Highway Trip Matrix Development: Update

Denvil Coombe

SATURN USER GROUP

28 November 2013

Introduction

- Usual assignment user classes:
  - Cars on employers' business trips (in work time: IWT)
  - Cars on other trips (out of work time: OWT)
  - Taxis
  - LGVs
  - HGVs
- The focus in this presentation is on car trip matrices
- The ideas and advice contained in this presentation have been developed over the last few years during work for Transport for London and other clients

Car Trip Matrices: Steps 1 to 5

1. Collection, editing and expansion of intercept survey data
2. Collection, editing and reconciliation of count data
3. Synthesis of matrix cell values in the non-interviewed directions
4. Creation of partial (‘observed’) trip matrices
5. Analysis of the accuracy of the partial trip matrices at sector level

Car Trip Matrices: Step 6

6a Assembly of synthesized trip ends

6b Assembly of generalized cost matrices

6c Assembly of trip cost distributions

6d Trip matrix synthesis

6e Assembly of external-to-external trip matrices

Car Trip Matrices: Steps 7 to 11

7 Factoring of daily or period trip matrices to assignment hours

8 Adjustments to the prior trip matrices in the light of the prior trip matrix tests

9 Matrix estimation to ensure greater consistency of the trip matrices with the count data

10 Adjustments to the prior trip matrices if the changes brought about by matrix estimation are regarded as significant

11 Adjustments to the prior trip matrices in the light of the journey time validations

Scope of Today's Presentation

- A more complete specification of the 11 steps can be found on the SATURN website under the 2011 UGM
- Today I will
  - reiterate some of the more important steps (1, 2, 4, 5, 6)
  - introduce a refinement which further recognises the inaccuracies in two of the main data inputs
  - discuss some aspects of using data from new sources (mobile phones, GPS tracking systems, satellite navigation devices)

Prior Matrix Tests

Stage	Test	Comparison	Measure	Criterion	Acceptability guideline
Final model		Total assigned flows and counts across RSI, calibration and validation screenlines, by time period.	Flow differences	< 5%	All or nearly all
Partial trip matrices	A	Flows and counts of trips across **RSI enclosures**, for peak/inter-peak periods separately or 12 hours or 24 hours, depending on periods used for gravity model calibration and trip synthesis.	Flow differences	< 5%	All or nearly all
Synthetic trip matrices	B	Flows and counts of trips across **RSI enclosures**, for peak/inter-peak periods separately or 12, 16 or 24 hours, depending on periods used for gravity model calibration and trip synthesis.	Flow differences	< 7.5%	All or nearly all
Prior trip matrices	**C**	**Total** assigned flows and **total** counts in both directions across **RSI, calibration and validation screenlines**, for each modelled hour.	Flow differences	< 7.5%	All or nearly all

Step 1: collection, editing and expansion of intercept survey data (1)

- Enclosures (cordons) and screenlines must be ‘watertight’
- For screenlines, movements which could partially route around the ends of the screenlines should be omitted
- Gaps on major roads because surveys were not feasible or permitted or abandoned should be treated as follows
  - Either use old RSI data re-expanded to new counts, providing that land-uses have not changed materially
  - Or synthesize RSI trip records by means of Select Link Analysis using an existing model
  - For both sources, low weights should be applied in the averaging process in the creation of the partial trip matrices (Step 4)

Step 1: collection, editing and expansion of intercept survey data (2)

- Flows on minor residential streets might form a material proportion of the total cordon or screenline flows but flows on individual roads might be too low to justify RSIs
- The trip end estimates will be estimates of total trips and so the partial matrices should be comparable
- Therefore gaps on minor roads should be treated as follows
  - Group gap sites serving the same area and obtain the total flow
  - Extract appropriate trip records from RSI sites on nearby roads and expand to the total gap flow – exclude trips with Os or Ds which are unlikely to be served by the gap roads
  - Merely expanding the trip records at the surveyed sites to the total flow at the surveyed sites and associated gaps would yield distorted trip patterns

Step 2: collection, editing and reconciliation of count data (1)

- Counts are required for:
  - expanding samples of new roadside interviews
  - re-expanding samples of old roadside interviews
  - calibrating matrices by means of matrix estimation
  - validating the model

Step 2: collection, editing and reconciliation of count data (2)

- General points
  - The traffic counts to be used for calibration and validation need to be specified at an early stage and should be of the same general quality
  - Calibration screenlines need to focus on synthesized or unobserved movements
  - Validation screenlines need to be independent of RSIs and calibration screenlines
  - Usual approach: Single-day MCCs to give vehicle proportions but indexed to ATCs conducted for two weeks to give more accurate totals – but note that vehicle proportions are still based on a single day's count
  - Alternative approach: Five-day MCCs and no ATCs

Step 2: collection, editing and reconciliation of count data (3)

