Developing enhanced route choice models for heavy goods vehicles using GPS data

Stephane Hess[1] – Mohammed Quddus[2] – Nadine Rieser[3] – Andrew Daly[1,4]

[1] Institute for Transport Studies, University of Leeds

[2] School of Civil and Building Engineering, Loughborough University

[3] Institute for Transport Planning and Systems, ETH Zurich

[4] RAND Europe

INTRODUCTION

GPS data has established itself as a key tool for measuring route choices and the subsequent modelling of these choices in random utility models. This technique has been successfully applied across different countries, primarily for car travel but also for walking and cycling. The level of sophistication of the estimated models as well as the choice set generation approaches varies widely across existing studies. The present paper presents an application that is somewhat different in scope, looking at the modelling of route choices for heavy goods vehicles, which typically make longer journeys (up to 500km in the present dataset) and where the decision making is potentially underpinned by different priorities from those used by car drivers, as an example. Furthermore, while many previous route choice studies have been conducted at the level of metropolitan areas or at best small countries, the present application uses the entire road network of England, which contains some 4.5 million individual links.

DATA PROCESSING

About 8.7 million GPS observations from the movements of 709 HGVs over a month (April 2010) were matched onto a road network database to form journeys in which a journey made by a vehicle is defined as being between the time when the ignition is turned on and the time when the ignition is turned off. There are a total of 68,403 unique journeys (within and between HGVs) in the journey database and these journeys represent the movements of HGVs across the entire road network of England (see Figure 1). During a journey, a vehicle travels on average through 126 links, some of which are quite long, with a 95th percentile of 1,080m.

Figure 1: Junctions travelled by 709 vehicles in April 2010

The first challenge was to turn these journeys into trips in which a trip is defined as a set of journeys grouped together, with trip end points being either pick-ups or deliveries. Therefore, a trip can either be a single journey (if a journey starts and ends at a delivery point or a pick up point or the vehicle's depot) or a combination of sequential journeys (if journeys have STOPS due to the fact that the vehicle stops at a truck-stop for refuelling or a rest break). Additionally, any gaps in the GPS recording (due to unavailability of GPS signals) would once again result in separate journeys. Our initial task therefore was to group together any journeys deemed to belong to a single trip, where we used the algorithm summarised in Figure 2.

