Towards the implementation of an activity-based travel demand model for London: Development of the household activity scheduling module

Scott Le Vine, Dr. Aruna Sivakumar, Professor John Polak
Centre for Transport Studies
Imperial College London
Professor Matthew Roorda
University of Toronto
UTSG Conference, 7 January 2010
Activity based modelling

- Classical transport analyses: people’s trips as unit of analysis

- But we know travel is [in general] a “means to an end” – a derived demand

- Through activity modelling, we seek to model the “ends” [people’s activities, distributed in time and space]
  - Where? -- e.g. SW10 1AA
  - When? -- e.g. 8:30 – 16:00
  - Why? -- e.g. Care for granddaughter

- Infer travel from a person’s “activity pattern”
Motivation

- Urban Energy Systems research project
  - Social & economic behaviour \(\rightarrow\) energy demands
  - Cross-sectoral linkages (retail products, transport, in-home consumption, commercial processes, etc.)
  - Optimisation, through understanding

- Choice of activity modelling
  - Thought to more closely mimic travel behaviour than alternative [trip-based] techniques
  - Heightened sensitivity to policy agenda
TASHA model

• Travel Activity Scheduler for Household Agents [TASHA]

• Simulates people’s out-of-home activities & travel during a representative weekday [24-hour period]

• Activity generation – What? How often?

• Activity scheduling – When? For how long?

• “Joint” activities with other household members
TASHA’s scheduling mechanics

(a) Draw activity frequency from marginal PDF

(b) Draw activity start time from feasible region in joint PDF

(c) Draw activity duration from feasible region in joint PDF
TASHA’s scheduling mechanics (2)

• Activity “conflicts” arise if the algorithm predicts a person to be doing two activities at once

• Deterministic rules for which activity takes priority, and how schedule is revised

• Conflict-resolution algorithms based on a small-scale, in-depth survey
TASHA’s scheduling mechanics (3)

• Work & school locations taken as input from survey data (as opposed to predicted by TASHA)

• Locations of all other out-of-home activities are estimated via simple location choice modules

• For instance:
  – Person will go shopping once...
  – Then...shopping will start at 19:00
  – Then...shopping will last for 1h45 mins
  – Then...shopping will be at Brent Cross shopping centre
London data

• *London Area Travel Survey* – Population & employment

• *London Transportation Studies* models
  – Spatial zoning system
  – Zone-to-zone travel time estimates

• Specialised dataset needed to develop TASHA

• Can implement pre-existing TASHA algorithms in London with a standard travel survey
Preliminary results

• Distributions of activity frequency, start time, & duration sourced from 1996 Toronto survey

• London data adjusted to match TASHA data structure (journey purpose, employment category, etc.)

• These results are diagnostic in nature
Activity start time distribution

• TASHA outputs, for all activity classes
Activity start time distribution (2)

- TASHA outputs, only education activities
Activity start time distribution (3)

- Observed survey data, only education activities
Activity start time distribution (4)

- TASHA outputs, all activity classes except education
## Observed v. predicted journeys/day

<table>
<thead>
<tr>
<th>TASHA predicted journeys per day</th>
<th>LATS observed journeys per day</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5+</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td>15%</td>
<td>1%</td>
</tr>
</tbody>
</table>
Comparison of datasets

- TASHA takes usual place of work [school] as an input

<table>
<thead>
<tr>
<th></th>
<th>Toronto survey data</th>
<th>LATS survey data</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of population employed</td>
<td>49%</td>
<td>45%</td>
</tr>
<tr>
<td>% of population employed <em>and</em> with an observed usual place of work</td>
<td>49%</td>
<td>33%</td>
</tr>
<tr>
<td>% of population in education</td>
<td>25%</td>
<td>22%</td>
</tr>
<tr>
<td>% of population in education <em>and</em> with an observed usual place of school</td>
<td>25%</td>
<td>15%</td>
</tr>
</tbody>
</table>
Summary

• Challenge of transferring a model system developed for a different spatial-network-social context

• Coupling between model specification and data availability

• Relaxing assumptions one-by-one time consuming, but yields valuable insights
Ongoing...

- Complete application of TASHA algorithms for London
- Incorporate into the research project’s cross-sectoral modelling system (called “SynCity”)
- Refine TASHA algorithms
  - What is a scheduling “conflict”?
  - How do people perceive them?
  - How do ICTs change how we experience demands on our time?
  - Do our data collection techniques adequately observe them?

