Estimation of Travel to work Matrices

Table of Contents

Key Messages of Paper

Purpose

This paper presents an overview of travel to work estimation using aggregate spatial modelling approaches. The model produces estimates of travel to work matrices at small area level (Middle Layer Super Output Area, MSOA, for the first release) and at higher frequency (i.e., annually for the first release) than that currently available (i.e., Census travel to work matrices, published every 10 years).

The purpose of the first release of the model is to:

a) generate an alternative estimation of the travel to work matrices to the one reported in the 2021 Census. Our model assumes pre- and mid- coronavirus (COVID-19) Pandemic travel behaviour and is complimentary to the Census 2021 travel to work data collected over the pandemic when many were working from home or furloughed
b) develop and test a spatial modelling framework as the basis for future work- i.e. a more disaggregated and segmented model
c) develop model-based estimations of travel to work which can be used as a benchmark for quality assuring and understanding the statistics extracted from mobile phone data and other large data sources on travel to work

The first release of the travel to work matrices for 2021 have been published on the Office for National Statistics (ONS) websites as experimental data [1], alongside a technical report [2] and blog [3]. Prior to publication, a paper was submitted to Methods and Research Assurance Group (MaRAG 030, May 2023).

The longer-term vision involves developing a scenario-based modelling approach to generate regular and highly segmented travel to work matrices. The model will augment a wide range of ONS or available to ONS data, such as household, jobs, and travel surveys (e.g. Census, Inter-Departmental Business Register IDBR, National Travel Survey, etc), geospatial data (connectivity, accessibility, rural-urban classification, zonal boundaries, etc), and large datasets (Valuation Office Agency VOA, Mobile Phone Data, financial transaction data, etc) and allow to forecast travel to work under different social, economic or political scenarios.

Key Asks of MARP

The key asks of the panel for this paper are summarised in the relevant sections of the document and provided with more detail here:

- **Q1 – Can the panel advise on related projects across Government and Academia?** We are aware of other related areas of work within ONS, such as Population 24/7 and the Dynamic Population Model (DPM). Can the panel recommend other related projects or work programmes that share similar aims or use similar datasets or methodologies that we should engage with?
- **Q2 – Can the panel recommend additional datasets currently missing from our framework?** The current method relies on certain assumptions specifically on the proportion of people working from a fixed place of work and the growth in residential population and jobs at workplaces. The assumptions are based on models and datasets such as the National Trip End Model (NTEM) planning data, which provides

estimates of growth in workplaces/jobs (the NTEM itself is based on ONS population mid-year estimates, housing data from Local Authority local plans, etc). What other datasets would the panel recommend using for real time estimation and forecasting? Specifically, jobs data segmented by occupation and industry type is hard to locate.

- **Q3 – Can the panel advise on the appropriateness of the use of individual level data?** Given the other projects going on in the ONS on population, jobs, household and housing estimations, would the panel recommend for us to generate more detailed sub national population estimation in-house as opposed to relying on outside ONS resources such as Department of Transport (DfT)'s NTEM? Would access to more secure but individual level data at ONS potentially improve the assumptions we are currently using for estimating growth and proportion of those working at the fixed place of work?
- **Q4 – Are there alternative data or methods we should investigate to improve the calibration?** As Census 2021 travel to work matrices cannot be used as ground truth due to issues with collection during the pandemic, is there other data or methods we should investigate to calibrate the cost function to reflect the relationship between cost, such as distance vs. number of commuters.
- **Q5 – Can the panel recommend alternative methods to estimate the travel to work matrices?** If the panel has any expertise in geospatial population or transport modelling methods, is there any work we have not referenced herein that we should be aware of? Or any potential for collaboration with academic research groups in the panel's institutions? Can the panel advise of any other modelling methods we can use to estimate the travel to work matrices? If yes, we can compare the results between two models.
- **Q6 – Can the panel recommend alternative validation approaches?** One of our major challenges is how we validated our model estimates (we have undertaken a sense check by comparing Census 2011 and 2021 outputs, reported in Section 4 of our technical report [2]). The procurement of mobile phone data is currently ongoing but mobile phone data should not be the only data source for model validation. Mobile phone data can itself be subject to potential biases and one of the main purposes of these models is to test these. What other source of data or what alternative approaches do the panel suggest for validation? Should we consider alternative mobility data, such as from Google Mobility? Can comparison against sources of data such as Census trials in certain areas can be helpful? Is it reasonable to consider defining specific small scale ONS surveys against which the travel to work matrices can be validated?
- **Q7 - Can the panel advise on how we communicate increased uncertainty to end users?** Modelled estimations of travel to work matrices will likely contain larger errors than those derived from self-reporting, i.e., from the Census 2011, particularly in initial versions of the model.
- **Q8 - Can the panel comment on our planned future steps?**

