Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 26 (2017) 39–52

Transportation Research Procedia

www.elsevier.com/locate/procedia

44th European Transport Conference 2016, ETC 2016, 5-7 October 2016, Barcelona, Spain

Origin-Destination Trip Matrix Development: Conventional Methods versus Mobile Phone Data

Reza Tolouei[a,*], Stefanos Psarras[a], Rawle Prince[a]

[a]*Model Development and Research, Transportation, AECOM, St. Albans,United Kingdom*

Abstract

Conventionally, trip matrices have been derived by a combination of roadside interviews (RSIs) and the application of trip-end and gravity models (to extrapolate and infill unobserved movements), followed by matrix estimation methods to incorporate evidence from supplementary traffic counts. More recently, mobile phone positioning data are being used increasingly by the transport planning community to develop 'prior' demand matrices as an alternative approach to RSI data or synthetic methods. There are a number of known strengths and weaknesses associated with each of these approaches. However, there is lack of robust evidence to suggest whether use of any of these approaches results in a matrix that performs better (or worse) overall. This study provides such evidence through a structured and systematic comparison of trip matrices developed using mobile data, RSI data, and a gravity modelling approach (i.e. synthetic matrices). In addition to comparison of assigned flows with traffic counts across a range of independent screenlines, the following aspects of the matrices are also compared with independent observed data: 1) correlation of trip-ends with estimates based on the UK National Trip-End Model (NTEM); 2) consistency of trip rates with estimates based on Great Britain National Travel Survey (NTS) data; and 3) comparison of trip length distributions with estimates from the NTS. The results suggest that, overall, the outcome of using the mobile phone data, when systematically refined and adjusted using independent data sources to address various known limitations and biases, does not seem to be either biased or less accurate than conventional methods. It has also been observed that areas of the model where no RSI data or other similar observed data are available, use of mobile data could result in a more consistent estimate of trips, benefiting from a significantly larger sample size.

Peer-review under responsibility of the Association for European Transport

Keywords: Trip matrices; RSI data; mobile phone data; Gravity models

* Corresponding author.
E-mail address: reza.tolouei@aecom.com

2352-1465 © 2017 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of Association for European Transport
10.1016/j.trpro.2017.07.007

1. Introduction

Conventionally, trip matrices have been derived by a complex process involving a combination of methods. This includes the use of roadside interviews (RSIs) to observe movements across defined screenlines, the application of trip-end and gravity models to extrapolate and infill unobserved movements, followed by matrix estimation methods to incorporate evidence from supplementary traffic counts. This process is widely accepted as the preferred approach to using synthetic matrices. More recently, mobile phone positioning data (referred to as 'mobile data' in this paper) are being used increasingly by the transport planning community to develop 'prior' demand matrices as an alternative to the RSI and synthetic matrix development approach.

There are a number of known strengths and weaknesses associated with both RSI data and mobile data, which have been documented in detail in a number of recent studies (for example, see Tolouei, et. al., 2015). However, there is lack of robust evidence to determine whether the use of mobile data results in matrices that perform better or worse overall than matrices developed using alternative, conventional methods (i.e. RSI data or synthetic methods based on gravity modelling). In this study, we provide such evidence through a structured and systematic comparison of two sets of trip matrices: matrices developed using mobile data, and matrices developed using a combination of RSI data and a gravity modelling approach (i.e. synthetic matrices).

Origin-Destination (OD) matrices estimated from mobile phone data have potentially certain strengths compared to conventional sources of OD information such as RSI data. These mainly include wider geographical coverage, higher sample size, capturing day-to-day variability of trips, and potential time and cost savings for data collection and processing.

However, this is a relatively new type of data which are not collected exclusively for the purpose of transport planning. There are therefore key weaknesses and uncertainties associated with OD matrices derived from mobile phone data which should be recognized and addressed. These include definition of trips and trip-ends, spatial resolution and data accuracy, identification of short trips, identification of vehicle types and vehicle occupancy, identification of trip purpose and mode, and expansion of mobile data. A comparison of the characteristics of RSI data and mobile data, sourced from an earlier paper (Tolouei, et. al., 2015), is shown in Table 1.

Table 1: Possible Drivers of Trends in Trip Rates

Attribute / Consideration	RSI Data	Mobile Phone Data
Type of raw data	Cross-sectional (a sample from a single day)	Longitudinal (cross-sectional data collected over a period of time)
Sampling approach	Specified locations for selected roads; Random sample of drivers at these locations	Full population of Operator's subscribers
Sample rate (for a given road)	10% to 20% (individual sample)	~30% (repeated sample over several days)
Variation of trips observed in the data	Spatial variation	Spatial and temporal variation
Data bias	Potential for response bias, this could be minimised through careful survey design and sampling strategy	Potential for bias towards the profile of 'subscribers' if different, bias could be corrected largely if identified properly
Expansion of data	Relatively straightforward using count data and statistical analysis where journeys traverse more than one sample site.	More complicated, requiring information on how the mobile phone users relate to total population
Identify trip purposes	Straightforward; survey question	Need to be inferred through assumptions/rules/other data sources (including RSIs if available).
Identify vehicle type	Straightforward; survey observation	Need to be inferred through assumptions/rules/other data sources (including RSIs if available).
Identify vehicle occupancy	Straightforward; survey observation	Need to be inferred through assumptions/rules/other data sources (including RSIs if available).
Geographical scope of data	Only those movements intercepted by screenlines / cordons (see Figure 1)	In theory all movements, though short trips may be omitted
Proportion of unobserved OD trips in the matrix	Relatively large, depending on number of RSI sectors	None or very low (short trips)

The main objective of this study is to compare the performance of trip matrices that are developed using conventional methods (i.e. using RSI data) with those based on emerging techniques using mobile phone data.

Different aspects of these matrices have been compared against various data sources, and the consistency of trip patterns in the developed matrices has been examined.

2. Study Data and Approach

AECOM has recently undertaken a major update to the Leicester and Leicestershire Integrated Transport Model (LLITM) to develop the LLITM 2014 Base model. As part of this update, highway demand matrices have been developed using mobile data (processed by Telefonica into origin-destination trip matrices). For the purpose of the study, and prior to the development of mobile data matrices, Leicestershire County Council commissioned a significant programme of data collection, including RSI data for 155 sites (approximately 100 sites were surveyed in 2013/2014), local planning data, household survey data, and extensive traffic count data. The presence of these data sets provided a unique opportunity to develop alternative trip matrices using the RSI data, as well as purely synthetic matrices, and to compare these with the mobile data matrices, all of which have a consistent geographical scope and represent demand pattern for the same time period. The methodology used to develop these matrices is discussed in the next section.

