Modelling the school commute for 7.5m students over 4 years using data from the school census

Nick Bearman[1] & Alex Singleton[1]

[1]Department of Geography and Planning, School of Environmental Sciences, University of Liverpool, Gordon Stephenson Building, 74 Bedford Street South, Liverpool, L69 7ZQ

Email: n.bearman@liverpool.ac.uk

KEYWORDS: transport, routing, schools, CO_2, R, pgRouting, Routino

1. Introduction

Many studies have shown that rates of active transport (e.g. cycling or walking) to school are decreasing (McMillan, 2007; Trang et al., 2012). Impacts of this include negative impacts on the environment from increased emissions (van Ristell et al., 2013), increasing traffic congestion around schools (Collins and Kearns, 2001), health impacts as a result of the lower level of physical activity (Faulkner et al., 2009) and greater exposure to pollutants (McConnell et al., 2010). Additionally it is estimated that home to school travel in England contributes 658k tonnes of carbon each year (DCFS, 2010) and therefore there are potentially substantial environmental benefits if pupils adopt more sustainable travel behaviours.

There are many factors that impact home to school transport mode choice including direct factors such as actual and perceived distance to school (Ewing et al., 2004; Lang et al., 2011), road infrastructure and urban form (Panter et al., 2010), and indirect factors such as ethnicity (McDonald, 2008), socio-economic status (Wilson et al., 2010) and lifestyle factors (Babey et al., 2009). Additionally it is argued that political changes promoting parental choice in school selection results in longer commuting distances, which have a related impact on mode choice and emission levels (Marshall et al., 2010; van Ristell et al., 2013).

A number of studies have evaluated home to school travel and the possible causes of the variation seen (Marshall et al., 2010; van Ristell et al., 2013; Wilson et al., 2007). Only limited studies to date have utilised a near complete data set for a country-wide analysis, with van Ristell et al. (2012) examining this at a national scale (in a US context) using a sample. Van Ristell et al. (2012) used straight line distances between home and school to model CO_2 emissions. However straight line distances underestimate the distance travelled and this limits the accuracy of any CO_2 measures derived from the distances. While calculating the shortest or quickest path along the road or rail network provides a more accurate estimate, it is much more computationally intensive to calculate. Processing time is a particular issue for very large samples or total population surveys and is one of the main areas addressed in this work.

Singleton (2014) used a nationwide data set of pupil locations and modes of travel to calculate the length and CO_2 emissions for all pupil commuting routes from 2011 in England. This work builds on Singleton's research, initially expanding the coverage to four years (2008 to 2011), increasing the efficiency of the code and pupil linkage mechanism.

2. Methodology

The Department of Education performs a school census each term collecting a variety of information on the school and its pupils. For the years between 2007 and 2011, pupils were asked to provide their usual mode of travel to/from school such as walking, cycling, car, bus, tram or rail. The school census data includes the pupil postcode, so when combined with the school location information and mode of travel this allows us to model the pupil's route from home to school. The school census contains 5 years worth of travel mode data (2007 to 2011), however the data for 2007 had a large proportion of missing values for mode of travel (31.2%) so this year was excluded from the analysis.

Data for each pupil from each year can be joined using the anonymous pupil reference number provided with each record. This allows us to evaluate the change in pupil travel modes and distances over the period of 4 years. This will include transitions from Primary to Secondary school, a change which has a major impact on pupil travelling behaviour (van Ristell et al., 2013). This is a significant development on the approach used in Singleton (2014) which only looked at the aggregate choice between years for a single cohort.

The model used for this work is primarily written in R[1]. The calculation of the routes has been split into two separate sections, one for road routing (for walking, cycling, car and bus transport) and one for non-road routing (for rail and tram transport). The road routing makes use of a software library called Routino[2], which uses OpenStreetMap data[3] to calculate road based routes. Routino provides a number of different templates for walking, cycling and car based routing. The non-road routing used a PostGIS pgRouting[4] enabled database to calculate the shortest route for each rail or tram journey. For each journey, the closest station was calculated to the home postcode and this station was selected as the start point. If the distance from home to station was unusually large, the journey was excluded (typically ~1% of cases). The threshold was calculated from the Tukey outlier (Tukey, 1977) of the journey distances and set at 2.1km for tube and 2.8km for tram. Rail outliers were not checked in this way because there is much more uniform rail coverage across the country.

