance to travel behaviour in general. *PLoS one*. **8**(8), e71790.
- Goodwin, P. 2013. *Get Britain cycling: report from the inquiry*. [Online]. All Party Parliamentary Cycling Group. [Accessed 25 August 2014]. Available from: <http://allpartycycling.org/inquiry/>
- Graham, R. 1998. The delaying effect of stops on a cyclist and its implications for planning cycle routes. In: *Third IMA International Conference on Mathematics in Transport Planning and Control*.
- Greater Bristol Cycling City. 2011. *End of project report*. [Online]. [Accessed 25 August 2014]. Available from: <http://www.betterbybike.info/>
- Hamilton-Baillie, B. 2008. Shared space: reconciling people, places and traffic. *Built environment*. **34**(2), pp.161-181.
- Handy, S.L. and Xing, Y. 2011. Factors correlated with bicycle commuting: A study in six small US cities. *International Journal of Sustainable Transportation*. **5**(2), pp.91-110.
- Harland, K. Heppenstall, A. Smith, D. and Birkin, M. 2012. Creating realistic synthetic populations at varying spatial scales: a comparative critique of population synthesis techniques. *Journal of Artificial Societies and Social Simulation*. **15**(1), pp.1-15.
- Heinen, E. van Wee, B. and Maat, K. 2010. Commuting by bicycle: an overview of the literature. *Transport reviews*. **30**(1), pp.59-96.
- Heinen, E. Maat, K. and van Wee, B. 2013. The effect of work-related factors on the bicycle commute mode choice in the Netherlands. *Transportation*. **40**(1), pp.23-43.
- Iacono, M. Krizek, K.J. and El-Geneidy, A. 2010. Measuring non-motorized accessibility: issues, alternatives, and execution. *Journal of Transport Geography*. **18**(1), pp.133-140.

- Jones, T. 2012. Getting the British back on bicycles—The effects of urban traffic-free paths on everyday cycling. *Transport Policy*. **20**, pp.138-149.
- Kaparias, I.Bell, M.G.Miri, A.Chan, C. and Mount, B. 2012. Analysing the perceptions of pedestrians and drivers to shared space. *Transportation research part F: traffic psychology and behaviour*. **15**(3), pp.297-310.
- Keep, M. 2013. *Road cycling: statistics*. [Online]. London: House of Commons Library. SN06224. [Accessed 25 August 2014]. Available from: <http://www.parliament.uk/>
- Larsen, J. and El-Geneidy, A. 2011. A travel behavior analysis of urban cycling facilities in Montréal, Canada. *Transportation research part D: transport and environment*. **16**(2), pp.172-177.
- Larsen, J.Patterson, Z. and El-Geneidy, A. 2013. Build it. But where? The use of geographic information systems in identifying locations for new cycling infrastructure. *International Journal of Sustainable Transportation*. **7**(4), pp.299-317.
- Lawson, A.R. McMorrow, K. and Ghosh, B. 2013. Analysis of the non-motorized commuter journeys in major Irish cities. *Transport Policy*. **27**, pp.179-188.
- Lovelace, R.Beck, S.Watson, M. and Wild, A. 2011. Assessing the energy implications of replacing car trips with bicycle trips in Sheffield, UK. *Energy Policy*. **39**(4), pp.2075-2087.
- Lovelace, R.Ballas, D. and Watson, M. 2014. A spatial microsimulation approach for the analysis of commuter patterns: from individual to regional levels. *Journal of Transport Geography*. **34**, pp.282-296.
- Menghini, G.Carrasco, N.Schüssler, N. and Axhausen, K.W. 2010. Route choice of cyclists in Zurich. *Transportation research part A: policy and practice*. **44**(9), pp.754-765.
- Monsere, C.Dill, J.McNeil, N.Clifton, K.Foster, N.Goddard, T.Berkow, M.Gilpin, J.Voros, K. and van Hengel, D. 2014. *Lessons from the Green Lanes: Evaluating Protected Bike Lanes in the US*. [Online]. Oregon Transportation Research and Education Consortium. [Accessed 25 August 2014]. Available from: <http://www.otrec.us/>
- Moody, S. and Melia, S. 2013. Shared space: Research, policy and problems. In: *Proceedings of the Institution of Civil Engineers-Transport*: ICE.
- Nelson, A.C. and Allen, D. 1997. If you build them, commuters will use them: association between bicycle facilities and bicycle commuting. *Transportation Research Record: Journal of the Transportation Research Board*. **1578**(1), pp.79-83.
- ONS. 2014a. *NOMISWEB official labour market statistics*. [Online]. Office for National Statistics. [Accessed 25 August 2014]. Available from: <http://www.nomisweb.co.uk/>
- ONS. 2014b. *2011 Census Analysis - Cycling to Work*. [Online]. Office for National Statistics. [Accessed 25 August 2014]. Available from: <http://www.ons.gov.uk/>
- ONS. 2014c. *Statistical disclosure control for 2011 Census*. [Online]. Office for National Statistics. [Accessed 25 August 2014]. Available from: <http://www.ons.gov.uk/>
- ONS. 2014d. *Open Geography Portal*. [Online]. Office for National Statistics. [Accessed 25 August 2014]. Available from: <https://geoportal.statistics.gov.uk/geoportal/>
- ONS. 2014e. *Neighbourhood Statistics*. [Online]. Office for National Statistics. [Accessed 25 August 2014]. Available from: <http://www.neighbourhood.statistics.gov.uk/>
- OpenStreetMap. 2014. *OpenStreetMap*. [Online]. [Accessed 25 August 2014]. Available from: <http://www.openstreetmap.org/>
- Parkin, J. 2004. *Determination and measurement of factors which influence propensity to cycle to work*. Thesis, University of Leeds.
- Parkin, J.Wardman, M. and Page, M. 2007a. Models of perceived cycling risk and route acceptability. *Accident Analysis & Prevention*. **39**(2), pp.364-371.
- Parkin, J.Ryley, T. and Jones, T. 2007b. Barriers to cycling: an exploration of quantitative analyses. *Cycling and society*. pp.67-82.
- Parkin, J.Wardman, M. and Page, M. 2008. Estimation of the determinants of bicycle mode share for the

