**Methods for using RTS and census commute data to estimate travel distributions by LA and by road type**

***Part 1: Estimate the volume of traffic on motorways from RTS***

We used annual average daily flow (AADF) estimates from RTS[[1]](#footnote-1) plus link lengths from RTS to estimate total annual travel on motorways. The roads are described in terms of links, and each link has an estimated AADF (which might be a count of one day, and might be last year’s count augmented by an ‘expansion factor’ based on other counts). We assigned road type by the last letter of the “road\_category” column.[[2]](#footnote-2) As each link has a link length (LL) and an AADF, we estimate total travel as 365\*LL\*AADF. These estimates coincide with those provided for motorway travel by RTS.

*- NB there are a lot of count point ids in the AADF dataset that don’t have a corresponding entry (count point id) in the link lengths dataset. E.g the M10 and the M41 don’t have lengths but they do have counts. Missing count points include about 100 A-road links from the Bristol city region in the years 2010 to 2015, for which we reached the same total as RTS. This amounts to 4.3% of total counts for cars\_and\_taxis, 2.7% for cycles, 4% for motorbikes, 4.1% for buses, 4% for vans, 2.5% for HGVs. Considering A roads alone, it’s 6.5% for cars, 6% for HGVs, 5.8% for vans, 4.9% for motorbikes, 4.8% for buses, and 2.7% for cycles. At present they are excluded from the A/M totals, which seems to match what RTS does. [To follow up with RTS…? Or just delete this]*

***Part 2: route Census data through graphhopper***

***Select commute origin-destination pairs from Census 2011***

* Take a random sample of 5k commuters per LA, selecting 1k each from foot; bicycle; car (driver or passenger); motorbike; or bus. Limit selection to those who have a fixed workplace. If fewer than 1k commuters using the mod in question are available, then just use the full sample. This gives 1.4 million commuters selected out of a total of 23 million.
* Use the LSOA:LSOA OD dataset, as this is the finest grain resolution dataset that also gives mode of travel.
* Use the postcode centroid for each start point, and the population-weighted centroid of the LSOA for the end point.
* Exclude OD-pairs which have Euclidean distance above the 98% percentile for commute trips by that mode observed in the National Travel Survey. This corresponds to a distance of 19.3 km for cycling, 4.8km for walking, 72.4km for car, 59.5km for motorbike and 27.4km for bus. These excluded trips are assumed mostly to reflect a person commuting from a different address (e.g. from a second home near their work).

***Route origin-destination pairs***

* Use graphhopper[[3]](#footnote-3) to route separately for walking, cycling, car, motorbike, bus. The routing decisions is made to provide the "fastest" option; graph hopper does not have the same option as cyclestreets to choose ‘quieter’ routes, although fastest option for bikes does also include “a complicated mix of some heuristics to also provide a rather safe route.”
* Graphhopper routes by mode. For bicycle use the option ‘bike’. For car/motorbike, use the option ‘car’.[[4]](#footnote-4) For bus use the mode ‘truck’, as the closest mode option available.[[5]](#footnote-5)
* Graphhopper provides information on road type, drawing on OpenStreetMap. Roads are divided into ‘motorway’ and ‘other’.

***Part 3: Process the graphhopper estimates to estimate travel by road type for all trips, and use RTS to scale relative distances between motorways and other roads***

1. For each home local authority, for each mode, calculate the distance in each English local authority. Sum these distances across all home local authorities to get the proportion of total distance in each local authority.
2. For all travel in a given local authority (done by people living anywhere in England), for each mode, calculate the distance on motorway versus other roads.
3. Compare the proportion of motorway travel generated in Step 2 above with the proportions derived from RTS counts (see Part 1). This RTS data used pooled data across 2010-2015 inclusive. Generate odds ratios scaling factors to scale the ‘motorway’ and ‘other’ road figures accordingly. Do this step at the city region level, i.e. the same scaling factor is used for all local authorities in a given city region.

The matrices generated by the first two steps can be applied sequentially in order to take estimates of travel done by residents living in a given local authority (from the scenarios), and turn these into estimates of travel done within each local authority, and on each road type. The odds ratios scaling factors calculated in Step three can then be applied, before calculating final proportions.