The length of each journey is recorded in the data set, with journeys removed if they are missing mode data, have the same home and school station or are unroutable for another reason (8.3%). Also, any journey outliers were removed using a modified Turkey outlier (13.5%), shown below. The pupil commute (*Kp*) was compared with the quartile (*Q*) and inter-quartile range (*IQR*) for all distances aggregated by year group (*a*), local authority (*g*) and mode type (*m*). This was necessary to account for geographic differences in transport infrastructure, different typical travelling distances in each transport type, and differences in travel behaviour for different age groups.

$$\left[\left(Q1_{mgq} - 1.5IQR_{mga}\right) < K_p < \left(Q3_{mga} + 1.5IQR_{mga}\right)\right]$$

[1] http://www.r-project.org/
[2] http://www.routino.org/
[3] http://www.openstreetmap.org/
[4] http://pgrouting.org/

3. Results

For this pilot a 10% subset of the 7.5 million record data set covering 25 local authorities over a range of urban and rural settings was included. A total of 815,676 records were processed, with 67,772 (8.3%) unable to be routed and 110,247 (13.5%) outside the Tukey outlier range. This left a total of 637,657 routes, 78.2% of the original total.

Initial analysis of mode choice and distance showed consistency with previous results (Singleton, 2014). As shown in figure 1 (below), non motorised travel (walking and cycling, blue) is the most common mode up to 2.5km, showing a negative exponential decay with increasing distance. Car (green) is the most common from 2.5km to 4.5km, with bus (red) being the most common for distances over 4.5km. Train (& tram) has increasing importance for longer distances (purple), but the absolute number of journeys is low compared to the other methods.

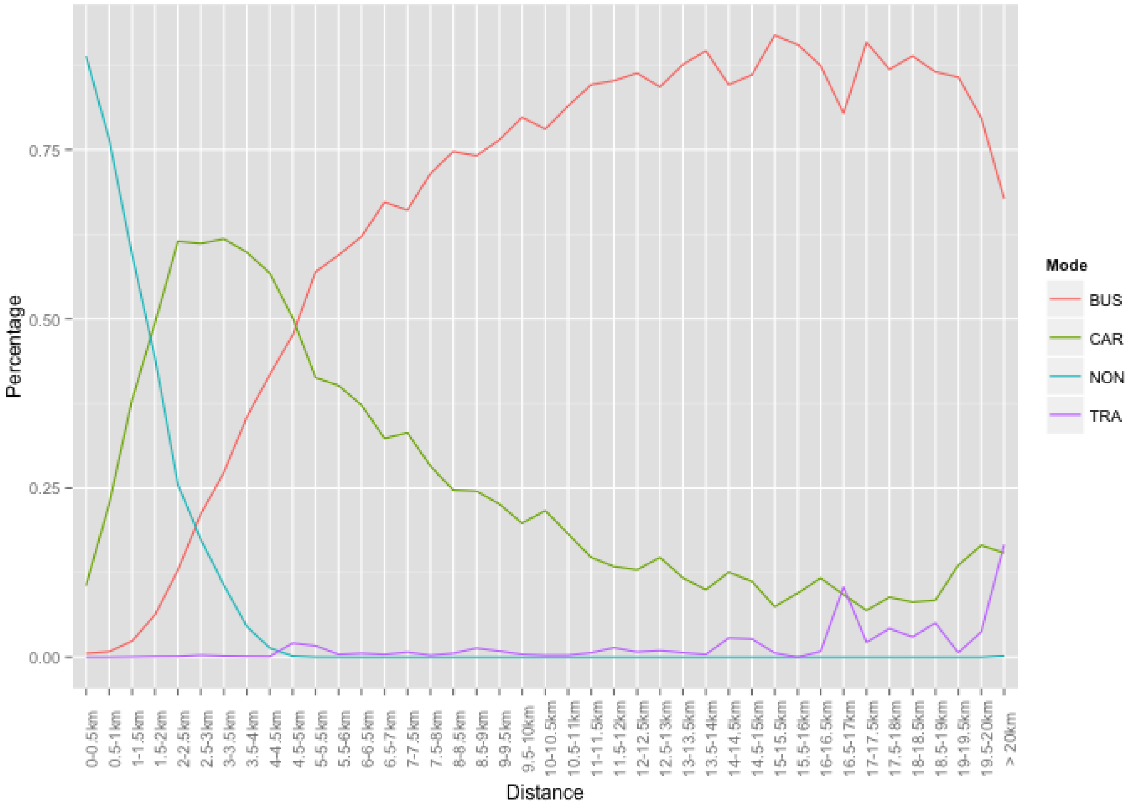

Figure 1. Proportion of mode choice by distance for 2008-2011.