Primary contact:  a.sivakumar@imperial.ac.uk
Dr. Aruna Sivakumar
TOWARDS THE IMPLEMENTATION OF AN ACTIVITY-BASED TRAVEL DEMAND MODEL FOR LONDON: DEVELOPMENT OF THE HOUSEHOLD ACTIVITY SCHEDULING MODULE

Scott Le Vine  
PhD student  
Imperial College London

Dr. Aruna Sivakumar  
Post-doctoral Research Associate  
Imperial College London

Dr. Matthew Roorda  
Assistant Professor, Department of Civil Engineering  
University of Toronto

Prof. John W. Polak  
Professor, Head of Centre for Transport Studies, Director of MSc Course  
Imperial College London

Abstract  
In recent years considerable interest has focused on the development of activity-based travel demand modelling techniques as an alternative to conventional trip-based approaches. Amongst the benefits claimed for these approaches are greater behavioural realism and credibility in prediction of travellers' behavioural response, especially to non-marginal policy measures. Work in North America and in some European countries has led to the development and implementation of significant activity-based demand modelling systems, which are being used for operational policy analysis. In the UK however, the practical application of activity-based modelling techniques has been much slower. In this paper, we report the first large scale empirical application of an activity-based demand modelling system for Greater London. The initial research focuses on the development of a household-based activity scheduling model. The modelling approach is based on TASHA [Travel Activity Scheduler for Household Agents], which was originally developed by Miller and Roorda (2003) at the University of Toronto. We evaluate portability of the model, and propose to extend the model along several fronts. This research is part of a broader line of enquiry into understanding the ways in which people's social and economic behaviour in cities result in aggregate energy demands, as an interim step towards developing strategies for improving urban energy efficiency.

1. Introduction and Objectives  
Activity-based travel demand models constitute the state-of-the-art in modelling travel demand. The activity-based paradigm is based on the fact that travel demand is derived from the needs of individuals, households and businesses (more generally, agents) to participate in activities distributed over space and time (see Jones et al., 1990, Axhausen and Gärling, 1992, Bhat et al., 2004). Correspondingly, changes in activity patterns, be they a result of policy or other social/technological trends, lead to changes in travel patterns.

Unlike the traditional trip-based approaches, activity-based models retain the spatial and temporal connections between all the activities and travel undertaken by an agent and are therefore more behaviourally realistic. The need for realistic representations of behaviour in travel demand modelling has become increasingly important as the emphasis has shifted from evaluating long-term investment-based capital improvement strategies to understanding travel behaviour responses to less capital-intensive demand management policies such as alternate work schedules, telecommuting, and congestion-pricing.

The conceptual appeal of this approach originates from the realisation that the need and desire to participate in activities is more basic than the travel that some of these participations may entail. By placing primary emphasis on activity participation and focusing on sequences or patterns of activity behaviour (using the whole day or longer periods of time as the unit of analysis), such an approach can address demand management issues through
an examination of how people modify their activity participations (for example, will individuals substitute more out-of-home activities for in-home activities in the evening if they arrived early from work due to a work-schedule change?).

The value of the activity-based approach is even more pronounced in today's environment, with impending energy security and air quality issues, rapid technological developments and socio-economic changes. Even as system integration is explored as a means of efficiency gains, it is necessary for transport planners and policy-makers to embrace a more holistic framework of understanding and modelling travel demand.

This research is part of a broader line of enquiry, in the Urban Energy Systems (UES) project at Imperial College London, into understanding the ways in which people’s social and economic behaviour in cities result in aggregate energy demands, as an interim step towards developing strategies for improving urban energy efficiency. As part of the UES project, we are developing SynCity, a prototype integrated model of an urban area as an energy system. At the heart of SynCity is an agent-based land use-transport modelling system (ABMS), which is based on the activity-modelling paradigm. In this paper we focus on the development of an activity-based model of daily travel demand that will be integrated into SynCity.

The specific objectives of this piece of research are

- to adapt an existing activity scheduling model for integration into SynCity, with particular focus on the UES project objective of modelling energy demand in an urban area
- to update the activity scheduling model along several lines, both conceptual and econometric, as required to meet the project objectives
- to develop a London implementation of an activity-based travel demand model
- to examine the spatial and temporal transferability of the model

Section 2 describes the research context beginning with a definition of key terms within the activity-based paradigm, followed by a brief discussion of the different modelling methodologies adopted in activity-based systems. This is followed by an overview of TASHA, an activity scheduling model system developed at the University of Toronto (Miller and Roorda, 2003), and a brief discussion of the reasons for the choice of TASHA for the UES project. We then present our research, with a short description of the data sources in section 3 and the preliminary results in section 4. Section 5 concludes this paper with a summary of further research being undertaken.

2. Research Context

The activity-based travel demand modelling paradigm, as described above, views travel as a derived demand; derived from the need to pursue activities distributed in space and over time. With a focus on passenger travel, the word ‘activity’ represents the many categories of time use, for instance, shopping, work, entertainment, social etc. Activities can be undertaken by a single individual (individual activities) or by multiple individuals together (joint activities, such as shopping with the family). Activities are characterised by a large set of attributes; several frequently observed attributes in empirical datasets include frequency of participation, individual or joint, time of day, location (in-home, IH, or out-of-home, OH), and mode of travel if location is OH.

The objective of an activity-based model is to predict the ‘string’ of activity and travel patterns (e.g. activity #1, followed by travel #1, followed by activity #2, etc.), and their attributes, for each individual and household in the study area, typically over a 24-hour period. Bhat et al. (2004) present diagrammatic illustrations of daily activity-travel patterns for workers and non-workers that represent all the processes that one wishes to model within an activity-based system.