- journey to work using census data. *Transportation*. **35**(1), pp.93-109.
- Parkin, J. and Meyers, C. 2010. The effect of cycle lanes on the proximity between motor traffic and cycle traffic. *Accident Analysis & Prevention*. **42**(1), pp.159-165.
- Parkin, J. and Rotheram, J. 2010. Design speeds and acceleration characteristics of bicycle traffic for use in planning, design and appraisal. *Transport Policy*. **17**(5), pp.335-341.
- Philips, I., Watling, D., Timms, P., 2013. A conceptual approach for estimating resilience to fuel shocks, in: *Selected Proceedings. Presented at the 13th World Conference on Transportation Research, Rio, Brazil*.
- Pooley, C.Tight, M.Jones, T.Horton, D.Scheldeman, G.Jopson, A.Mullen, C.Chisholm, A.Strano, E. and Constantine, S. 2011. Understanding walking and cycling: Summary of key findings and recommendations. [Online]. Understanding Walking and Cycling Project. Lancaster University. [Accessed 25 August 2014]. Available from: <http://eprints.lancs.ac.uk/>
- Pooley, C.G.Horton, D.Scheldeman, G.Mullen, C.Jones, T.Tight, M.Jopson, A. and Chisholm, A. 2013. Policies for promoting walking and cycling in England: A view from the street. *Transport Policy*. **27**, pp.66-72.
- Prato, C.G. 2009. Route choice modeling: past, present and future research directions. *Journal of Choice Modelling*. **2**(1), pp.65-100.
- Pucher, J.Dill, J. and Handy, S. 2010. Infrastructure, programs, and policies to increase bicycling: an international review. *Preventive medicine*. **50**, pp.S106-S125.
- Rietveld, P. and Daniel, V. 2004. Determinants of bicycle use: do municipal policies matter? *Transportation Research Part A: Policy and Practice*. **38**(7), pp.531-550.
- SAP. 2010. Greater Bristol Cycling Strategy 2011-2026. [Online]. Greater Bristol Cycling City Stakeholder Advisory Panel. [Accessed 25 August 2014]. Available from: <http://www.betterbybike.info/>
- Schoner, J.E. 2013. *Catalysts and Magnets: built environment effects on bicycle commuting*. [Online]. Thesis, University of Minnesota. [Accessed 25 August 2014]. Available from: <http://nexus.umn.edu/>
- Schoner, J.E. and Levinson, D.M. 2013. The Missing Link: Bicycle Infrastructure Networks and Ridership in 74 US Cities. In: *Transportation Research Board 92nd Annual Meeting*.
- Sherali, H.D.Sivanandan, R. and Hobeika, A.G. 1994. A linear programming approach for synthesizing origin-destination trip tables from link traffic volumes. *Transportation Research Part B: Methodological*. **28**(3), pp.213-233.
- Stinson, M.A. and Bhat, C.R. 2003. Commuter bicyclist route choice: Analysis using a stated preference survey. *Transportation Research Record: Journal of the Transportation Research Board*. **1828**(1), pp.107-115.
- Tobler, W.R. 1970. A computer movie simulating urban growth in the Detroit region. *Economic geography*. pp.234-240.
- UK Data Service. 2014. *Experian Demographic Data, 2004-2005 and 2008-2011*. [Online]. [Accessed 25 August 2014]. Available from: <http://discover.ukdataservice.ac.uk/>
- UK Government. 2014. *National Travel Survey (NTS)*. [Online]. [Accessed 25 August 2014]. Available from: <https://www.gov.uk/>
- Vandenbulcke, G.Dujardin, C.Thomas, I.Geus, B.d.Degraeuwe, B.Meeusen, R. and Panis, L.I. 2011. Cycle commuting in Belgium: spatial determinants and 're-cycling' strategies. *Transportation research part A: policy and practice*. **45**(2), pp.118-137.
- Wardman, M.Tight, M. and Page, M. 2007. Factors influencing the propensity to cycle to work. *Transportation Research Part A: Policy and Practice*. **41**(4), pp.339-350.
- West of England Partnership. 2011. Joint Local Transport Plan 3 (JLTP3). [Online]. [Accessed 25 August 2014]. Available from: <http://www.travelplus.org.uk/>
- Yeboah, G. and Alvanides, S. 2013. Everyday cycling in urban environments: Understanding behaviours and constraints in space-time. In: *21st GIS Research UK (GISRUK), Liverpool University, UK*.