Subsequent analysis of mode choice by school year shows a substantial difference in change of mode choice between primary (years 1-6) and secondary (years 7-11) schools (figure 2, over). It is well known that pupils travel further to secondary school than primary school (van Ristell et al., 2013) and this is reflected in the mode choice. In both, non-motorised is the most common mode (blue), but for primary years this swaps with car (green) at around 1.25km, where as for secondary walking is most common up until 3km. Additionally, buses (red) are rarely used by primary pupils, where as they are much more important for secondary pupils, similar to car for journeys up to 3km, and more common than car for journeys above 3km.

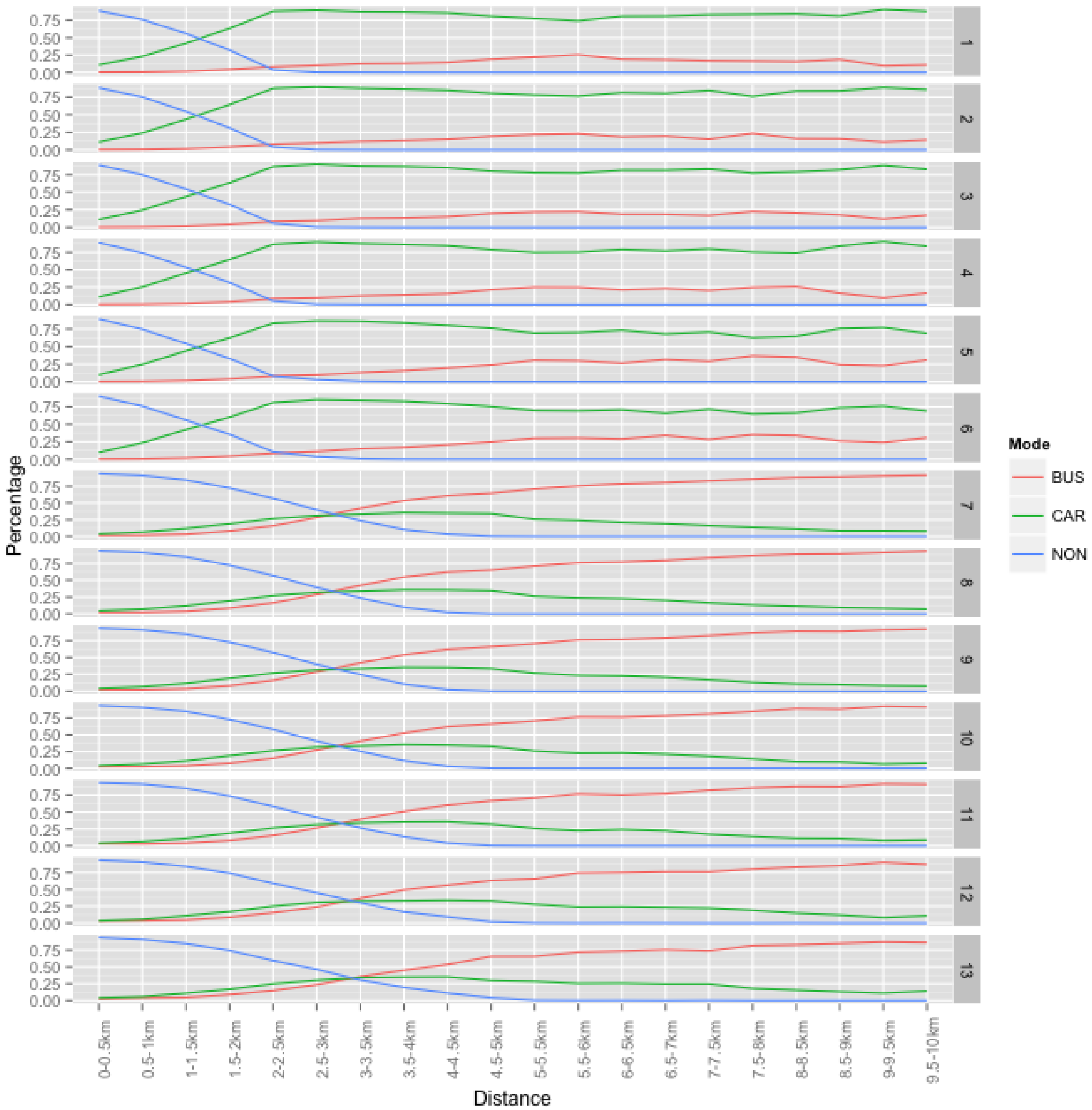

Figure 2. Mode type (bus, car or non-motorised) over distance, split by school year (right). Differences can be seen between primary (years 1 – 6) and secondary (years 7 – 13).

CO_2 emissions were calculated from the distances of the routes for 2008 and 2011. CO_2 rates for cars were based on the average CO_2 levels emitted from cars registered with the DVLA in each LSOA. This highlights the geographical variation in CO_2 emissions, which would be missed if a national average was used. Regional values were used for tram and tube transport, with a 'London' and 'non-London' value used for buses because of the difference between the two geographical areas.

Initial calculations showed a wide variation of values (see table 1, over). Of the 25 Local Authorities (LAs) included in this subset, Birmingham had the highest absolute value (~84 tonnes) for 2011, followed by Sheffield (~18t) and Wiltshire (~12t). When comparing the data from 2008, Birmingham has increased by 33%, Sheffield by 58% and Wiltshire by 82%. Absolute values may be of interest from a policy perspective, when trying to evaluate the impact of authority wide changes. However, mean values (number of journeys in each LA divided by total carbon in LA) provide more information about the characteristics of the local area. For example, Birmingham's mean value changed from 324g CO_2 in 2008 to 481g CO_2 in 2011. These change values indicate some of the comparisons that can be done using this data set. With this subset, the amount of information that can be gained from this data is limited, because changes in pupil numbers and missing routing data are not taken into account.