- Treatment of gaps
  - Flows on all major roads must be counted
  - Flows on all minor roads may be estimated based on some sample counts and a road typology
  - Either counts or estimates are required for ALL roads crossing cordons and screenlines

- But the inclusion of counts on ALL roads does NOT mean that all the minor roads should be included in the network – the network density should be consistent with the zoning system (if problems in the assignment are to be avoided)

Step 4: creation of partial trip matrices (1)

- The format of the partial trip matrices will need to accord with the format employed in the trip synthesis (in Step 6)
- Current best practice is specified in the ERICA5 Manual although it may be preferable to develop and use other software
- This method derives weighted averages where there is more than one estimate available for particular cell values, with the weights being based on indices of dispersion (normalised variances)
- Key section is 3 b viii in the ERICA5 Manual on combining different sources of error – however, this specification, as written, is incomplete and further information is required in order to implement the method

Step 5: analysis of the accuracy of the partial trip matrices at sector level (1)

- TAG Unit 3.10.2 advises
  - "*1.6.11 The process of "introducing observed data" must then make allowance for the statistical accuracy of that data, based essentially on sampling theory (see guidance in DMRB 12.1). This could be done along the following lines. For each observed cell of the matrix..., the "prior" value would be tested as to whether it lay outside of the confidence region of the observed data: if so, the prior data will need to be modified."*
  - At zonal level, the prior (synthesized) value will almost always lie within the 95% confidence intervals and so little or no use would be of the 'observed' data

Step 5: analysis of the accuracy of the partial trip matrices at sector level (2)

- TAG Unit 3.10.3 advises
  - "*1.5.21 However, by taking a weighted average of the observed and synthetic matrices empty cells are eliminated and greater weight is given to cells where there are more observed trips than expected from the locally-calibrated synthetic model. Relative weights should reflect the relative accuracy of the two forms of estimates. If these are not known then a rough guide would be to use 90% of the observed estimate and 10% of the synthesised estimate."*
  - Again, no mention of the level of spatial detail at which these processes should be carried out
  - No advice on how to determine weights which reflect the relative accuracy of the 'observed' and synthesised values (although this can be done)
  - The 90%/10% advice means that the lumpiness of the 'observed' matrices (due to sampling variability) would be retained (undesirably)

Step 5: analysis of the accuracy of the partial trip matrices at sector level (3)

- Common practice
  - It is not uncommon for analysts to use ‘observed’ trip matrices at zonal level in one of the three approaches advised in WebTAG
    - In the Unit 3.10.2 approach, this would lead to little or no use of the ‘observed’ trip estimates
    - In the first method in Unit 3.10.3 which weights the estimates by reliability, greater weight would be attached to the ‘observed’ data than the synthesised
    - In the second method in Unit 3.10.3, the 90%/10% method, little use would be made of the ‘synthesized’ data

Step 5: analysis of the accuracy of the partial trip matrices at sector level (4)

- Common practice
  - It also not uncommon for analysts to compress the zonal level partial matrices to sector level before using them to ‘control’ the synthesized matrices
  - Usually, the sector systems used are those defined by the RSI cordons and screenlines
  - In this approach, typically
    - the majority of the sector level cell values will have 95% confidence intervals which are too wide to be useful because the numbers of trip records are too few
    - the minority of sector level cell values will have satisfactory 95% confidence intervals, often based on more trip records than necessary

Step 5: analysis of the accuracy of the partial trip matrices at sector level (5)

- A better approach
  - Define a 95% confidence interval which would be regarded as acceptable – say, 20% to 30% of the cell value
  - Re-design the sectors with the aim that each sector level cell has at least sufficient trip records to exceed the 30% threshold and no more than is necessary to achieve the 20% threshold
  - If necessary, define movements (rather than sector level cell values) in terms of one sector to many others, many sectors to one sector, or many sector to many others
  - Vary the sector and movement definitions by purpose, as necessary
  - Use the resulting estimates of trips making the defined sector level movements as constraints in the matrix synthesis (Step 6)

Step 6: synthesis of complete trip matrices (1)

- Complete trip matrices need to be synthesized because
  - the trip matrices derived from the RSI survey data are partial and estimates of the movements not intercepted in the surveys are required
  - at the zonal level, the sampling variability of the ‘observed’ (partial) trip matrices will be very large and some means of smoothing out that variability is required

Step 6: synthesis of complete trip matrices (2)

- Complete trip matrices may be synthesized using
  - either a gravity model
  - or a destination choice model
- The aims are the same, namely to calibrate model parameters so that
  - the partial matrix trip cost distributions are replicated
  - the statistically reliable sector level movements in the partial matrices are reproduced
- The assumption is made that a model which meets these aims will provide satisfactory estimates of non-surveyed movements
- The calibrated distribution model is to be used only to synthesise base year matrices and is not intended to be used for forecasting future year demands

Step 6: synthesis of complete trip matrices (3)