Appendix I: Glossary of Acronyms

UA – Unitary Authorities

OA – Output Area

LA – Local Authority

MSOA – Middle layer Super Output Area

LSOA – Lower layer Super Output Area

WZ – Workplace Zone

PWC – Population-Weighted Centroids

BUA – Built Up Areas

ONS – Office for National Statistics

NTS – National Travel Survey

ID – Indices of Deprivation

DfT – (UK Government) Department for Transport

ABM – Agent Based Model

Appendix II: GenSynthFlow (Java software)

GenSynthFlow is a Java implementation of the origin-destination flow synthesis algorithm described in section 5.4 on page 31 of the Methodology. The UML class diagram in Figure 23 gives an overview of each of the classes (one per source file), the data objects and methods they consist of, and how the classes are interdependent. The basic operation is as follows:

- The top level ‘main()’ method (residing in class **GenSynthFlow**) reads in CSV-format census data files describing the commuter flow counts and coordinates for each commuting origin OA and commuting destination WZ.
- Each OA and each WZ will have its own object (of class **ZoneFlow**) to store its flow count and coordinates.
- These are grouped into lists *oaFlowList* and *wzFlowList* (of class **ZoneFlowList**).
- ‘main()’ then creates a list for each WZ destination of all the candidate OA origins that are within the currently selected range of it (e.g. 2-5km) and stores these in *candOAsForAllWZs* (of class **CandOAsPerWZ**).
- It then creates an initial synthetic population (via the constructor of class **SynthPop**).
 - ‘SynthPop()’ creates an array of objects (of class **Worker**) representing the commuting population. The array is sized to equal the total of all the WZ flow counts.
 - The WZs of each worker are fixed (for eternity) at this point.
 - Random OAs are assigned to each worker, based on those found in the list of possible candidates (*candOAsForAllWZs*).
- ‘main()’ then iterates the population by calling method ‘iterSynthFlowDist()’ of class **SynthPop** to juggle OAs assigned to each Worker to reduce mismatches. ‘iterSynthFlowDist()’ keeps track of the current total flow counts for each origin OA in array *AreaCount* (within class **AreaWorkingCounts**).
- Finally ‘main()’ converts the population to an origin-destination flow table by calling the ‘popToFlowMatrix()’ method (within class **SynthPop**) and writes the data out to a CSV file (via class **CSVWriter**).