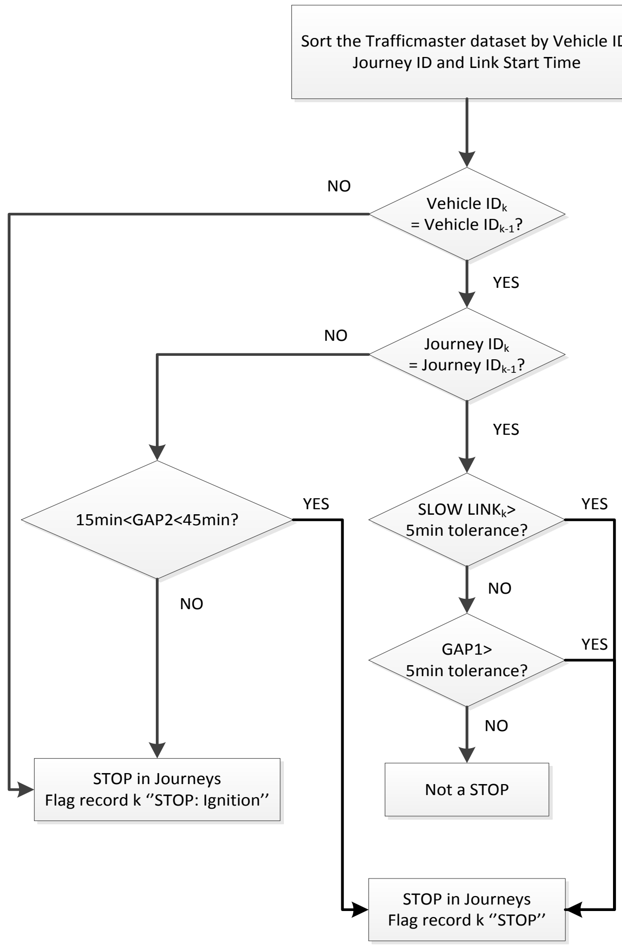

Figure 2: A flowchart to identify a STOP in journeys

Using this approach, we identified records when the vehicle stops. We now need to identify which of these stops is most likely due to a TRIP END or whether we're unable to determine why the vehicle may have stopped. The records flagged as STOP:IGNITION are when the vehicle ID has changed or the journey ID has changed for a stop outside of a 15-45 minute window. Both these cases indicate the vehicle is stopping for a delivery or pick-up i.e. these are TRIP ENDs. The records flagged as STOP are where the is a SLOW LINK or a GAP in the data of at least 5 minutes or where the journey ID has changed for a STOP of 15-45 minutes. These stops could be for a number of reasons, either a TRIP END (delivery or pick-up), a BREAK (rest break), TRAFFIC (heavy congestion or a traffic incident), or something else – e.g. a driver stops to check directions or take a phonecall.

We can identify BREAKs from comparing the vehicle location against a database of truck stops and motorway service areas and from an understanding of driver rules on rest breaks. Where we can identify BREAKs we can remove the STOP flag on the data as the records are parts of the same TRIP.

Using the above rules, the original set of 68,403 journeys were turned into 65,405 separate trips.

CHOICE SET GENERATION

The next step consisted of constructing choice sets of possible routes for the origin destination pairings for each of these trips. Generating a large set of possible routes is crucial with a view to producing unbiased results, as it is very unlikely that when making their route choice, a driver only considers the cheapest, fastest, and shortest options. A first step consisted of converting the data (i.e. network and map-matched routes) into a format required to run the choice set generation, which also involved imputation of any missing link coordinates, which was an issue in a network of the size used in this study. Data cleaning was then performed by removing links and nodes that are not connected to the rest of the network, and checking connectivity of the routes, i.e. ensuring no gaps in the data.

For choice set generation, we used the breadth first link elimination approach described by Rieser-Schüssler et al. (2012). We produced choice sets of different sizes to allow us to test the sensitivity of the model results to choice set generation settings, using choice sets of 10 routes, 15 routes and 20 routes. Such smaller numbers (compared to some intra-urban studies) are justified on long distance journeys and also due to the fact that heavy goods vehicles primarily make use of the key main roads in the network.

For each of the routes included in the choice set, we calculated a number of key statistics to be used in model estimation, namely total distance, time and fuel consumption, as well as distance and time spent on each type of road, and at a more disaggregate level, distance and time spent on one of 161 key roads in England. Additionally, we computed the path size value to allow us to quantify route overlap, using the second formulation by Ben-Akiva and Bierlaire (1999).

MODELLING

A number of different techniques are currently being applied to model the choice of route, from basic logit structures through to using error components models as put forward by Freijinger & Bierlaire (2007). This latter model is especially appealing in the context of longer distance route choices by commercial vehicles as it allows us to capture the strong correlation between different routes using key parts of the strategic road network, such as for example the M25 ring road around London, and the key motorways leading from London out to different parts of the country. Initial model results are very encouraging, yielding highly significant parameter estimates. Strikingly, the value of time estimates obtained from our initial models are almost perfectly in line with official recommended measures for valuation as put forward in the UK Department for Transport's web-based guidance for practitioners (www.dft.gov.uk/webtag/).

REFERENCES

Ben-Akiva, M. and M. Bierlaire (1999), Discrete choice methods and their applications in short term travel decisions, in R.Hall, ed., 'The Handbook of Transportation Science', Kluwer, Dordrecht, the Netherlands, pp. 5–33.

Frejinger, E. and Bierlaire, M. (2007), Capturing correlation with subnetworks in route choice models, Transportation Research Part B 41(3), pp. 363-378.

Rieser-Schüssler, N., M. Balmer and K.W. Axhausen (2012) Route choice sets for very high-resolution data, Transportmetrica.