The LLITM highway assignment model is used to assign the two developed matrices (i.e. RSI-based matrix and mobile data matrix) onto the highway network. The assigned flows are compared with traffic counts across a range of long screenlines and cordons. Whilst this was the key focus of the study, the following aspects of the matrices were also compared with independent observed data:

- correlation of trip-ends with estimates based on the National Trip-End Model (NTEM);
- consistency of trip patterns between RSI and mobile data matrices; and
- comparison of trip length distributions with estimates from the NTS.

Various sources of data used to undertake the above analyses are described below.

2.1. Sources of Data

The following sources of data have been used to develop and evaluate the OD matrices discussed above:

- Mobile phone positioning data;
- Roadside interview data (RSI);
- Traffic counts;
- UK National Travel Survey (NTS); and
- National Trip-End Model (NTEM).

The primary source of LLITM prior matrices are mobile phone data, processed into origin-destination matrices by Telefónica. AECOM and Telefónica have worked together collaboratively to develop and refine the approach to produce highway demand matrices from mobile data and to address some of the known limitations. More details on the processing and verification of mobile phone data is given in Tolouei, et. al., 2015.

UK NTS data was used to provide various estimates used to both develop and validate the matrices. Trip length distribution from NTS was used to calibrate the synthetic matrices, and to verify trip pattern in both mobile data and RSI data matrices. Trip rates were used to verify and adjust the mobile data matrices. NTS was also used to estimate the return time probabilities, used to estimate trips in non-interview directions in the processing of RSI data. The next section provides more details on these processes.

The LLITM data collection programme included collection of traffic count data for nearly 665 sites. A subset of these was used to expand the RSI data, whilst others were used during the calibration and validation of highway matrices.

NTEM produces forecasts of transport planning information, including estimates of trip ends for a range of years. NTEM 6.2 was used as the basis of providing trip-end estimates, used in the development of synthetic matrices. A bespoke version of NTEM was developed for this purpose, with updated local planning data and disaggregated into model zones.

3. Matrix Development Approach

As noted earlier, alternative trip matrices were developed from mobile data and RSI data (referred to as “mobile data matrices” and “RSI matrices” in the remainder of this paper). Synthetic matrices were also developed separately, which are usually required in order to complement both RSI and mobile data matrices. The process of developing these matrices is summarised below.

3.1. Synthetic Matrices

Synthetic matrices are commonly developed for transport models to complement observed trip information such as roadside interviews. While they may be based on reliable observed data about trip making behaviour and land use information, they only estimate the likely trip patterns. The absence of observed movements means they are referred to as synthetic matrices. In brief, the process firstly requires the preparation of a set of inputs for a distribution model. These inputs take the form of an observed cost profile, a cost skim of the modelled area and a set of trip-ends for each zone. The distribution model calibration process is then undertaken and its outputs converted into car assignment matrices. The process is run separately for each trip purpose, time period and direction.

In order to develop synthetic matrices for LLITM, two sources of data were required for each LLITM segment: estimates of trip origins and destination in each model zone, and trip length distribution. As mentioned earlier, trip-ends estimated based on NTEM 6.2, separately by time period and trip purpose, as well as observed trip length distribution, estimated from the NTS, were used.

To develop the synthetic OD matrices for each segment in LLITM zones, a gravity modelling approach was used. The deterrence function was estimated based on RSI data and using a log-normal density function. A doubly constrained matrix balancing process was used, and iterations were performed to update the deterrence function parameters so that the final synthetic matrices by purpose match the total origins and destinations (doubly constrained), and also the trip length distribution of the observed data. It should be noted that no sector constraints (also known as k-factors) was used in the calibration process.

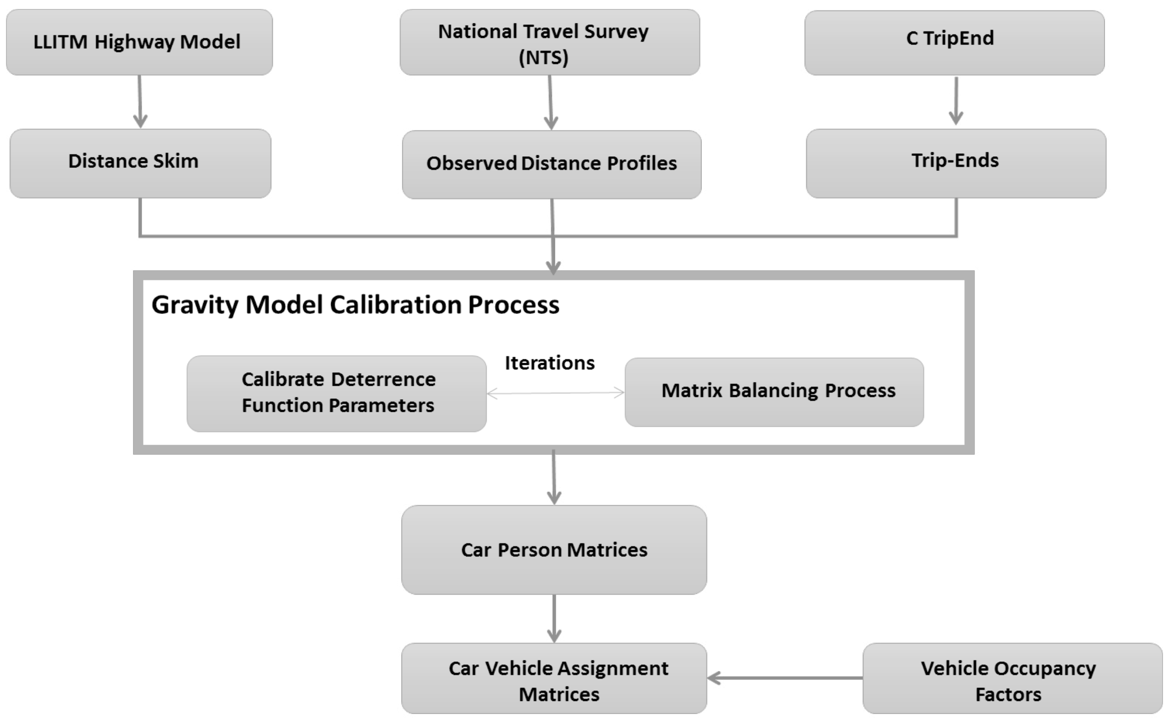

Figure 1: Overview of Synthetic Matrix Development Process

3.2. Mobile Phone Data Matrices

The data is primarily sourced from mobile phone data, processed into OD matrices by Telefonica. Following a detailed verification process, the matrices were subject to a detailed disaggregation, segmentation, and adjustment process before they were used in LLITM as prior matrices. The methodology to develop LLITM prior matrices from mobile phone data is detailed in a separate paper (Tolouei, et. al., 2015), and therefore is not discussed in this paper. Nevertheless, the key steps taken to develop these matrices are shown in Figure 2 .