Table 1. CO_2 emissions for the 25 local authorities in this pilot. The table contains both the total amount of CO_2 emitted as a result of all home to school journeys (left) and the average CO_2 emitted per journey (right) for 2008, 2011, and the percentage change between the two years.

Local Authority	**Total (tonnes of Carbon)**			**Average g CO_2 per journey**		
	2008	2011	% change	2008	2011	% change
Birmingham	56,888,410	84,515,510	33%	324	481	33%
Bradford	10,764,392	20,297,180	47%	290	375	23%
Breckland	8,003,222	5,141,964	-56%	498	562	11%
Bristol	2,991,850	10,745,580	72%	280	380	26%
Broadland	6,736,280	3,659,490	-84%	373	396	6%
Chester & West Chester	4,893,318	12,314,552	60%	412	471	13%
Cornwall	17,632,652	18,354,768	4%	450	467	4%
East Devon	3,356,114	5,654,832	41%	511	522	2%
Great Yarmouth	2,573,174	3,364,868	24%	326	430	24%
King's Lynn & West Norfolk	3,059,028	5,522,470	45%	470	526	11%
Kingston upon Hull	23,714,196	4,032,334	-488%	176	238	26%
Kirklees	5,513,442	15,463,018	64%	342	440	22%
Leeds	3,454,852	22,540,054	85%	333	423	21%
Liverpool	7,879,004	22,695,366	65%	452	649	30%
Manchester	2,544,722	13,427,564	81%	308	459	33%
Mid Devon	6,060,548	3,610,908	-68%	508	545	7%
North Devon	4,612,876	3,388,308	-36%	418	463	10%
North Norfolk	9,008,934	5,844,532	-54%	714	951	25%
Northumberland	12,037,528	9,719,592	-24%	316	385	18%
Sheffield	15,748,004	35,600,762	56%	321	481	33%
Solihull	15,783,410	6,946,472	-127%	375	387	3%
South Gloucestershire	17,725,880	7,865,480	-125%	264	342	23%
South Norfolk	4,428,530	5,343,398	17%	576	663	13%
West Devon	20,401,360	3,123,140	-553%	595	721	18%
Wiltshire	4,519,894	25,631,636	82%	571	715	20%

Figure 3 (below) shows the average CO_2 emissions per journey for primary and secondary school journeys by home LSOA for Norfolk. A number of patterns can be seen, with typically higher CO_2 emissions in rural areas and lower CO_2 emissions in urban areas.