The complete source code and detailed JavaDoc documentation is viewable online at:

<http://richard-thomas.github.io/GenSynthFlow/>

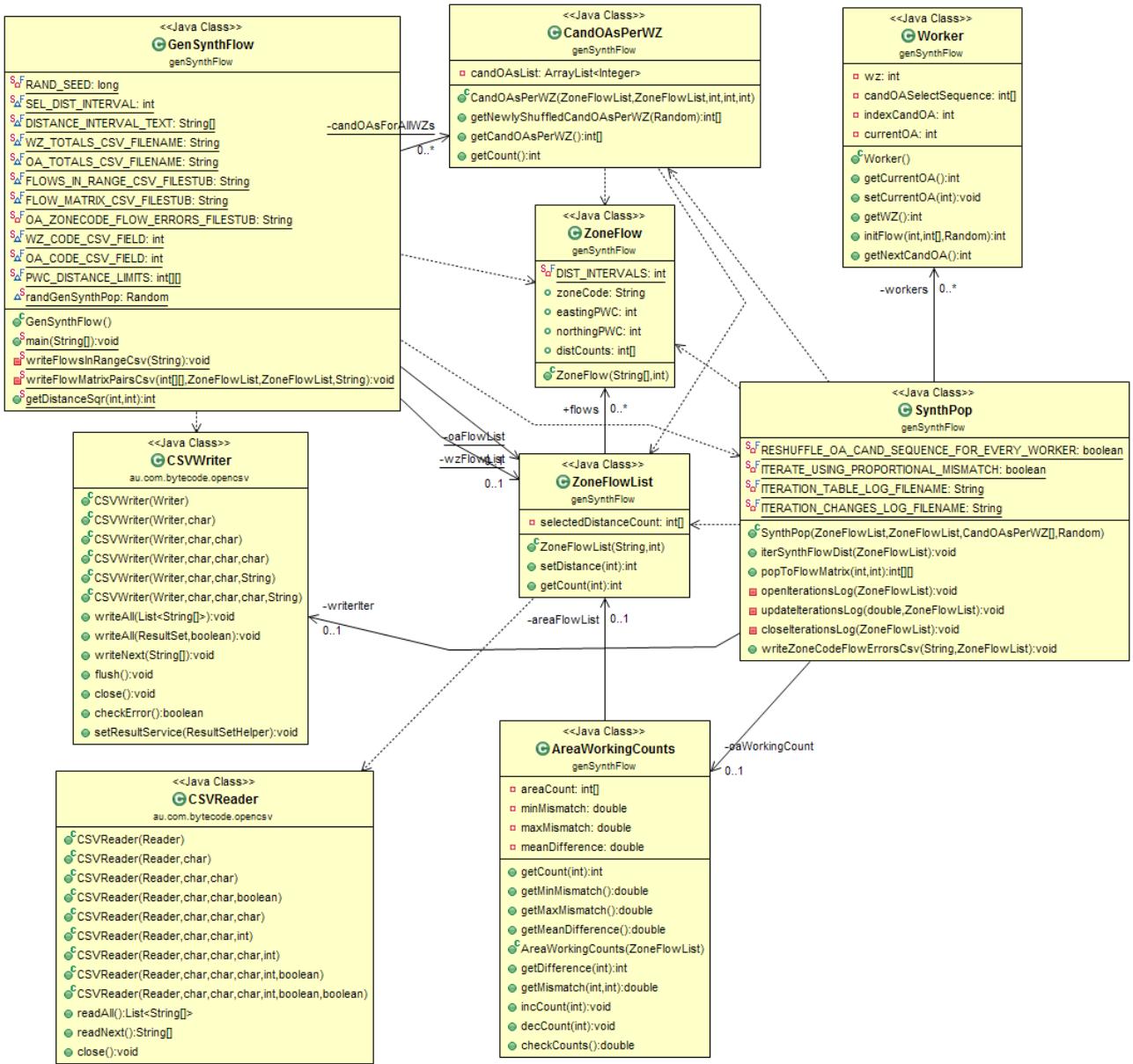


Figure 23: UML Class Diagram for GenSynthFlow Java software

Appendix III: Correlations between 'Independent' Variables

The following tables detail all the (often unexpected) correlations between some of the dependent variables used in the 3 linear regressions in the study. (More obviously correlated variables like median distance and average journey time have been omitted). Mixing such correlated variables was generally avoided as can indicate collinearity which might ill-condition the regression. Correlations greater than 0.8 can be problematic, though ones greater than 0.7 have also been highlighted as ones to be wary of.

Table 16: MSOA Route-based Correlations between Nominally Independent Variables

		EffortRatio	Traffic	SpeedKmh
EffortRatio	Pearson Correlation	1	.700	<u>-.903</u>
	Sig. (2-tailed)		.000	.000
Traffic	Pearson Correlation	.700	1	<u>-.703</u>
	Sig. (2-tailed)	.000		.000
SpeedKmh	Pearson Correlation	-.903	-.703	1
	Sig. (2-tailed)	.000	.000	
N		6887	6887	6887

Table 17: Notable MSOA Area-based Correlations between Nominally Independent Variables

		MeanEuc DstKm	AvSpeed Kmh	AvEffort Ratio	AvTraffic	2-5km (%)	5-10km (%)	10-20km (%)
MeanEucDstKm	Pearson Correlation	1	.362	-.237	-.325	-.412	<u>.817</u>	<u>.756</u>
	Sig. (2-tailed)		.002	.044	.005	.000	.000	.000
AvSpeedKmh	Pearson Correlation	.362	1	<u>-.936</u>	<u>-.849</u>	-.140	.517	.081