***Example output matrices, for Bristol city region local authorities for car/van travel*** *[in reality these matrices are often larger, including very small amounts of travel in a large number of local authorities, so a simplified version shown here]*

***1) For each home LA, for car/van trips, the proportion of distance in each English LA from Graphhopper***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | LA of travel |  |  |
| Home LA | E06000019 | E06000022 | E06000023 | E06000024 | E06000025 | E06000030 |
| E06000022 | 0.0% | 56.7% | 6.6% | 3.9% | 11.2% | 0.3% |
| E06000023 | 0.0% | 5.5% | 48.1% | 9.8% | 26.5% | 0.6% |
| E06000024 | 0.0% | 1.0% | 10.6% | 69.3% | 10.1% | 0.0% |
| E06000025 | 0.3% | 6.0% | 14.9% | 3.1% | 66.6% | 0.0% |

*NB cells show the percentages, e.g. for people living in E06000022 they do 56.7% of car travel within E06000022*

***2) for each LA of travel, for car/van trips, the total distance by road types from Graphhopper***

|  |  |  |
| --- | --- | --- |
|  | Road type |  |
| LA of travel | motorway | other |
| E06000019 | 0% | 100% |
| E06000022 | 0% | 100% |
| E06000023 | 21% | 79% |
| E06000024 | 28% | 72% |
| E06000025 | 32% | 68% |
| E06000030 | 37% | 63% |

***3) Compare relative proportions on motorways between graphhopper and RTS, to generate scaling weights: worked example for car/van travel in Bristol City Region***

In Bristol city region as a whole, the proportion of car/van travel done on motorways was 33.0% in RTS versus 26.8% after applying the above graphhopper routing method. This corresponds to an odds ratio of (0.330 / (1 – 0.330)) / (0.268 / (1 – 0.268)) = 1.34 to scale car/van motorway travel to RTS levels. The corresponding odds ratio for non-motorway roads is 1 / 1.34 = 0.74.

***Part 4: Process the graphhopper estimates to estimate travel distance by road type for stratified trip types, including applying RTS scaling factors for motorways***

* We stratified trips in each LA by

1. the urban/rural status of the home LSOA of participants, in local authorities with the least 10% each in rural and urban areas.
2. trip distance, using categories:

* for car/van and motorbike: 0-4.9km, 5-14.9km, 15-39.9km, 40km+
* for bicycle and bus: 0-4.9km, 5-14.9km, 15km+
* for walking: no sub categorisation by distance
* For each stratified subtype, we produced matrices as described in Steps 1 and 2 of Part 3, i.e. first LA of travel by LA of residence; and then road type by LA of travel.
* The second set of stratified matrices, road type by LA of travel, were calculated to incorporate within them the city-region-level scaling weights estimated in Step 3 of Part 3. Thus, for example, the odds ratio scaling weight of 1.34 applied to car/van motorway travel all trips subtypes, even though the initial proportion of motorway trips varied considerably (from negligible for shorter urban trips to substantial for long rural trips). The method for applying an odds ratio scaling weight to an initial percentage was as follows:
  1. convert the initial percentage ‘p1’ to an initial odds ratio ‘or1’, as or1 = p1 / (1-p1).
  2. multiply the initial odds ratio by the relevant odds ratios scaling weight to generate an updated odds ratio ‘or2’.
  3. convert the updated odds ratio to an updated percentage ‘p2’ as p2 = or2 / (1+ or2). This updated percentage is then used in the matrix.

***Part 5: Convert the matrices showing estimated travel distance by road type into matrices showing estimated travel duration***

* We took the matrices showing the proportion of travel distance in each LA done on motorways. We converted these into equivalent matrices for travel duration based on estimates as to how travel speed on motorways compares to travel on other roads.
* To do this, we first used NTS to select trips <5km, which we assumed to be entirely off motorways. On these short trips we calculated average speed at the stage level to be 20.9km/hr for cars, 22.2km/hr for motorcycles and 11.8km/hr for buses.
* We then compared these to the average speed of 94.9 km/hr (59 miles/hour) reported on England’s Strategic Road Network in 2018.[[6]](#footnote-6) This comparison gave scaling factors of 94.9/20.9=4.5 for cars; 94.9/22.2=4.3 for motorcycles; and 94.9/11.8=8.0 for buses. In other words, we estimate car travel speed to be 4.5 higher on motorways than on other roads
* We used the speed ratios to adjust the proportions in our distance matrices for road type. For example, in a local authority where 20% of car distance was estimated to be on motorways and 80% on other roads, we a) divided 20% by 4.5 = 0.044, and then b) rescaled this fraction to be a percentage as 0.044/(0.044+0.8) = 0.052, or 5.2% of duration estimated to be on motorways.