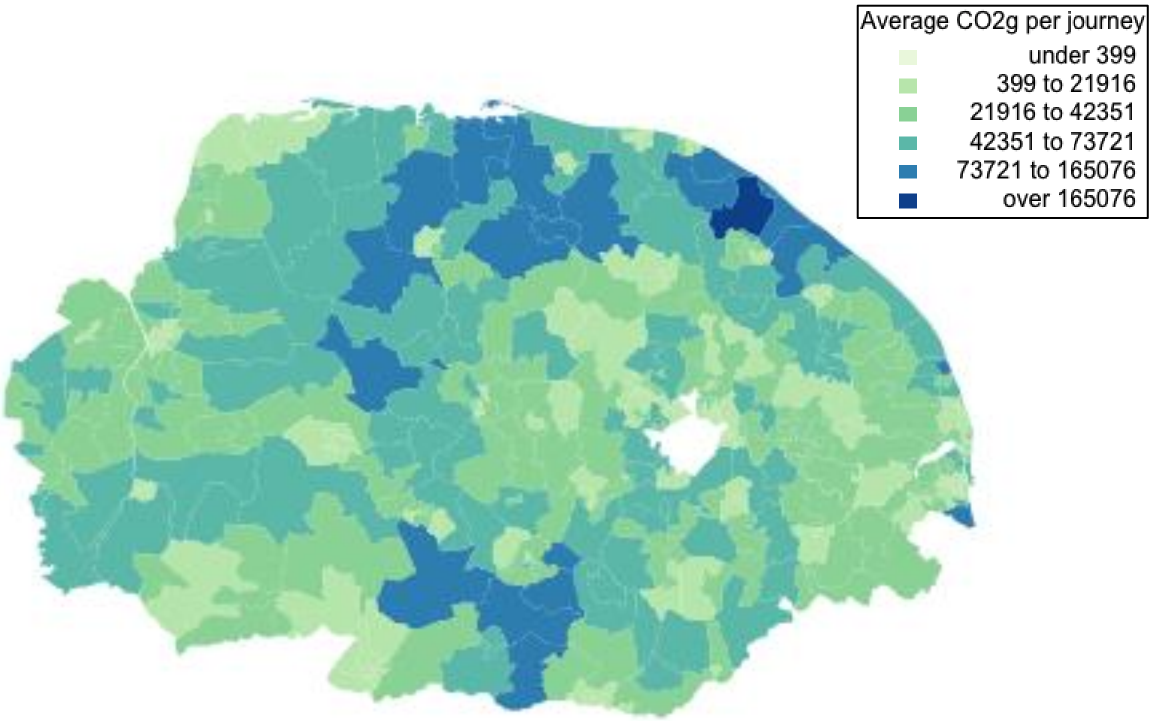

Figure 3. Map of LSOAs in Norfolk (excluding Norwich), showing average CO_2 per journey to primary and secondary school by home LSOA, 2011.

4. Discussions and Conclusions

Further results from this work will be calculated before the conference, with the results presented for the whole data set. This work has considered how home to school travel has changed over the four years 2008-2011, however there are a number of other areas where this work could be developed.

Firstly, the processing power required to calculate each of the 7.5 million routes for each year is substantial. Efficiency improvements have been made over the original coding, but processing time is still a limiting factor in this work. Therefore this work will explore the potential of using cloud computing services to enable faster processing of the data. There are a number of ethical and legal requirements to be considered before this approach can be used, given the nature of the school census dataset.

Additionally, as this work is modelling all of the routes taken by school children, it is possible to look at the total impact of the resultant traffic on the peak time road activity. For example, we can calculate how many cars on the road are there as a result of the school commute and we can also calculate the total CO_2 emissions on each segment of road along a specified route. We can also use this information to evaluate the impact of the school commute on those using different transport methods along the same routes.