Executive Summary

- We are developing aggregate spatial models to fuse various data and estimate travel to work matrices, at MSOA level for England and Wales.
- Currently, estimates of travel to work matrices for England and Wales are derived from Census counts of individuals' response to travel to work questions associated with their individual and household information in the Census questionnaires. Although Census provides a good approximation of travel to work matrices, the data is only collected every 10 years with no information available to fill in the gaps. Furthermore, it is unclear how useful the travel to work data from the 2021 Census is for policy makers and wider stakeholders given the significant impact of the pandemic over which period the data for the 2021 Census was collected [4].
- Our project aims to address the above issues and estimate the travel to work matrices annually from 2012 to 2021.
- Following discussions with our expert panel, we have identified the gravity model as a potential avenue for investigation.
- The gravity model has been widely used in transportation modelling to assign trips between origin and destination zones. We follow the Census travel to work definitions and have formalised the gravity model to answer our research question.
- We have implemented three core modules in our processing pipeline: trip end generation module, calibration of the cost function and doubly constrained module.
- The outcomes from the latest model are:
    - The population 16 years old and over who are in employment and travel to a fixed workplace, at residential zones, at MSOA level in England and Wales, annually for 2012-2021.
    - The number of employees 16 years old and over attracted to a fixed workplace, at MSOA level in England and Wales, annually for 2012-2021.
    - Travel to work matrices, at MSOA level in England and Wales, annually for 2012-2021,
    - Travel to work matrices for 2021 assuming pre- and mid- pandemic travel behaviour respectively.
- There are challenges in the validation of the outputs of the model due to the lack of ground-truth datasets.
- We have investigated the data quality of existing input data: National Travel Survey (NTS) and National Trip End Model (NTEMP) from Department of Transport (DfT), which demonstrate low confidence at fine spatial granularity (smaller than Region, such as Local Authorities and MSOAs) due to their sample size, response rate and coverage. This can affect the spatial granularity and reliability of the results the model can estimate. We have started to set up cooperation with DfT to improve the input data quality. Meanwhile we are considering and preparing to add several datasets to this framework specifically for better estimation of population and job growth as major inputs to the model, including Labour Force Survey (LFS), Business Register and Employment Survey (BRES) and mobile phone data.
- We have planned a medium-long term roadmap to improve the model for the future versions.

1. Introduction and Background

Travel to work (TTW) matrices show movement of people from their home location (origin) to their place of work (destination) at an aggregated level. Information on travel to work informs both national and local transport services and policies. It provides a basis for transport planning, for example, whether new public transport routes or changes to existing routes are needed. Additionally, it allows the measurement of environmental impacts of commuting, such as traffic congestion and pollution, and how these might change over time, for example because of changes in commuting modes, such as a shift from car to bicycle.

The Office for National Statistics (ONS) collects travel to work data from the census every 10 years, with the most recent being 2021. This travel to work data helps us generate travel to work matrices for census years, for instance, at 10-year intervals with no updates for years in-between.

