Progress towards a free, open-source, spatially detailed, multimodal, and all-purpose transport model for Great Britain

Morgan M[*1], Philips I[†1]

[1]Insitute for Transport Studies, University of Leeds

January 11, 2022

Summary

Knowing how, why, when, and where people travel is essential for many forms of transport analysis. However, data is often hard to find as it is either expensive (mobile phone data), only considers specific journey purposes (census travel to work data), or specific modes (train ticket data). This paper reports on producing an open-source transport model covering all modes and journey purposes. We highlight novel uses of existing datasets and present preliminary results applicable to the transport planning and research communities.

KEYWORDS: Open source, transport, demand model, planning, travel time matrix

1. Introduction

Transport models are a core part of transport analysis inside and outside academia. We usually lack complete detailed data between all geographical areas on how, why, when, and where people travel, so transport models fill this gap using survey data as an input. While there are many different techniques for producing a transport model, a common approach is a four-stage process: 1) *Trip generation* – the number of trip origins and destinations in each area is ascertained based on population and land use. 2) T*rip distribution* – origins and destinations are linked to identify trips from one place to another. 3) *Mode choice* – trips are allocated to a mode of travel, e.g. car or bus. 4) *Route assignment* – trips are assigned to specific roads or public transport routes (Ortúzar and Willumsen, 2011).

This paper describes work to produce the first three stages of an open transport model for Great Britain. The main objectives for the model are:

- *Free and open-source*: While transport modelling principles are well established, it is unusual for models and data to be openly published and licenced for reuse.
- *National coverage and spatially detailed*: Many transport models only focus on a city/region, while national models often lack spatial detail. This model enables the consideration of both local and long-distance travel at a consistent level of spatial detail.
- *Multimodal:* The model considers six different modes of travel walking, cycling, bus/coach, light/heavy rail, car driver, and car passenger.
- *All trip purposes:* The model considers a wide range of trip purposes. This contrasts to previous national models, which often focus on a limited number of trip purposes such as commuting.

2. Method

The model construction is as follows. Firstly, the pre-existing National Trip Ends Model (NTEM) was used for the trip generation and mode choice. NTEM is produced by the UK Department for Transport and provides the number of trip origins and destinations disaggregated by mode and purpose for each

[*] M.Morgan@leeds.ac.uk
[†] I.Philips@leeds.ac.uk

of 7,700 zones across Great Britain. Secondly, a travel time matrix was constructed for each mode. Thirdly, an analysis of the National Travel Survey (NTS) (Department for Transport, 2021) produced decay curves indicating how the proportion of trips varies with travel time disaggregated by mode and trip purpose. Finally, a gravity model distributes the trips to produce a set of desire lines indicating the number of trips per day for each mode of travel.

2.1. NTEM

The NTEM is a long-established tool and data source for transport planning in the UK (Department for Transport, 2017; Tolouei et al., 2017). For this research, we focus on the NTEM's baseline scenario of travel demand in 2016. The NTEM has several characteristics that make it well suited as a starting dataset for this work. The model covers England, Scotland, and Wales while also having a high level of geographical detail. The analysis zones are Middle Super Output Area (MSOA), with each zone having a population between 5,000 and 15,000.

The number of trip origins and destinations for each zone can be extracted from the NTEM using the Tempro software, providing an Excel spreadsheet disaggregated by mode, journey purpose, and whether the trip was home-based.

2.2. Travel time decay curves

The Trips and Household tables from the NTS were joined. Data from 2016-19 was included (n = 860,499 trips). Trip related variables retained included duration, the mode and purpose. We derived a variable for whether the trip included a start or finish at home or whether it was a non-home based trip. Mode and purpose categories in the NTS were merged as appropriate to match the mode and purpose categories in the NTEM. The data used contains limited geographical information (as part of the anonymisation). We used data discerning the English region and whether the respondent lives in a rural or urban location.