***Notes on potential limitations of this approach***

* We are using “trips made by people who live in a given LA **and are starting from that LA**” [from the census] as a proxy for “trips made by people who live in a given LA **but might be starting their trip anywhere**” [from the travel diary]. This could lead to underestimation of travel in LAs that attract a lot of people in from outside, e.g. you go from St Albans to Westminster, travel around in Westminster during the day for meetings/shopping etc, and then return home – the middle stage would not be captured. Although this may be partly offset by places that are big attractors of trips from outside, like Westminster, also disproportionately being commuting destinations.
* Relatedly, we are using **commute trips as a proxy for all trips** when it comes to estimating cross-border flows and % travel on motorways. In the NTS the mean distance for a commute trip by a given mode is similar to the mean distance for all trips by that mode (see table below). The variance is larger for all trips, however, especially when it comes to trips made by car/motorbike. This reflects a higher proportion of very long trips for non-commute purposes (e.g. 2.1% of non-commute trips by car/can/motorbike are over 100 km, as opposed to 0.7% of commute trips). These data indicate that commute trips are an imperfect proxy for all trips, and may underestimate the amount of time spent on motorways.

|  |  |  |
| --- | --- | --- |
|  | Mean distance (SD) in km | |
|  | Commute trips | All trips |
| Walking | 1.4 (1.1) | 1.2 (1.3) |
| Cycling | 5.5 (4.8) | 5.5 (6.8) |
| Car/van/motorbike | 15.8 (19.5) | 14.5 (30.0) |
| Bus | 8.4 (6.2) | 7.4 (7.3) |

* However comparisons with RTS and with LTDS data from London suggest discrepancies of the level we can live with (see Appendix). Specifically, agreement in terms of road type distribution was fairly good in the RTS comparison and very good in the London LTS comparison, while agreement in terms of the distribution across local authorities was generally good in the RTS comparison and moderate in the LTDS comparison.
* In addition, we believe these problems are substantially mitigated by the fact that we are (a) performing our calculation stratified by trip distance, and many trips switched in Metahit scenarios will shorter trips; and (b) we are adjusting the both total travel volumes and the scaling proportion for motorways to match RTS.

**APPENDIX: comparison of alternative datasets against the method of synthetic population + graph hopper routing + matrix transformation**

*Comparison of the census commute method with Road Traffic Data, in South West England*

* As previously discussed, NTS (and therefore our synthetic population) may underestimate total volumes of travel in general, as compared against the Road Traffic statistics (RTS). This is why we have generated these scaling weights, by region \* mode, to adjust upwards total volumes of travel.
* One can also compare the synthetic population and RTS in terms of the **patterning** of travel across local authorities and across road types. For South-West England, I compared a random sample of the synthetic population (200 routes in graphhopper per local authority per mode) with RTS data from 2011-2015. I found the two methods
  + agree well in terms of the correlation of volume at local authority level (r =0.81 for cycling, r=0.98 for motorbike travel, r=0.97 for car travel)
  + agree fairly well in terms of road type (e.g. RTS estimates that 17% of car distance is on motorways, versus 11% from the synthetic population/GraphHopper method. For motorcycles the equivalent percentages are 6% versus 9%).

*Comparison of the census commute method with all-trip origin destination data in London*