References

Babey, S.H., Hastert, T.A., Huang, W., and Brown, E.R. (2009). Sociodemographic, Family, and Environmental Factors Associated with Active Commuting to School among US Adolescents. J Public Health Pol *30*, S203–S220.

Collins, D.C., and Kearns, R.A. (2001). The safe journeys of an enterprising school: negotiating landscapes of opportunity and risk. Health & Place *7*, 293–306.

DCSF (2010). Road to zero carbon, Department for Children, Schools and Families: London.

Ewing, R., Schroeer, W., and Greene, W. (2004). School Location and Student Travel Analysis of Factors Affecting Mode Choice. *1*, 55–63.

Faulkner, G.E.J., Buliung, R.N., Flora, P.K., and Fusco, C. (2009). Active school transport, physical activity levels and body weight of children and youth: A systematic review. Preventive Medicine *48*, 3–8.

Lang, D., Collins, D., and Kearns, R. (2011). Understanding modal choice for the trip to school. Journal of Transport Geography *19*, 509–514.

Marshall, J.D., Wilson, R.D., Meyer, K.L., Rajangam, S.K., McDonald, N.C., and Wilson, E.J. (2010). Vehicle Emissions during Children's School Commuting: Impacts of Education Policy. Environ. Sci. Technol. *44*, 1537–1543.

McConnell, R., Liu, F., Wu, J., Lurmann, F., Peters, J., and Berhane, K. (2010). Asthma and School Commuting Time. Journal of Occupational *52*, 827–828.

McDonald, N.C. (2008). Children's mode choice for the school trip: the role of distance and school location in walking to school. Transportation *35*, 23–35.

McMillan, T.E. (2007). The relative influence of urban form on a child's travel mode to school. Transportation Research Part A: Policy and Practice *41*, 69–79.

Panter, J.R., Jones, A.P., Van Sluijs, E.M.F., and Griffin, S.J. (2010). Neighborhood, Route, and School Environments and Children's Active Commuting. American Journal of Preventive Medicine *38*, 268–278.

Singleton, A. (2014). A GIS approach to modelling CO2 emissions associated with the pupil-school commute. International Journal of Geographical Information Science *28*, 256–273.

Trang, N.H.H.D., Hong, T.K., and Dibley, M.J. (2012). Active Commuting to School Among Adolescents in Ho Chi Minh City, Vietnam: Change and Predictors in a Longitudinal Study, 2004 to 2009. American Journal of Preventive Medicine *42*, 120–128.

Tukey, J.W. (1977). Exploratory data analysis (Reading, Mass.: Addison-Wesley Pub. Co.).

van Ristell, J., Quddus, M., Enoch, M., Wang, C., and Hardy, P. (2013). Quantifying the transport-related impacts of parental school choice in England. Transportation *40*, 69–90.

Wilson, E.J., Wilson, R., and Krizek, K.J. (2007). The implications of school choice on travel behavior and environmental emissions. Transportation Research Part D: Transport and Environment *12*, 506–518.

Wilson, E.J., Marshall, J., Wilson, R., and Krizek, K.J. (2010). By foot, bus or car: children's school travel and school choice policy. Environment and Planning A *42*, 2168–2185.

Biography

Nick Bearman is a Research Associate at the University of Liverpool and has previously worked at the University of Exeter Medical School after completing his PhD at UEA. He is working the ESRC funded project "Leveraging the Google cloud to estimate individual level CO_2 emissions linked to the school commute". This work uses open source software to estimate the routes and CO_2 emissions of every school child in England between 2008 and 2011. He is interested in the use of GIS to solve novel problems, often with the use of customised design and programming to create solutions. His previous work has included sonification and epidemiology applications working with a wide variety of academics.

Alex Singleton is a Reader in Geographic Information Science; his research concerns how the social and spatial complexities of individual behaviour can be represented and understood within a framework of quantitative social science and computer modelling. In particular, his research extends from a geographic tradition of area classification and has developed a broad critique of the ways in which geodemographic methods can be refined through modern scientific approaches to data mining, geographic information science and quantitative human geography.