For each of the combinations of mode, purpose, region and rural/urban, we calculated the proportion of trips in each 5-minute bin. The data were then fitted to a decay curve using Non-Linear Least squares regression Using R (https://www.r-project.org/) with the following formula.

$$\text{proportion of trips} \sim \mathrm{a} \times \mathrm{b} \times \mathrm{c} \times e^{-a \times \boldsymbol{time}} \times \left(\frac{1}{e^{-a \times \boldsymbol{time}}}\right)^{b+1}$$

The left of Figure 1 illustrates a curve fitted to the NTS data. Notice the single point high above the peak of the curve; this point includes journeys of 30 minutes and may be overrepresented due to NTS respondents approximating their travel time. Equally, journeys of 25 or 35 minutes appear underrepresented. The right side of Figure 1 shows the significant variation in the shape of the decay curves, highlighting the importance of accounting for mode, purpose, and location in the model.

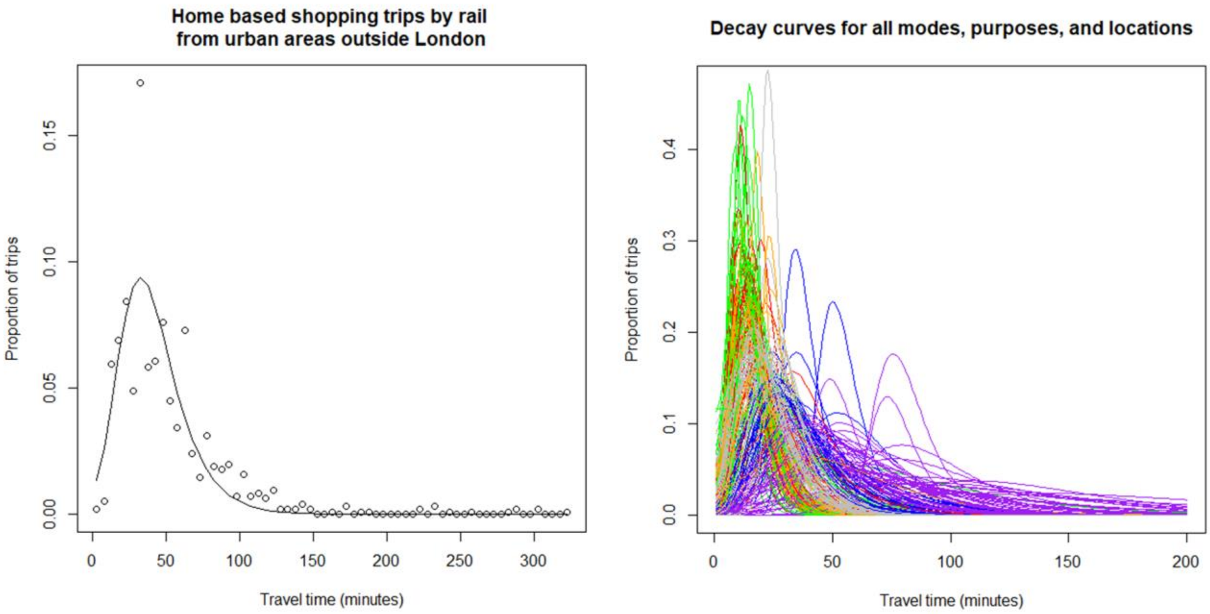

Figure 1 An example decay curve and the observed data (left), all 336 decay curves used in the model coloured by mode of travel (right).

2.3. Travel time matrix

The gravity model described below requires weighting values to ascertain the proportion of trips between each pair of zones. Ideally, this is done using a generalised cost calculation that accounts for the time and cost of travel. However, fare data for the UK is incomplete, so the simplifying assumption was taken to only consider travel time as a necessary precursor to calculating the generalised cost of travel.

Multiple tools exist to calculate travel time matrixes; however, most produce a sparse matrix that only considers travel times up to a pre-set threshold, often two hours. This threshold has the practical purpose of limiting the maximum computation time as the computation time increases exponentially with maximum travel time. Therefore, it is appropriate when considering certain journey purposes (e.g. commuting) or small geographical areas (e.g. a city). However, ignoring travel over two hours is not appropriate for a national model for all trip purposes.