Census 2021 was undertaken during a period of rapid change, during the coronavirus (COVID-19) pandemic. Government guidance was to stay at home where possible and avoid public transport. There was a shift to home working for those in industries and occupations that could, and approximately 5.6 million jobs in England and Wales were supported by the furlough scheme around the time of Census 2021 [5].

This means that travel to work data from Census 2021 is a combination of pre- and mid-pandemic travel behaviours, and it is unclear how furloughed respondents answered census questions. Changes such as hybrid and home working may be here to stay, but there will also be some return to pre-pandemic travel behaviours. It is therefore unclear how representative and useful Census 2021 TTW data is to policy makers given the context. More information is available on the quality of the travel to work census data [4].

Other sources of travel data are available, such as the National Travel Survey (NTS) produced by Department for Transport (DfT), or travel to work modules in the Opinions and Lifestyle Survey (OPN). However, neither source provides comprehensive travel to work matrices or reliable data below regional level, which is important for travel planning. For this reason, ONS is embarking on research to understand where novel modelling methods could be used to improve the timeliness and resolution of travel to work estimates.

Using aggregate spatial modelling approaches, the Data Science Campus has produced an alternative estimation of the travel to work matrices which bridges this gap. We have produced experimental data for each year from 2012 to 2021 and have published modelled estimates for 2021. When new data becomes available, the model can be updated to include 2022 onwards.

This will provide complementary statistics to the Census 2021 travel to work data collected during the pandemic, which contains a mixture of pandemic and pre-pandemic behaviours by estimating, for example, separate travel to work matrices using both pre– and mid-pandemic commuting behaviour. Modelled estimates for 2021 [1] are the first release in a planned work programme of incremental improvements to the model and outputs.

We have developed a detailed roadmap, this report and resulting experimental data are the first stage in this road map and are published to compliment the Census 2021 travel to work data, and to gather feedback from stakeholders on the technical specification and outputs. An

overview of the technical specification of the methodology is provided in Section 3 with the summary of input data provided in Section 2. Initial results are shown in Section 4. We discuss our methodology in Section 5. A broad summary of future stages to build on the current gravity model and to move towards a scenario-based recursive spatial equilibrium model in our vision is provided in Section 6.

Question 1 for panel

Can the panel advise on related projects across Government or Academia?

2. Data

The input datasets used in the first release of the model are listed in Table 1.

Table 1. List of the data sources used in the latest pipeline of the gravity model.

Dataset Name	Description and source	Purpose
wf02ew_oa	Census 2011 travel to matrix, England and Wales ONS https://www.nomisweb.co.uk/census/2011/wf02ew	To extract 2011 travel to work O-D matrices, population and employment as ground truth
TS066_Economic_activity_status	Census 2021 employment status table ONS https://www.nomisweb.co.uk/datasets/c2021ts066	To estimate 2021 working population
core_planning_growth_factor_v8	National Trip End Model (NTEM) v8 core scenario planning data 2011-2021 The National Trip End Model forecasts the growth in trip origin and destination up to 2051 for use in transport modelling. The Core_planning growth_factors_v8 dataset includes the growth of employment and population up to 2051. DfT https://www.data.gov.uk/dataset/11bc7aaf-ddf6-4133-a91d-84e6f20a663e/national-trip-end-model-ntem	The Core_planning growth_factors_v8 dataset is used to estimate the growth of workers and jobs between 2011 and 2021.
Various ONS Geography tables	Census 2011 definitions Census 2021 definitions Lookup tables between Census11 and 21 definitions Census 2011 Population Weighted Centroids ONS https://geoportal.statistics.gov.uk/	Census definitions and lookup tables are used to define zone translations between 2011 and 2021. Population weighted centroids are used to calculate the distance matrix used in the cost function