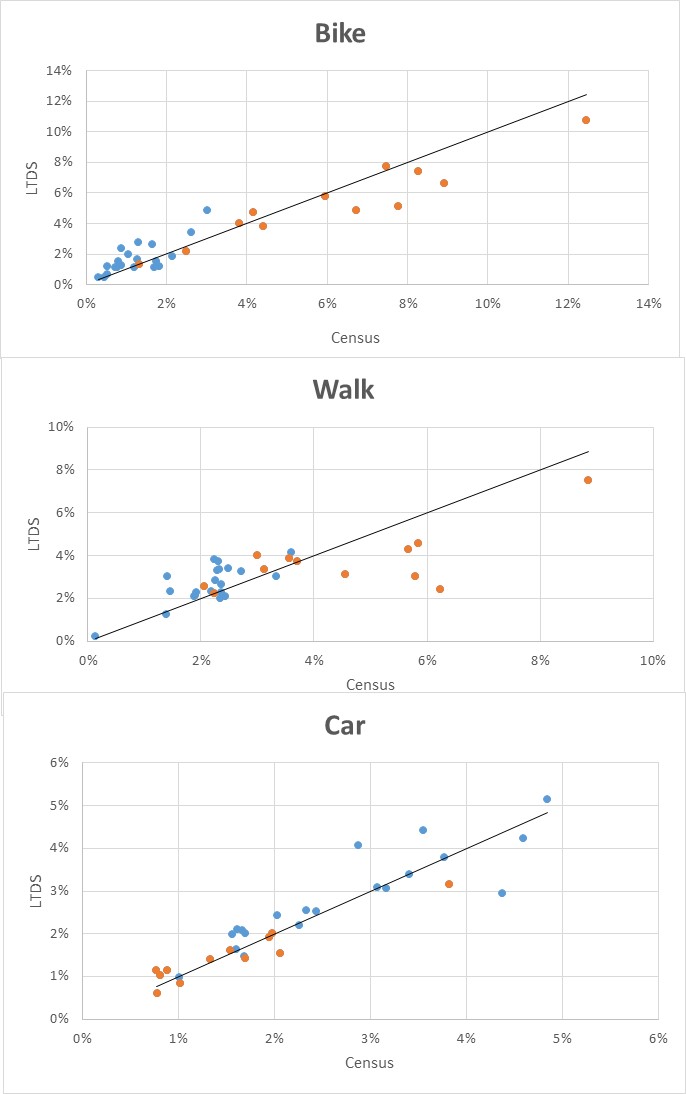
* In London, we have origin-destination trip data on all trips from LTDS. Comparing this to the census we find

1) in terms of road type distributions, agreement between the Census and LTDS generally very good in London:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | % motorway | % A road | % B/C/other |
| Bike | Census | 0% | 22% | 78% |
|  | LTDS | 0% | 21% | 79% |
| Walk | Census | 0% | 18% | 82% |
|  | LTDS | 0% | 17% | 83% |
| Car | Census | 23% | 60% | 17% |
|  | LTDS | 20% | 58% | 22% |

2) In terms of the proportion of travel in different local authorities,

* agreement between the Census and LTDS good for cycling and car driving: correlations r=0.95 for bike, r=0.99 for car (or r=0.92 excluding ‘outside London’ travel).
* Agreement less good for walking, r=0.76. See scatter graphs on next page
* For cycling and especially walking, tendency for the Census estimate to have relatively more travel in inner London (e.g. Islington, Camden) compared to the LTDS estimate [i.e. orange dots below black line]. For walking, the points further from the line are Islington and Tower Hamlets, followed by Hackney, Southwark, Westminster, Camden.



Blue=outer London/outside London, orange = inner London. Black line is the line of perfect agreement between Census and LTDS. The dots correspond to the 32 London local authorities (grouping Westminster and City of London) plus ‘outside London’ as a further 33rd group. Car graph excludes ‘outside London’ for car travel, which is 28% from the census data and 26% in LTDS.

1. <https://data.gov.uk/dataset/208c0e7b-353f-4e2d-8b7a-1a7118467acc/gb-road-traffic-counts> [↑](#footnote-ref-1)
2. Note that some A roads, for example the A1, is M in some places and A in others. [↑](#footnote-ref-2)
3. https://www.graphhopper.com/ [↑](#footnote-ref-3)
4. an option ‘motorbike’ is not available from graphhopper except via an external provider who does not provide ready access to road type information [↑](#footnote-ref-4)
5. Note that injury risks on buses are low and we are not modelling the addition of new buses to our fleet, and so this rather crude approach will have little impact on our model. [↑](#footnote-ref-5)
6. Department for Transport 'Travel time measures for the Strategic Road Network and local ‘A’ roads: January to December 2018 report', accessed 10/12/2019 from https://www.gov.uk/government/statistics/travel-time-measures-for-the-strategic-road-network-and-local-a-roads-january-to-december-2018 [↑](#footnote-ref-6)