The solution was a two-step approach. Firstly, the OpenTripPlanner for R package (Morgan et al., 2019) was used to construct a sparse travel time matrix of up to two hours for each mode of travel. Then a new algorithm was developed to interpolate the sparse matrix into a full matrix. Figure 1 provides a simplified illustration of the process. The journey time between Cardiff and Exeter is unknown as it is more than two hours. Any travel time estimate based on straight-line distance will be poor, as it would not account for the significant diversion required due to the sea. However, many intermediate locations are within two hours of both locations. Thus, the fastest route via a known intermediate point estimates the actual travel time.

Figure 2 Example of estimating travel times for journeys over two hours from known intermediate points

While Figure 2 shows only three intermediate locations, the algorithm considers hundreds of possible locations. It is also possible to apply the algorithm recursively doubling the maximum travel time with each iteration. Figure 3 shows that the method works very well for driving times, but for public transport modes, it was necessary to seed the sparse matrix with a small number of long-distance journeys such as the train times between major city centre stations. However, the method still struggles with trips between remote rural locations with intermittent public transport, especially if the trip involves multiple connections. Nevertheless, estimates of up to about 10 hours of travel time are reasonably good, so the method is adequate for weighing the gravity model.

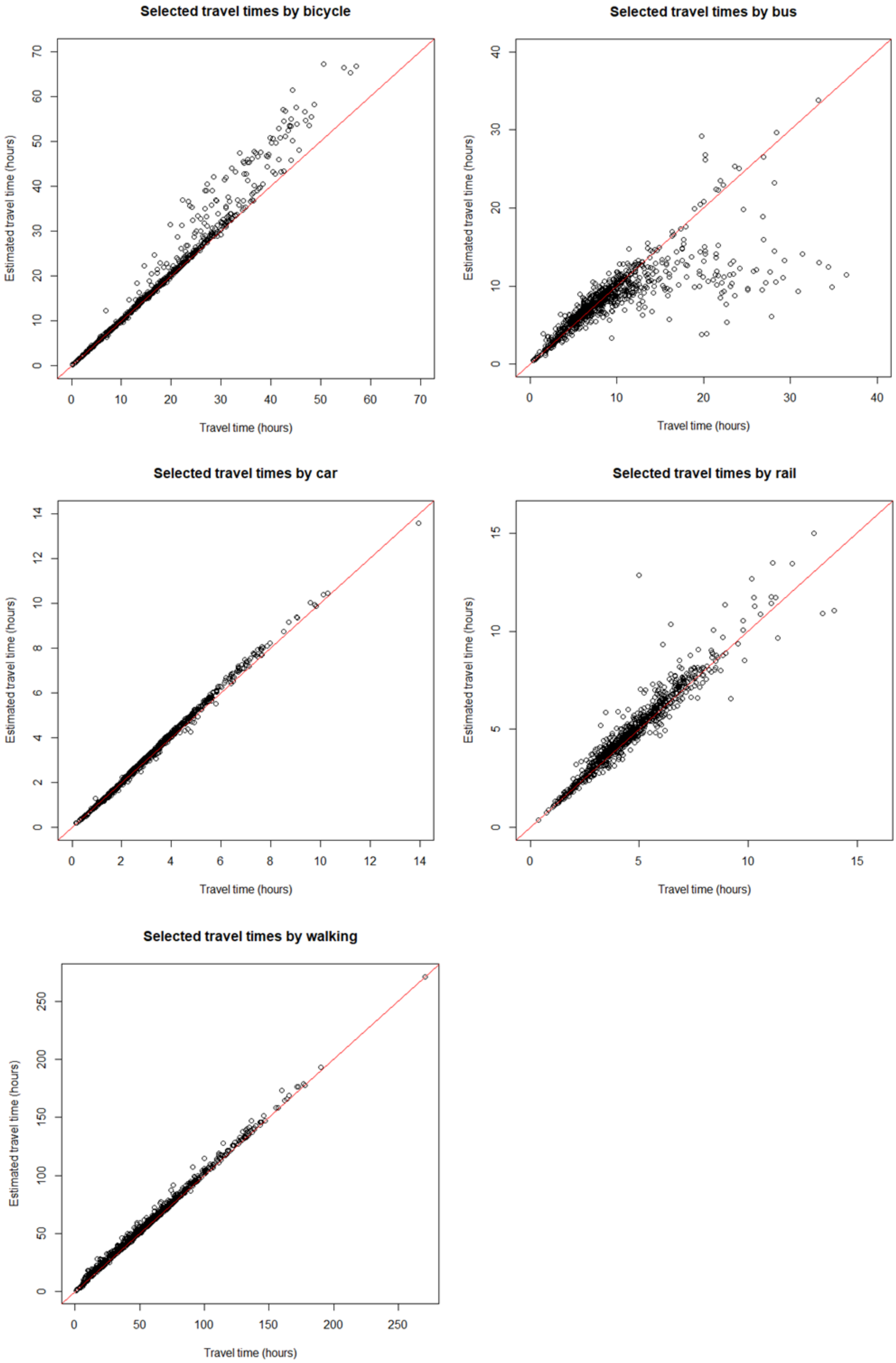

Figure 3 Estimated and actual travel times by mode for 1000 random origin-destination pairs.

2.4. Gravity model

Finally, the travel time matrix and the time decay curves were combined to provide weightings for a gravity model. The weightings were multiplied by the NTEM's count of trip origins for each zone and then applied to multiple iterations of balancing so that the number of trip origins and destinations matched as closely as possible. Finally, the journey purposes were aggregated to count the total number of trips on each origin-destination pair by mode of travel.

3. Initial results and conclusions

An alpha release of the travel time matrices, decay curves, and flow estimates has been published on GitHub (https://github.com/ITSLeeds/NTEM2OD/releases/tag/0.1). A preliminary analysis indicated a plausible distribution of trips consistent with the NTEM, see Figure 4.