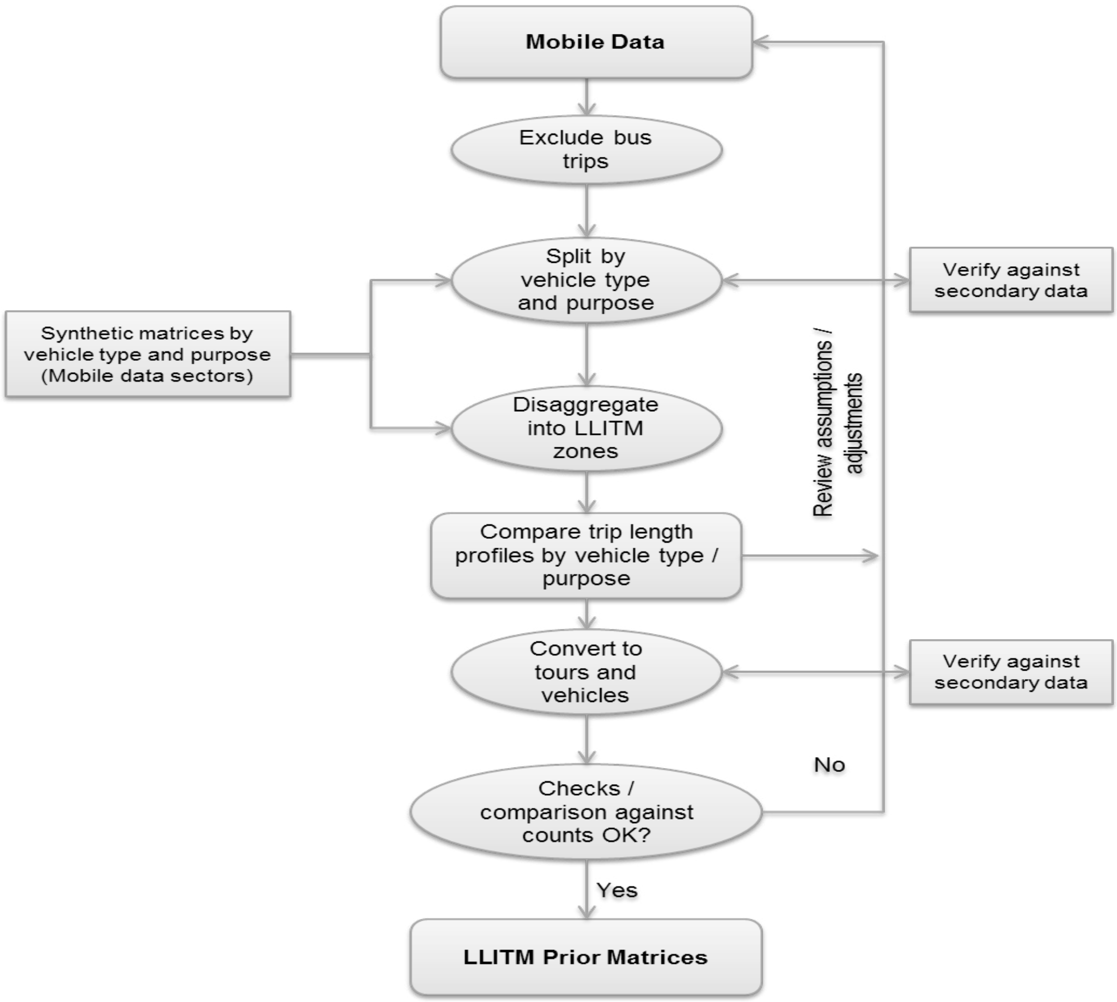

Figure 2: Overview of Prior Matrix Development Process based on Mobile Data.

As the figure shows, following the segmentation and disaggregation process, the matrices have been through an iterative refinement process where, depending on the outcome of verification process, a number of adjustments were made to the matrices, intended to correct for various shortcomings in the data. The matrix adjustments were undertaken in two different stages. In the first stage, the intension was to correct various biases using evidence from secondary data which were independent from traffic count data. The key adjustments in this stage included replacement of short trips (which are not fully represented in the mobile data) from synthetic matrices, and adjustments made to the trip rates, based on the NTS, to correct for possible expansion issues.

If, following the stage 1 refinements, the performance of the matrices across screenlines was not shown to be up to the standard recommended by WebTAG before they can be used for calibrating the highway model (i.e. all or nearly all modelled flows being within ±5% of count data across screenlines); further refinements were required. The intention was to address various errors in the estimated trip patterns, some of which were briefly discussed earlier in this paper. In order to refine the prior matrices and address the above issues, a methodology was developed to adjust the inter-sector movements based on an initial comparison of total observed and modelled flows across screenlines and cordons. In this update, all counts along a given screenline were used as a single constraint. This minimised the impact of any localised routing issues in the model at the time, and the results of this process were used to update the matrices at a sector level. An important issue that should be taken into consideration is the resulting changes in the developed matrices. The adjusted matrices must retain as much of the information as possible from the observed data. The application of a sector-based factoring, as compared to a cell-based factoring, would significantly reduce changes to the developed matrices and would retain as much of the information as possible from the observed data. It is also noted that the sectorised updates should be applied after the highway assignment network has been reviewed, and any significant routing discrepancy has been removed.

3.3. RSI Data Matrices

The RSI data for 159 sites were used to develop partially-observed matrices which were then merged with the synthetic matrices, following conventional methods and current best practice documented in the UK (Coombe, 2011), to develop alternative highway prior matrices for LLITM. Whilst it is not within the scope of this paper to describe in detail the process of developing prior matrices from the RSI data, below provides a brief summary of the approach taken.

The process of deriving the highway prior matrices for LLITM is outlined in Figure 3.