National Travel Survey. household and individual tables	National Travel Survey (NTS) 2011-2021 household and individual tables NTS is a household survey designed to monitor long-term trends in personal travel which collects information on how, why, when and where people travel as well as factors affecting travel by residents of England. DfT https://www.gov.uk/government/collections/national-travel-survey-statistics	To estimate the proportion of workers with a fixed workplace by latent geographic cluster
Latent geographic clusters	Land use clusters defined by land use characteristics including area type, population density and accessibility measures. Jahanshahi & Jin (2021) [6]	To estimate the proportion of workers with a fixed workplace using travel patterns from the NTS

As this work is incremental to improve the modelling and output quality with more data sources, we are preparing to add several datasets to this framework, including Labour Force Survey (LFS), Business Register and Employment Survey (BRES) and mobile phone data soon. More details can refer to Section 6: Future Steps.

Question 2 for panel

Can the panel recommend additional datasets currently missing from our framework?

Question 3 for panel

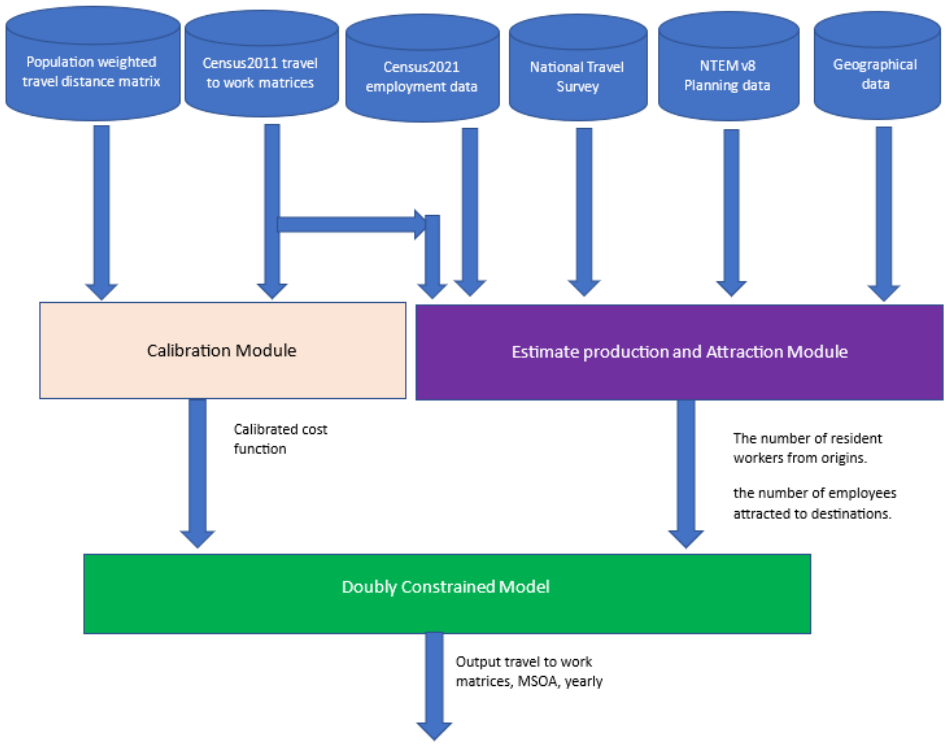
Can the panel advise on the appropriateness of the use of individual level data?

3. Methods

We have developed a gravity model, calibrated using the 2011 Census travel to work data, to estimate travel to work matrices. The classical gravity model described in Chapter 4-5, Modelling Transport [7] assumes that the number of trips made between two areas can be estimated by the number of resident workers at the origin (residential end), the number of employees working at the destination (workplace end) and the cost, such as travel time, travel distance or monetary cost (or a combination of these factors referred to as a generalised cost) of travelling between these two locations.

Figure 1 shows the processing pipeline, that is, how the different sources of data feed into the model and generate the estimation of travel to work matrices. The pipeline consists of three core modules: Production and Attraction Estimation Module, Calibration Module and Doubly Constrained Module.