	Sig. (2-tailed)	.002		.000	.000	.239	.000	.495
AvEffortRatio	Pearson Correlation	-.237	-.936	1	<u>.853</u>	.044	-.328	-.091
	Sig. (2-tailed)	.044	.000		.000	.713	.005	.444
AvTraffic	Pearson Correlation	-.325	-.849	.853	1	.166	-.390	-.187
	Sig. (2-tailed)	.005	.000	.000		.160	.001	.113
2-5km (%)	Pearson Correlation	-.412	-.140	.044	.166	1	-.528	-.323
	Sig. (2-tailed)	.000	.239	.713	.160		.000	.005
5-10km (%)	Pearson Correlation	.817	.517	-.328	-.390	-.528	1	.353
	Sig. (2-tailed)	.000	.000	.005	.001	.000		.002
10-20km (%)	Pearson Correlation	.756	.081	-.091	-.187	-.323	.353	1
	Sig. (2-tailed)	.000	.495	.444	.113	.005	.002	
N		73	73	73	73	73	73	73

Table 18: Notable OA Area-based Correlations between Nominally Independent Variables

		MeanEuc DstKm	AvSpeed Kmh	AvEffort Ratio	0-2km (%)	5-10km (%)	10-20km (%)
MeanEucDstKm	Pearson Correlation	1	.439	-.426	-.710	.712	.795
	Sig. (2-tailed)		.000	.000	.000	.000	.000
AvSpeedKmh	Pearson Correlation	.439	1	-.882	-.380	.558	.145
	Sig. (2-tailed)	.000		.000	.000	.000	.000
AvEffortRatio	Pearson Correlation	-.426	-.882	1	.374	-.516	-.154
	Sig. (2-tailed)	.000	.000		.000	.000	.000
0-2km (%)	Pearson Correlation	-.710	-.380	.374	1	-.512	-.353
	Sig. (2-tailed)	.000	.000	.000		.000	.000
5-10km (%)	Pearson Correlation	.712	.558	-.516	-.512	1	.261
	Sig. (2-tailed)	.000	.000	.000	.000		.000
10-20km (%)	Pearson Correlation	.795	.145	-.154	-.353	.261	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	
N		1951	1951	1951	1951	1951	1951