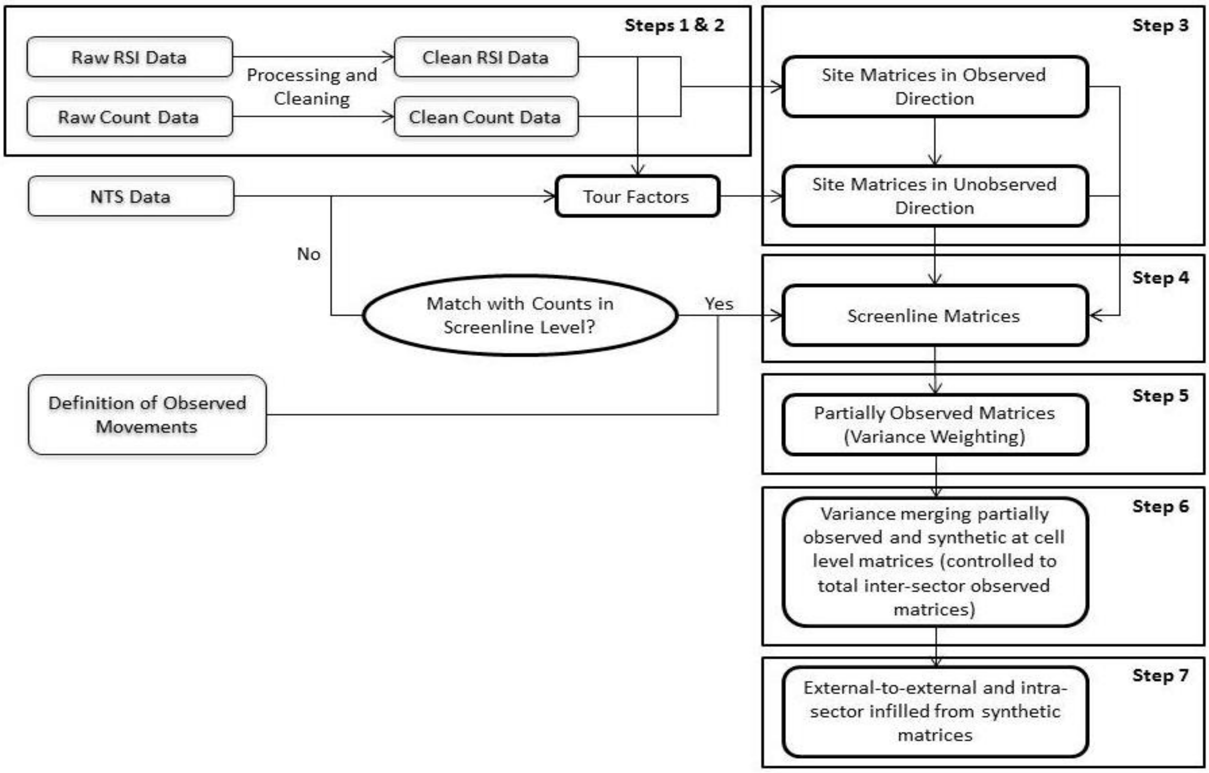

Figure 3: Overview of Prior Matrix Development Process based on RSI data

In summary, the following steps have been taken.

RSI data have been processed to create demand matrices for each RSI site. These have been built by purpose of travel, vehicle type and time period, as appropriate. The site matrices have been expanded to traffic counts and then combined across cordons and screenlines to create cordon and screenline demand.

Generating expanded trip matrices for each RSI site in the observed direction is relatively straightforward; however, a separate process has been used to estimate the OD trips for the unobserved direction using information on OD patterns in the observed direction, counts in the unobserved direction, and trip outbound / return time probabilities, estimated from the NTS data.

Proxy site data have been used where no RSI data exist, creating pseudo-sites. Most of these inherit data from nearby sites to create 'proxy' sites. When data were not available for a nearby site, select link analysis has been used to estimate trip distribution, before being expanded to the count data.

The cordon and screenline matrices have been merged using a variance-weighting method to eliminate double-counting in such a way as to place most confidence in estimates with larger samples. Other characteristics of the data such as age, direction observed in, and type of count used to expand have been also used in calculating variances.

The observed RSI and synthetic matrices have been merged using the variance-weighting method at OD level, constrained to total inter-sector trips from the RSI data. For this, an estimate of the variance of observed and synthetic demand was required to be estimated. The variance of the observed trips has been calculated using the same methodology used in the partially-observed matrix development (sourced from DMRB, 1999), which is estimated by:

$$\boldsymbol{\sigma_{ij}^2 = O_{ij}\left({}^{E_{ij}}\!/_{O_{ij}} \right)^2}$$

where:

- σ_{ij}^2 is the variance of observed vehicles between zones i and j;
- O_{ij} is the number of observed vehicles between zones I and j; and
- E_{ij} is the number of expanded vehicles between zones I and j

The variance of the corresponding synthetic matrices has been estimated using the findings of research carried out by Haskey (1973) which indicated that trip synthesis using gravity model introduces additional error with a coefficient of variation of approximately 0.5. On this basis have estimated that the variance of the synthetic trip matrix is:

$$\boldsymbol{\left(\widetilde{\sigma_{ij}^S}\right)^2 = Var(S_{ij}) = \left(\sigma_{IJ}^T\right)^2 + \left(0.5\,.S_{ij}\right)^2}$$

where:

- $\left(\widetilde{\sigma_{ij}^S}\ \right)^2$ is the variance of the synthetic matrix between zones i and j;
- $\left(\sigma_{ij}^T\right)^2$ is the variance of the observed vehicles between zones i and j;
- S_{ij} is the synthetic trip matrix between zones i and j;

The relative variances from the two sources of data have been used to weight the demand from each source to create a merged demand matrix. Assuming that the distributions of variances are approximately normal, an application on Bayes's Theorem gives the following equation for merging the observed matrices:

$$\boldsymbol{\widehat{T_{ij}} = \frac{T_{ij}\left(\sigma_{ij}^S\ \right)^2 + S_{ij}\left(\sigma_{ij}^T\right)^2}{\left(\sigma_{ij}^S\right)^2 + \left(\ \sigma_{ij}^T\right)^2}}$$

where

- $\widehat{T_{ij}}$ is the merged matrix between zones I and j;
- T_{ij} is the pre-merged observed matrix between zones i and j
- S_{ij} is the pre-merged synthetic matrix between zones i and j
- σ_{ij}^T is the variance of the observed matrix between zones i and j; and

- σ_{ij}^2 is the variance of the synthetic matrix between zones i and j.

Following the merging of the observed and synthetic demand matrices, the merged demand has then been re-constrained to the sector to sector movements captured by the RSI surveys. By applying this constraint, it is ensured that the merged demand in the fully observed part of the matrix is consistent with the observed demand, which although subject to sampling and other errors, is an actual observation.
Synthetic matrices have been used to infill external-to-external trips as well as intra-sector movements within the matrices, which were not observed from the RSI data.

3.4. Comparative Performance of Matrices

This section summarises the key findings on the performance of the matrices developed using mobile data, as well as those based on conventional methods using a combination of RSI and synthetic data. Some key aspects of the matrices are compared with secondary data sources; these are described below. The results are presented for two modelled time periods: AM Period (07:00-10:00) or AM Peak (08:00-09:00), and PM Period (16:00-19:00) or PM Peak (17:00-18:00). It is also noted that, unless stated otherwise, the comparison results below are given for the matrices prior to any sectorised updates (see Section 3.23.2).