Figure 1: The processing pipeline of the gravity model



The Production and Attraction Estimation Module estimates the number of resident workers travelling to a fixed workplace at each origin, and the number of employees working in each destination, using these assumptions:

- the number of resident workers for years 2012 to 2020 is estimated using growth factors in workers taken from DfT's National Trip Ends Model (NTEM), applied to 2011 Census values to estimate future year resident workers
- the number of resident workers for the year 2021 is estimated using Census 2021 resident employment data
- the proportion of resident workers travelling to a fixed workplace is estimated from DfT's National Travel Survey (NTS), 2011 to 2021, segmented by latent geographical clusters defined by land use characteristics (refer to Table 1 and our technical report [2] for more detail).
- the proportion of resident workers travelling to a fixed workplace for the year of 2021 is estimated for pre-pandemic, using NTS 2018-19 commuting travel patterns, and mid pandemic, using NTS 2020-21 commuting travel patterns.
- the number of employees at each destination is estimated using growth factors in jobs from NTEM, applied to 2011 Census values as Census 2021 workplace employment data is not available at the time of writing

In the Calibration Module, the cost of travelling between the origin and destination is currently estimated using straight-line distance. The model is calibrated using a cost function that aims to replicate the relationship between distance between home and work and the number of commuters, as seen in the 2011 Census travel to work matrices. In later versions of the model, we plan to replace distance with better estimations of costs, in terms of time and monetary expenses, of travelling between home and work.

Question 4 for panel

Are there alternative data or methods we should investigate to improve the calibration?

An iterative process, called the Doubly Constrained Module, then ensures that the estimated number of resident workers at all origins matches the number of employees working at all destination

For more detailed information, please refer to Section 3 of our technical report [2].