- General principles
  - Only inputs to the distribution model which are acceptably accurate should be used
    - The corollary is that use of inputs that have been developed by applying inaccurate factors should be avoided - this consideration applies particularly to the level of segmentation of the trip ends
  - The segmentation of the trip distribution model should be commensurate with the level of the segmentation of the inputs that can be derived with adequate accuracy
    - This means that a high level of segmentation should not be used in the distribution model if adequately accurate trip ends cannot be produced at that high level of segmentation
    - The distinctions in model parameters for the various segments in such a model may be spurious and at least partly due to errors in the trip ends

Step 6: synthesis of complete trip matrices (4)

- General principles (continued)
  - The RSI data should be used in a statistically sound way
    - This means that the data should not be disaggregated by sector, time period and trip purpose to the point where the matrix cell values are not statistically reliable.
    - Because spatial detail is important for the trip matrices in an assignment model, priority should be given to the detail of the sector system over the other two dimensions.
  - Use of inaccurate factors should be avoided in the derivation of the matrices for assignment from those created by the distribution model
    - Working in units of person trips in PA format at the 24-hour level will require factors to convert to vehicle trips in OD format for the assignment hours

Step 6: synthesis of complete trip matrices (5)

- Options for trip synthesis
  - For periods (morning peak, inter-peak, evening peak), or 12-, or 16- or 24-hour days
  - By trip purpose, either the normal home-based employers' business, commute, education, shopping, and other and non-home-based employers' business and other or some grouping of these segments (such as assignment user classes)
  - In either person trips or vehicle trips
  - In either OD or PA format
- The choice depends only on the availability of reliable inputs

Some Refinements (1)

- Trip ends may only be available for in-work and out-of-work trips (rather for the full set of trip purposes conventionally used in trip matrix synthesis)
- Sector systems used to define statistically reliable movements are likely to be quite coarse for some of the smaller demand segments (trip purposes)
- These two considerations suggest that prior trip matrices should be developed for two demand segments only: in-work and out-of-work (on the grounds that these are the segments which need to be assigned separately)
- Tests have shown that the accuracy of the fully synthetic (or non-observed) movements would be much the same whether the gravity model was run for six or two purposes

Some Refinements (2)

- However, the first two points made on the previous slide remain:
  - the IWT trip ends are unlikely to be sufficiently accurate
  - the sector systems for reliable estimates of IWT movements are likely to be very coarse
- It is quite likely, therefore, that adjustments to IWT component may be required at some stage in order to improve the synthesis, and that some iteration will be required to find the best set of adjustments
- A simpler approach is to conduct the gravity model work at the total purpose level and to apply the IWT/OWT split as late as possible in the process, so that variations in that split can be tested easily

Some Refinements (3)

- Tests have shown that there would be little loss of accuracy in running the gravity model for all purposes combined as opposed to running the model for IWT and OWT separately
- The IWT/OWT splits can be derived for the most detailed sector system for which the IWT movement estimates are statistically reliable: although these will be approximate because of the coarseness of the sector system, because they are applied late in the process (just prior to assignment), they can be varied easily

Some Points About New Data (1)

- These comments should not be regarded as either comprehensive or necessarily definitive
- Consider three ‘new’ data sources:
  - Mobile phones
  - GPS tracking systems
  - Satellite navigation devices
- OD data can be obtained from all three sources but how should these data be used?

Some Points About New Data (2)

- The data from all three ‘new’ sources are samples, so the questions that arise include:
  - Are the samples biased and, if so, how may they be corrected?
  - How may the samples be expanded so that they represent total movements
- The further question that arises is:
  - How should data from the ‘new’ sources be merged with the prior trip matrices constructed in the conventional way from RSI data and trip synthesis?

Some Points About New Data (3)

- Mobile phone data may be available for selected (but not all) service providers, so we would need to know:
  - whether the profile of the users of the selected service provider(s) differs from the population as a whole (for bias)
  - whether there is a material proportion of mobile phone owners who do not always carry their phone in switched on mode (for bias and expansion)
  - how the sample of mobile phones providing records relate to the total travelling population (for expansion)

Some Points About New Data (4)

- GPS tracking data are available for only a sample of vehicles, so we would need to know:
  - How the sample of tracked vehicles relates to the total population of travelling vehicles (for bias and expansion)

- Satellite navigation data are likely to be very biased because the devices are likely to be used only for infrequent and unfamiliar trips – regular and familiar trips will not be recorded

Some Points About New Data (5)

- I would say:
  - Do not consider using satellite navigation data for OD trip matrices because the data are likely to be very biased
  - Check the patterns of trips derived from mobile phone and GPS tracking data against independent sources, such as RSI survey data, and consider adjustments if significant biases are evident
- I would also say that expansion of samples of trip records from mobile phones and GPS tracking devices is likely to be extremely difficult
- Research is in hand by a team led by Atkins for the HA into ways in which the sample data from the 'new' sources may be used to enhance conventionally constructed prior trip matrices