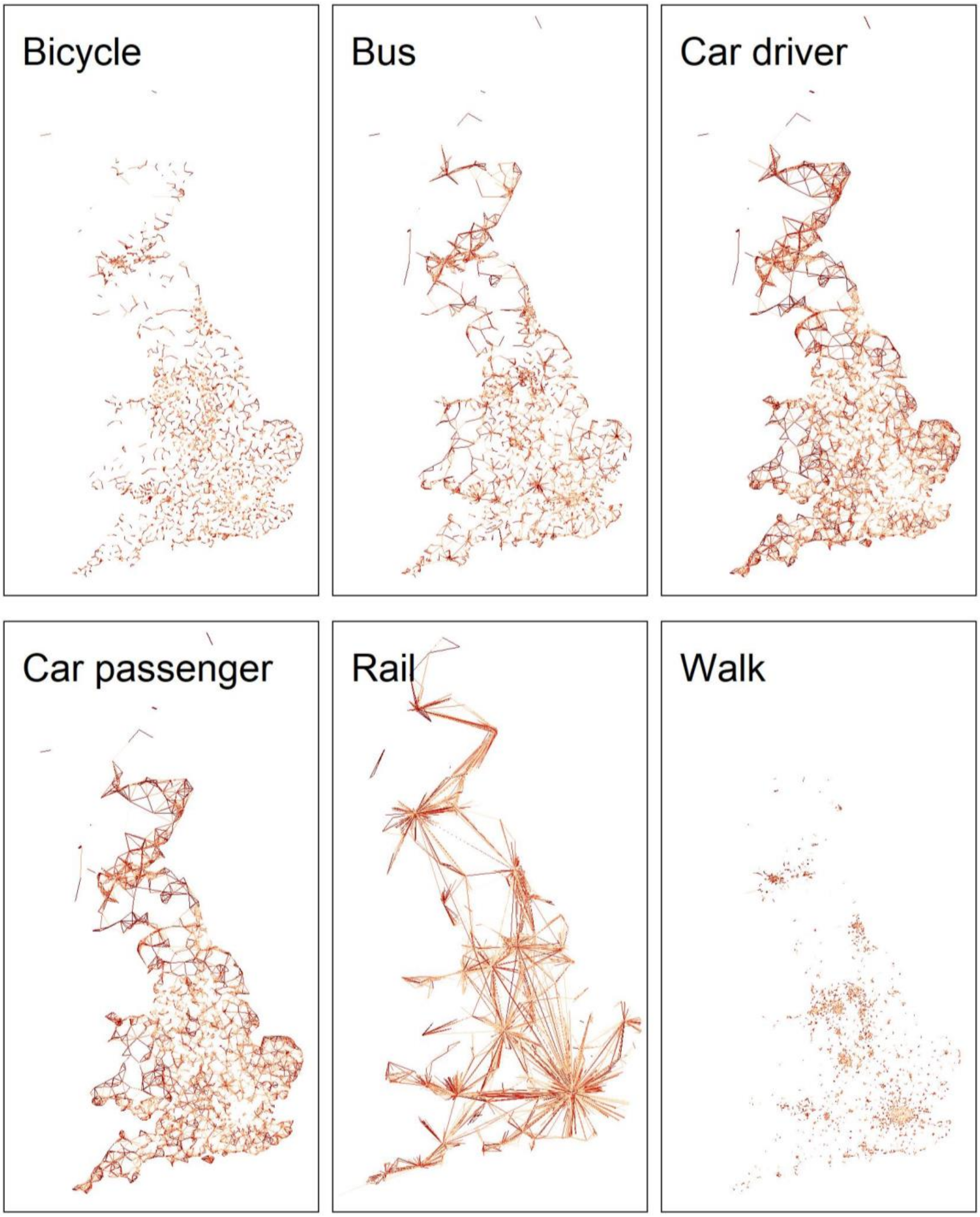

Figure 4 Maps of the top 1% of desire lines for each mode of transport.

This represents a significant contribution to free and open-source transport modelling. Though this is an early iteration, being open-source, it has the advantage of bringing to bear the input of a wide community to develop the nuance and accuracy of the model.

In future work, we hope to address the simplifying assumptions described, we also hope to address validation issues, such as the lack of available free data. Validation of the model is challenging, as there is a lack of comparison data. One possibility would be to use mobile phone data. However, this is expensive and usually comes with restrictive licences.

4. Acknowledgements

The authors would like to thank the Department for Transport for funding this work through the Shared Digital Carbon Architecture programme.

References

Department for Transport, 2021. National Travel Survey, 2002-2020. [data collection]. 15th Edition. https://doi.org/10.5255/UKDA-SN-5340-11

Department for Transport, 2017. National Trip End Model [WWW Document]. URL https://data.gov.uk/dataset/11bc7aaf-ddf6-4133-a91d-84e6f20a663e/national-trip-end-model-ntem (accessed 1.15.22).

Morgan, M., Young, M., Lovelace, R., Hama, L., 2019. OpenTripPlanner for R. J. Open Source Softw. 4, 1926. https://doi.org/10.21105/joss.01926

Ortúzar, J. de D., Willumsen, L.G., 2011. Modelling Transport, Modelling Transport. John Wiley & Sons, Ltd, Chichester, UK. https://doi.org/10.1002/9781119993308

Tolouei, R., Psarras, S., Prince, R., 2017. Origin-Destination Trip Matrix Development: Conventional Methods versus Mobile Phone Data. Transp. Res. Procedia 26, 39–52. https://doi.org/10.1016/j.trpro.2017.07.007

Biographies

Dr Malcolm Morgan is a Research Fellow at the University of Leeds, Institute for Transport Studies. He researches the spatial distribution of transport energy use and carbon emissions. He focuses on using data to develop tools that improve policymaking, such as the Place-Based Carbon Calculator www.carbon.place.

Dr Ian Philips is a transport geographer. Research interests include the use of spatial data and analysis to understand and promote sustainable and equitable transport. Projects have included analysis of car ownership and use, car dependence, socio-spatial patterns of vulnerability to motoring costs, E-bikes and their capability to reduce car CO_2 emissions.