3.5. Trip Length Distribution

Figure 4 shows comparison of trip length distributions for total person trips obtained from mobile phone matrices, RSI matrices and NTS data in the AM and PM periods.

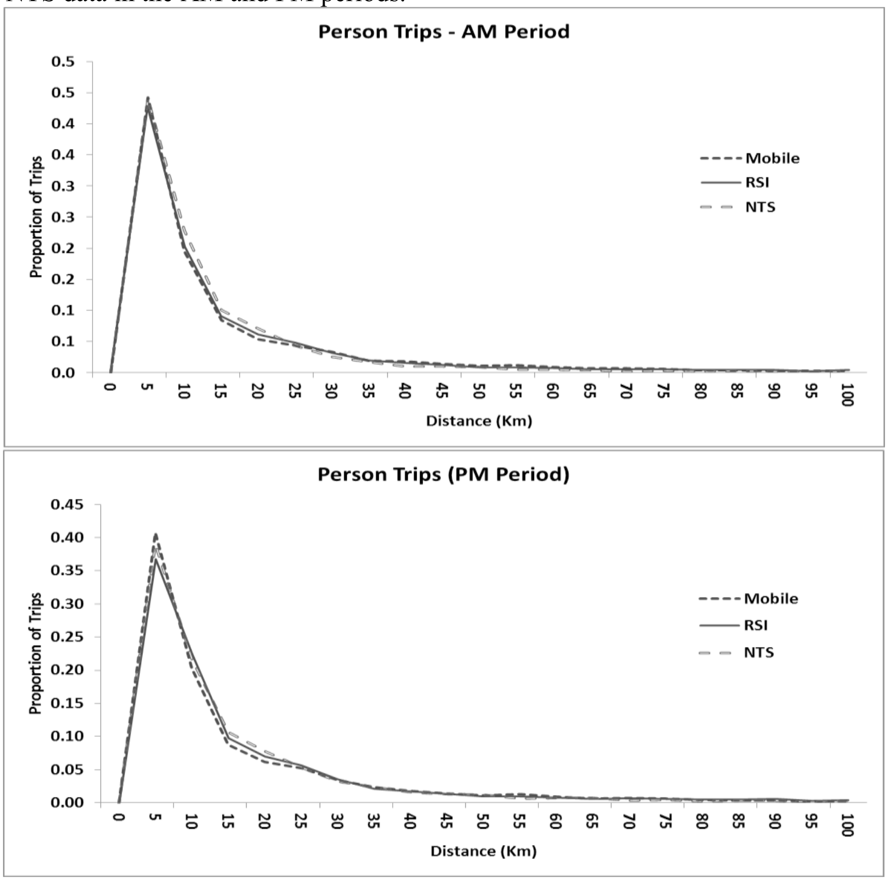

Figure 4: Comparison of Trip Length Distributions of Mobile Phone Matrices, RSI Matrices, and the NTS

In general, there is a good correspondence between the trip length distributions across all data sources, in both periods. It should be noted that in both RSI and mobile data matrices, short trips are mainly synthetic, which are calibrated to the observed trip length distribution from the NTS.

Figure 5 and Figure 6 show the correlation between all-day trip origins (total person trips), estimated from processed mobile phone data and RSI data, with data obtained from the NTEM, at model zone level. The results show that there is generally a good correlation between both mobile data and RSI data trip origins with the NTEM. It is important to note that in calculating trip origins, intra-sector trips have been removed; the estimates are therefore do not include synthetic trips and are independent from the NTEM.