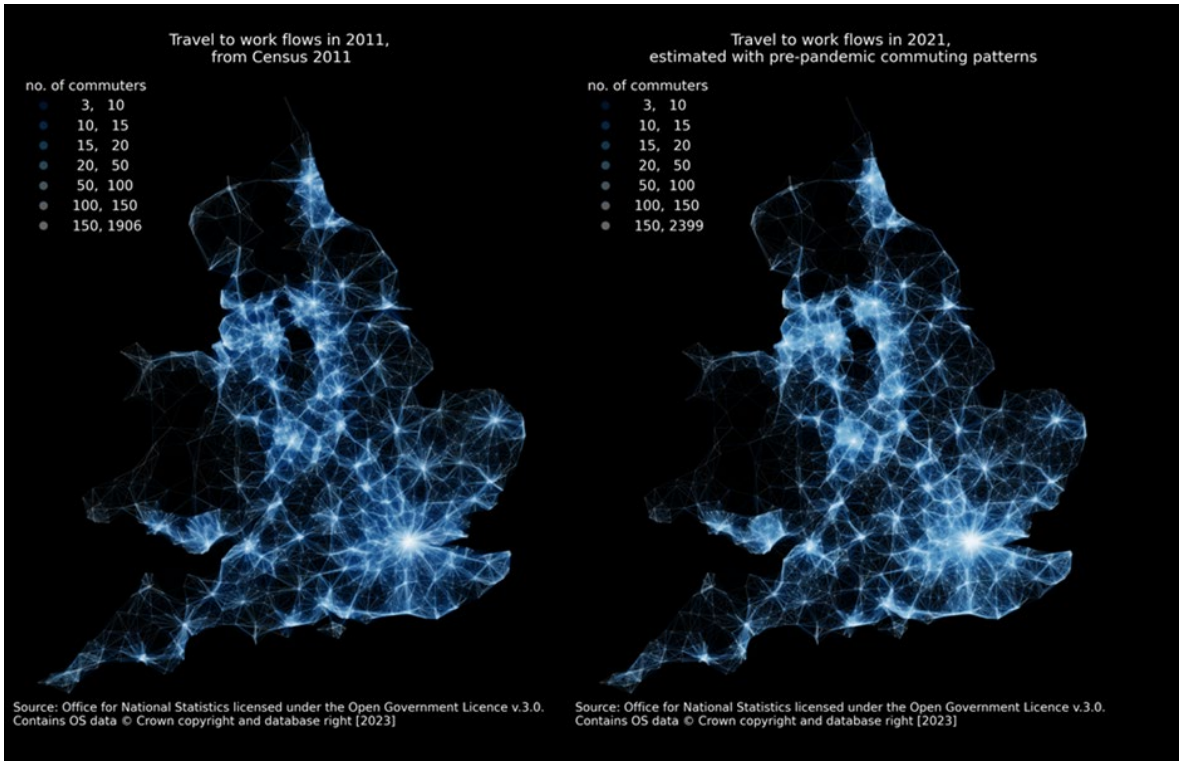
Question 5 for panel

Can the panel recommend alternative methods to estimate the travel-to-work matrices?

4. Results

More comprehensive discussion on results including calibration goodness of fit is provided in Section 4 of the technical report [2]. Here we have just included the visual comparison between travel to work in 2011 (from 2011 Census) and 2021 (under the assumption of pre-pandemic commuting patterns). It can be observed that general patterns are the similar as expected but the trend to work in bigger urban areas looks to be larger in 2021 estimate.

Figure 2: Travel to work flow map, MSOA, England and Wales, 2011 Census and Census 2021 assuming pre-pandemic commuting travel behaviours, respectively



5. Discussion

This section will discuss issues which affect the quality of results and influence the project.

5.1 Calibration

The cost function is calibrated based on the ground truth of Census 2011 travel to work data. In a short time period, the calibrated cost function can be stable; however, in the long term, the cost function will shift gradually; for example, the populations' commuting distance may change due to the road networks improvements, industry changes and social-economic changes. We are investigating how the cost function can be updated regularly to reflect the latest travel to work behaviour patterns.

5.2 Validation

An unresolved issue for this project is the validation of outputs. Outputs of such models are usually validated by comparison with observed data, such as from surveys. This has not been possible in the absence of a representative survey of travel to work for 2021, such as the context of the Census 2021. We are currently investigating potential datasets which can be used for validation. These include:

- mobile phone data, for example, aggregated footfall data derived from mobile phone users, the usefulness of which is subject to transparent methodology for data generation and representative data for the small area demographic statistics; This data is not yet available to us
- future releases of NTS, which will provide regional data on travel behaviour trends; however, the NTS is not representative at more detailed geographic granularity (for instance, at MSOA level) so cannot be used directly to validate travel to work matrices
- Census 2021 travel to work matrices; the quality is not sufficient for validation but can be used to compare model outputs to identify where substantial differences occur
- a 2019 census rehearsal was undertaken for four areas, Carlisle, Ceredigion, Hackney and Tower Hamlets
- travel surveys have been undertaken by local authorities such as Greater London Authority, Transport for London, and Greater Manchester; these may provide useful validation data

5.3 Segmentation

The model will be further improved by segmentation, such as travel distance, travel modes, type of jobs and socioeconomic characteristics. These segmentations will improve the calibration and the quality of travel to work matrices and will result in much higher confidence in the outputs. For instance, segmentation by travel distance allows fitting separate models for different distance bands given different potential travel behaviours and sensitivity to cost and time for different travel distances and modes. In addition, segmentation will enable travel to work matrices to be estimated by travel modes and population segments, such as age and occupation. Moreover, incorporating segmentation by socioeconomic and demographic characteristics can help evaluate policy constraints and changes in travel patterns (such as the effect of shifting to work from home or hybrid working by job type and other socioeconomic characteristics post-pandemic). This latter point can contribute into our long-term vision for a cloud-based interactive solution which allows scenario-based analysis.

Question 6 for panel

Can the panel recommend alternative validation approaches?