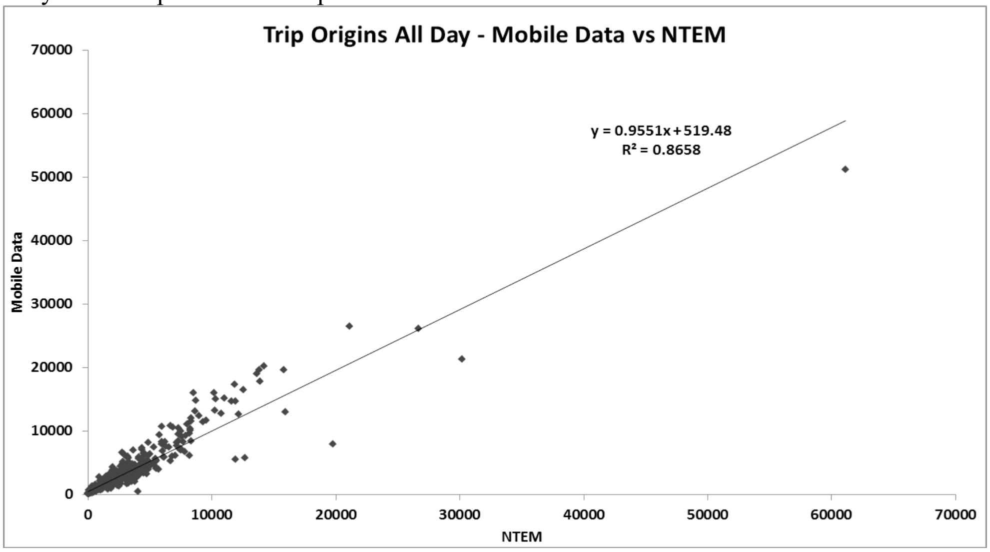

Figure 5: Relationship between NTEM and Mobile Data Trip Origins

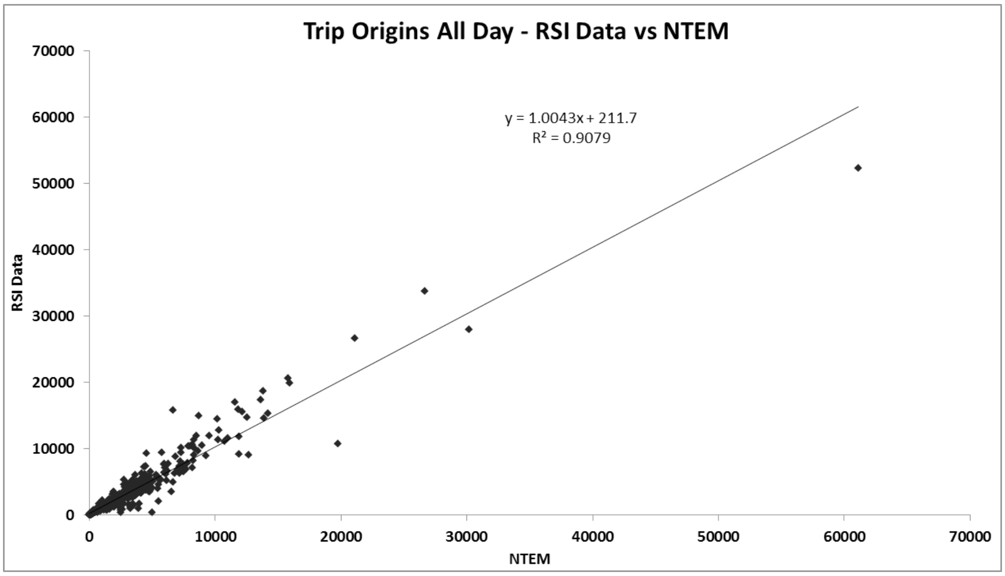

Figure 6: Relationship between NTEM and RSI Data Trip Origins

3.6. Sectored Matrices

The sectoring system defined to process the RSI data was used to compare trip patterns between the two sets of matrices. Figure 7 shows the relationship between trip origins from processed mobile phone and RSI data, in the defined 24-sector RSI sectoring system, for AM and PM periods. In general, there is a good correlation between the two sources of data, taking into account various sources of error and inconsistency between them.

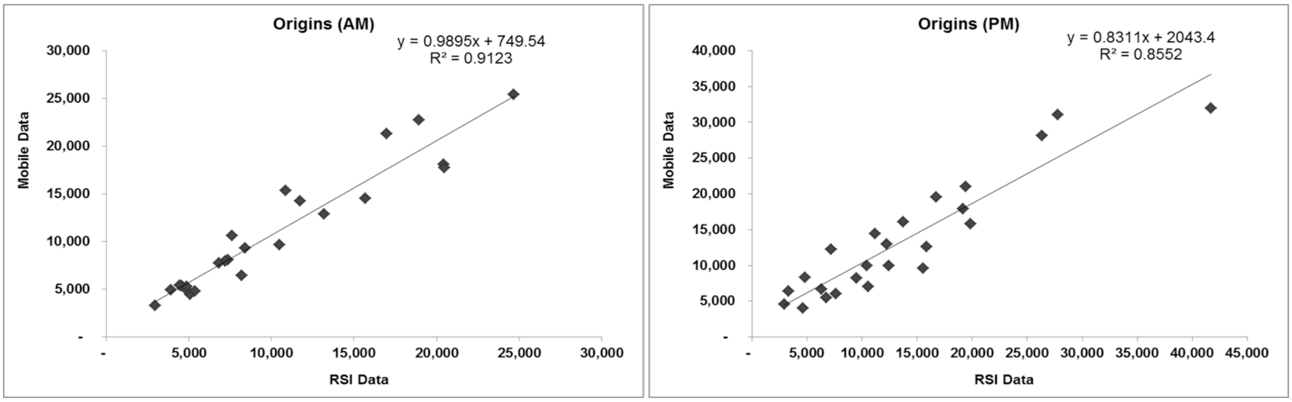

Figure 7: Comparison of Trip Origins between RSI and Mobile Data (AM and PM Periods)

In order to verify whether patterns of trips derived from mobile phone data are consistent with those from RSI trips, inter-sector trips were compared between the two matrices for AM and PM periods. The results for the 24 RSI sectors are shown in Figure 8.The results suggest that there is a reasonable correlation between mobile phone data and RSI data in terms of trip distribution pattern.

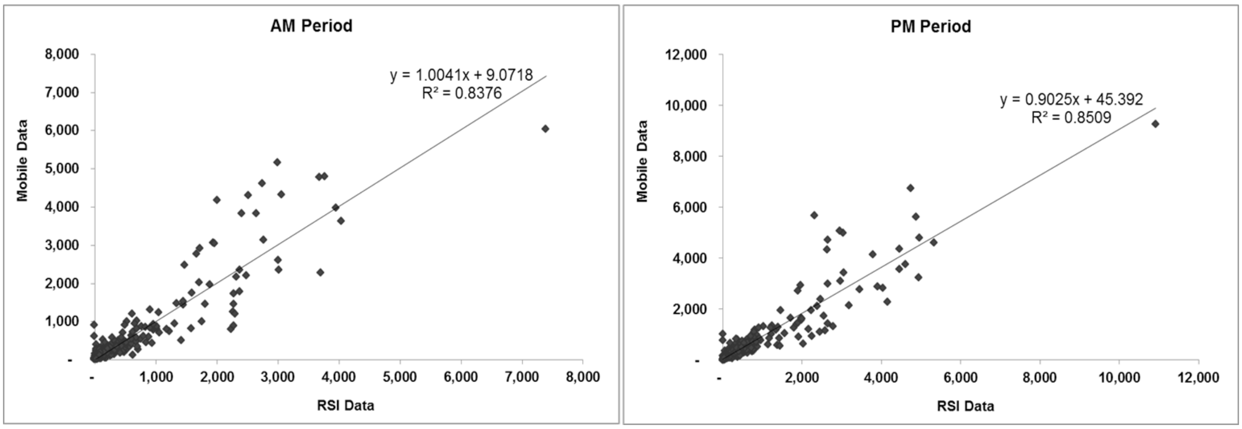

Figure 8: Comparison of Inter-sector Person Trips between RSI and Mobile Data (AM and PM Periods)

3.7. Screenline Performance against count data

The matrices were assigned on LLITM highway network and the modelled flows were compared with the traffic count data at screenline level. The comparisons were made against two groups of screenlines: those used to form the RSI cordons and those used for the purpose of calibration and validation (i.e. not used in the RSI processing); these are shown in Figure 9.

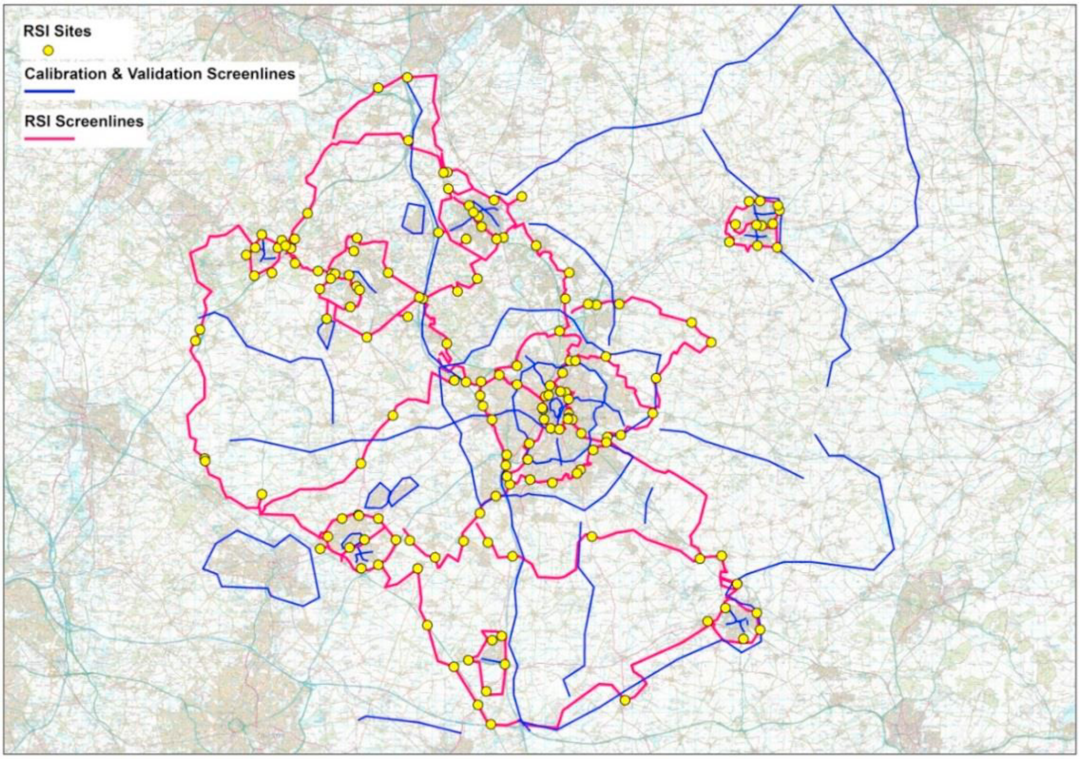

Figure 9: Overview of RSI and Calibration / Validation Screenlines

Table 2 summarises the performance of the matrices, measured as the proportion of total screenlines that pass UK WebTAG's calibration criteria (modelled flows being within 5% of traffic counts across screenlines). The results are shown for both sets of matrices (i.e. built from mobile data and RSI data), separately by screenline classification, for AM and PM peak periods for total vehicle flows.

Table 2: Proportion of Screenlines Passing WebTAG's 5% Test Criteria: RSI Matrices vs. Mobile Data Matrices

Screenline Classification	Number of Screenlines	RSI Data (AM)	Mobile Data (AM)	RSI Data (PM)	Mobile Data (PM)
RSI Screenlines	38	55%	53%	58%	47%
Cal/Val Screenlines	74	43%	55%	38%	49%
Total	112	46%	54%	45%	49%

For the RSI screenlines, matrices sourced from RSI data perform relatively better than mobile phone data matrices. This outcome was partly expected as RSI matrices for the majority of trips crossing these screenlines consist of observed trips, consistently expanded to the count data. The same counts are used for the RSI screenlines, so RSI matrices are expected to better match flows at these screenlines compared to the matrices built from mobile data. However, it should be noted that there are still errors of different sources in trip estimates based on RSI data; these include sampling error, assumptions made to estimate trips in unobserved directions, and errors in estimating trips using "links" crossing screenlines where no roadside interview has taken place (i.e. RSI 'holes').

Against the screenlines that have been defined for the calibration and validation process, mobile matrices show a better performance than RSI matrices. It should be noted that the former have gone through a process of adjustments and refinements using independent sources of data to address known limitations of mobile phone data. On the other hand, a large proportion of trips crossing these screenlines in the RSI matrices are unobserved trips which are sourced from synthetic matrices. Moreover, for many screenlines, especially those in close proximity of major motorways, the proportion of the through traffic in the matrices is not negligible. A high proportion of these are external to external trips which are infilled from synthetic matrices. The above arguments and the relative poor performance of synthetic matrices, together explain why RSI-based matrices perform worse than mobile matrices against the calibration and validation screenlines, which form the majority of screenlines.

3.8. Sectorised Refinements

Overall, the performance of both sets of matrices across screenlines, as shown in Table 3, is not up to the standard recommended by WebTAG before they can be used for calibrating the highway model; hence further sector-based refinements were undertaken to address remaining errors, as described in Section 3.2.The sectorised refinements were applied to both set of matrices built from mobile data and RSI data. The overall performance of the refined matrices at screenline level is summarized in Table 3.

Table 3: Proportion of Screenlines Passing WebTAG's 5% Test Criteria after Sectorised Refinements: RSI Matrices vs. Mobile Data Matrices

Screenline Classification	Number of Screenlines	RSI Data (AM)	Mobile Data (AM)	RSI Data (PM)	Mobile Data (PM)
RSI Screenlines	38	89%	89%	95%	92%
Cal/Val Screenlines	74	96%	98%	97%	94%
Total	112	94%	96%	96%	93%

As the results show, the performance of both sets of matrices has been improved significantly against the screenlines. The results and the comparisons between the two set of matrices provide similar insights and conclusions as the ones shown in Table 4. RSI matrices perform similar (AM) or better (PM) than the mobile data matrices against the RSI screenlines. Overall, following the sectorised updates, both sets of matrices show an acceptable performance when compared with the count data.

As noted earlier, it is important to closely monitor changes to the matrices as a result of sectorised refinements. Changes to number of trips at matrix cell level, as well as trip-end level, as well as changes in trip length distribution of matrices were analysed by comparing the relevant statistics before and after sectored updates. For these analyses, only a subset of trips with an origin in the internal model area was used. Table 4 shows the regression analysis statistics for matrix zone changes. In all cases, the slope is relatively close to 1 and R2 is about 0.9, suggesting limited changes in number and patterns of trips.

Table 4: Regression analysis of Matrix Zone Changes (Leicestershire Origins)

Regression Statistics	RSI Matrix (AM)	Mobile Matrix (AM)	RSI Matrix (PM)	Mobile Matrix (PM)
Intercept	0.00	0.00	0.00	0.00
Slope	1.06	1.04	1.04	1.04
R Squared	0.89	0.90	0.92	0.89

Figure 10 and Figure 11 show scatter plots of changes in trip origins for RSI matrices and mobile data matrices respectively, for trips originating from the model internal area. A strong correlation and a slope close to one suggest that overall, sectorised refinements have resulted in limited changes to total and distribution of trip-ends.