5.4 Data Caveats

Existing data, such as the National Travel Survey (NTS) and National Trip End Model (NTEMP) from DfT also demonstrate low confidence at fine spatial granularity (smaller than Regions, such as MSOA) from our exploration and investigation on the input data due to their sample size, response rate and coverage. This will affect the spatial granularity and reliability of the results the model can estimate. We have started to set up cooperation with DfT to improve the input data quality. Meanwhile we are planning to add several datasets to this framework including Labour Force Survey (LFS), Business Register and Employment Survey (BERS) and mobile phone data to improve the model and the quality of the output.

Question 7 for panel

Can the panel advise on how we communicate increased uncertainty to end users?

6. Future Steps

The longer-term vision involves developing a scenario-based modelling approach to generate regular and highly segmented travel to work matrices. The model will augment a wide range of ONS or available-to-ONS data, such as household, jobs, and travel surveys (including census, IDBR, NTS, and so on), geospatial data (connectivity, accessibility, rural-urban classification, zonal boundaries, and so on), and large datasets (VOA, Mobile Phone Data, financial transaction data, and so on) and allow scenario-based analysis to enable decision making under uncertainties on future housing market, jobs market, economy, and technology developments. This allows testing "what-if" policy questions and scenario-based analysis for a wider purposes and population segments.

Aligned with the above vision, the possibilities for future work on this project are extensive; these include:

- plan and implement a strategy for model validation (we see this as the highest priority)
- develop better estimation of jobs and population growth using various sets of detailed data and products across the ONS and through developing joined and collaborative plans
- segmenting the model by at least job type (for example, industrial classification, SIC, at workplace end), residential population type (for example, occupation type, SOC, at residential end) and travel distance bands (for example, short, medium, and long-distance trips) to improve model calibration and validation results
- further segmentation by main transport modes including public transport, private modes and non-motorised modes
- incorporate mobile phone mobility data for residential and workplace estimation and travel to work validation

- extend the developed model to include wider travel purposes including leisure and tourism
- build on current spatial model to develop a recursive spatial equilibrium model that can account for major development or restructuring. In the model in our vision, producer and consumer choices adjust to temporal changes in activity churn, background trends, estate development, and transport supply
- develop a scenario generator which can update model assumptions (for example, proportion of working from home for office-based SICs) under different user-specified scenarios
- developing better spatial visualisation of movements and population
- incorporate other deterrence functions for measuring connectivity in addition (we are currently using travel distance); the first attempt can be developing a generalised cost from combining travel time and distance given the value of time of different socioeconomic and demographic segments
- develop the fully reproducible model on the cloud which can read in the data (for example, mobile phone data), recalibrate the model and produce outcomes given requested scenarios in real time linked to a dashboard for dissemination and report generation

Question 8 for panel

Can the panel comment on future steps.

7. References

- [1] Estimation of travel to work matrices, England and Wales, https://www.ons.gov.uk/releases/estimationoftraveltoworkmatricesenglandandwales, accessed 2023-09-06.
- [2] Technical report: Estimation of travel to work matrices, https://datasciencecampus.ons.gov.uk/projects/technical-report-estimation-of-travel-to-work-matrices/, accessed 2023-09-06.
- [3] Blog: Estimation of travel to work matrices, https://datasciencecampus.ons.gov.uk/estimation-of-travel-to-work-matrices/, accessed 2023-09-06.
- [4] Travel to work quality information for Census 2021, ONS, 8 December 2022, https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/traveltoworkqualityinformationforcensus2021, accessed 2023-05-12.
- [5] Coronavirus and homeworking in the UK labour market: 2019, 24 March 2020, ONS, https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/coronavirusandhomeworkingintheuklabourmarket/2019 , accessed 2023-05-14.
- [6] Jahanshahi, K., Jin, Y., 2021. Identification and mapping of spatial variations in travel choices through combining structural equation modelling and latent class

analysis: findings for Great Britain. Transportation (2021) 48:1329–1359. https://doi.org/10.1007/s11116-020-10098

- [7] Juan de Dios Ortúzar, Luis G. Willumsen, Modelling Transport, 4th Edition, ISBN: 978-0-470-76039-0 April 2011.