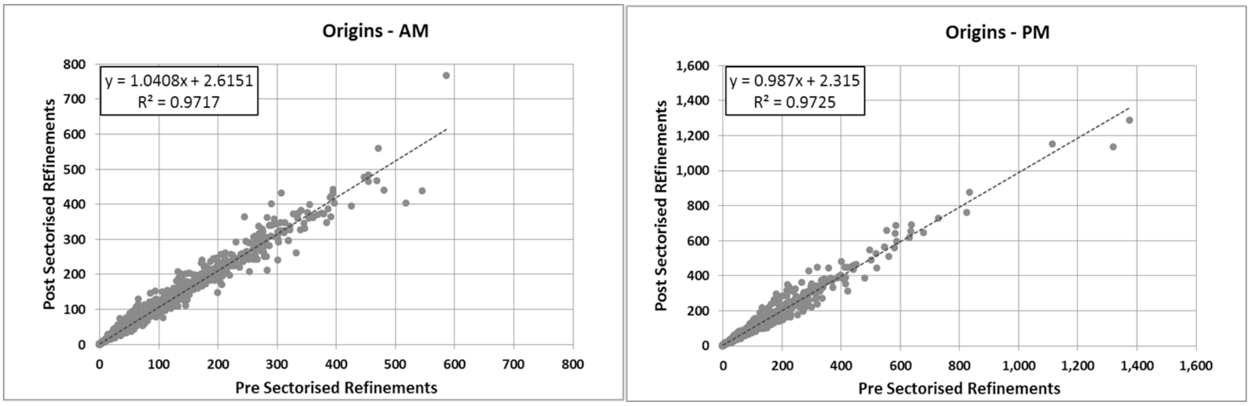

Figure 10: Analysis of Matrix Trip-end Changes for RSI Matrices

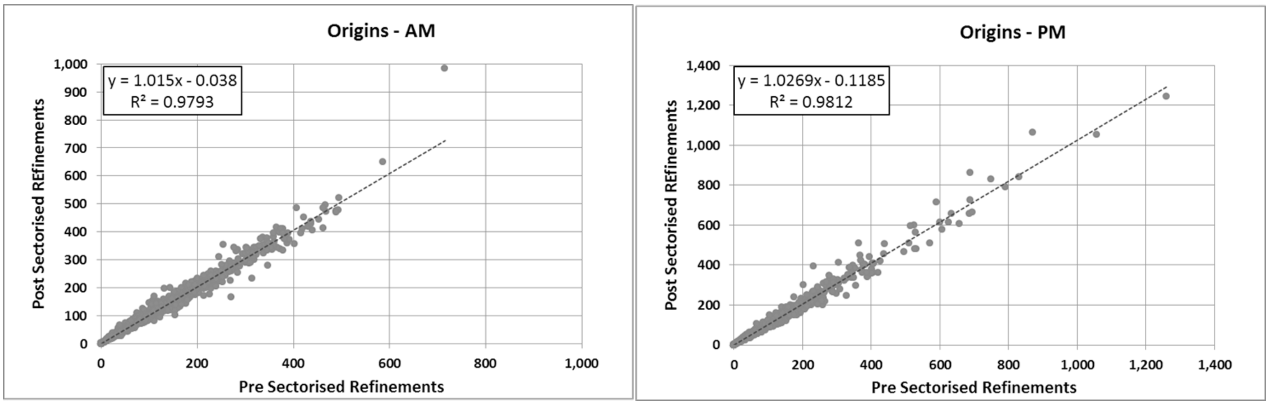

Figure 11: Analysis of Matrix Trip-end Changes for Mobile Data Matrices

It is noted that trip length distributions before and after sectorised refinements were compared as well. Changes in average and standard distribution of trip lengths were less than 5% for both set of matrices.

4. Summary and Conclusions

Mobile phone data have been used as the primary source of data to develop the highway prior matrices for the LLITM model. The data have been subject to a detailed verification process, as well as a number of refinements and adjustments to address various known limitations of the mobile data. These processes are described in detail in an earlier paper (Tolouei, et. al., 2015), as well as (to a lesser extent) in Section 3.2. The availability of detailed RSI data for about 150 sites in the study area provided an opportunity to develop alternative trip matrices using conventional matrix development techniques, and to compare their performance against mobile data matrices

In summary, mobile data and RSI data matrices both show a reasonable level of consistency with independent data in terms of trip length distribution and trip rates (through estimates of trip-ends). Furthermore, the number and pattern of inter-sector trips were found to be reasonably consistent and highly correlated between mobile data and RSI data matrices.

To further assess the performance of the matrices, the modelled flows were compared with traffic count data across a range of screenlines, separately for the screenlines directly used for RSI matrix development and the screenlines defined for calibration and validation purposes. The results showed that for the RSI screenlines, as expected, the RSI matrices perform better than mobile matrices. This is mainly the result of the fact that most of these trips are sourced from the observed RSI data, expanded to the same traffic count data used for comparison. However, there were still some differences with the count data, explained by various errors in the matrices as briefly discussed in Section 3.7.

On the other hand, for the rest of screenlines, modelled flows from mobile data perform better, showing a closer match to the count data, compared to those from the RSI matrices. For these screenlines, the majority of trips in the RSI matrices are either not intercepted by the roadside interviews and hence are synthetic, or have been merged with the synthetic demand at matrix cell level. They therefore include a large proportion of trips that are taken from the synthetic matrices, which are subject to relatively large errors. Trips sourced from mobile data, after a number of adjustments to address the weaknesses in the underlying data, perform considerably better, showing more consistency with traffic count data. This is mainly due to a significantly larger sample size of person trips in the mobile data matrices at matrix cell level, providing a full geographical coverage. In the case of LLITM, and probably the majority of similar models where RSI data and conventional methods are used to build the matrices, a large proportion of trips in the study area are not observed by the roadside interviews, and hence are synthesised.

The results reported in this paper suggest that, overall, the outcome of using the mobile phone data, when systematically refined using independent data sources to address various known limitations, does not look to be either biased or less accurate than the conventional method, using RSI data. In the areas of the model where no RSI data or other similar observed data are available, use of mobile data could even result in a more consistent estimate of trips, benefiting from a significantly larger sample size.

References

Coombe, D (2011) *Highway Trip Matrix Development: Current Best Practice* [PowerPoint slides]. Retrieved from: https://www.yumpu.com/en/document/view/39258896/highway-trip-matrix-development-current-best-practice-saturn

Haskey, J. (1973) *A General Investigation into the Efficiency of Trip Distribution Models*, MAU Note 244..

Great Britain Department for Transport (1999), *Design Manual for Roads and Bridges (DMRB)*, *Volume 12, Section 1, Part 1 (The Traffic Assessment Manual), Appendix D13.* [Online]: http://www.standardsforhighways.co.uk/ha/standards/dmrb/vol12/section1.htm

Tolouei, R., Álvarez, P., Duduta, N. (2015) Developing and Verifying Origin-Destination Matrices using Mobile Phone Data: The LLITM Case. P*roceedings of the European Transport Conference, Frankfurt*.