In practice, activity-based modelling systems broadly consist of two sub-systems: activity generation and activity scheduling. The objective of the activity generation step is to predict the activities that an individual is likely to undertake on a given day in order to satisfy both his/her personal needs and commitments as well as the needs of his/her household. Whether the activity is undertaken jointly with other household members or singly is also determined in the activity generation stage. The objective of the activity scheduling step is to
capture the temporal and spatial linkages between the various activities undertaken on the
day. This mirrors the activity scheduling process that an individual must undertake in order
to plan his/her day. A model that can mimic this scheduling behaviour closely will therefore
produce the most realistic predictions of activity-travel behaviour.

Different methodological approaches have been adopted in the development of activity-
based modelling systems. The literature suggests that these can be broadly classified into
econometric methods and rule-based (also known as computational process) methods.
Econometric techniques use systems of mathematical equations to model individual activity
participation behaviour. These models are trained on observed activity-travel data to capture
the relationships between the patterns and observed characteristics. Such econometric
models are very flexible and powerful tools based on microeconomic behavioural decision
theory (see, for example, Ben-Akiva & Lerman 1985, and Hensher & Button 2000). A recent
example of an econometric model of activity-travel demand is CEMDAP (Comprehensive
Econometric Micro-simulator of Daily Activity-Travel Patterns, Bhat et al., 2004). Rule-based
methods, on the other hand, are based on a conditional, decision-tree type of approach and
are typically implemented through heuristics and simple statistical models. These are more
focused on the behavioural and decision-making processes and try to capture behaviour
through a system of rules. A contemporary example of a rule-based model of activity-travel
demand is TASHA (Miller and Roorda, 2003). Most operational models are implemented
using a micro-simulation framework where the choices of the individuals are simulated
dynamically based on the underlying (econometric or rule-based) models. Such a micro-
simulation model operates on the population of the study area to forecast activity and travel
patterns at the disaggregate level.

The objective of the UES project is to develop an agent-based [micro-simulation] modelling
system that can forecast both passenger and freight transport and energy demand based on
the underlying disaggregate activity participation patterns. One of the elements of the agent-
based modelling system is therefore an activity-based passenger travel demand model. We
have chosen to adapt TASHA, an activity-travel demand model developed by Roorda et al.
at the University of Toronto, for this purpose. The choice of TASHA for our implementation

![TASHA Conceptual Framework](Source: Roorda and Miller, 2006)
was based on several factors. One, TASHA is available as an open-source implementation in Visual C#, and the TASHA development team support methodological extensions to it. Two, TASHA is a rule-based model that focuses on scheduling behaviour and is developed using data collected specifically to understand this behaviour. It is therefore more amenable to being transferred to a different spatial context, potentially with little modification. Three, TASHA is a fully operational model that has been tested and implemented within an integrated land use-transport modelling framework.

We now present a brief description of TASHA in order to set the context for the rest of this paper. For more details, see Miller and Roorda (2003) and Roorda and Miller (2005). Figure 1 presents an overview of the TASHA model system. It consists of four elements: the activity generation step, the activity location choice models, the activity scheduling model, and the tour mode choice model. The activity types to be implemented in the model are relatively broad: work, school, independent market, joint market, independent other, joint other, return–home and in–home, which is the default activity.

TASHA takes as inputs various sets of statistical distributions, which in the Toronto TASHA application are sourced from the 1996 edition of the Toronto region’s periodic travel survey, the Transportation Tomorrow Survey (TTS) (Roorda et al. 2008). The three key distributions are:

- Activity frequency per 24-hour day (0, 1, 2, etc.)
- Activity start time, conditional on frequency (in 15-minute increments)
- Activity duration, conditional on start time and frequency (in 15-minute increments)

![Figure 2](image)

**Figure 2.** The three principal distributions used as inputs by the TASHA activity scheduling algorithms [Source: Roorda et al. 2008]

The activity generation step uses the frequency distributions of activity duration, start time and frequency to randomly generate potential activity episodes for each activity type for each person in the sample population (see illustration in Fig. 2).

The activity generation step is followed by a set of location choice models. TASHA takes the location of work and school activities as exogenous inputs, using TTS respondents’ ‘usual place of work’ and ‘usual place of school’ as identified by them in the TTS household interview. The models of activity location choice for the remaining activity types are entropy models that estimate the probability of choosing an activity location among the entire set of traffic analysis zones.

The next step is one of scheduling the activities for each individual in the study area. Person schedules are then constructed for each individual by iteratively scheduling his/her activities. This process results in scheduling conflicts when two or more activities are generated at overlapping times. For instance, in the activity generation stage a person may be assigned to be at a work activity and a shopping activity concurrently. TASHA has a set of deterministic rules for how people would deal with such conflicts. Three different classes of conflict are
defined by Roorda (2005), as shown in Figure 3, each of which are addressed by TASHA’s conflict resolution algorithms.

Class 1: A competing activity being added to the schedule is added within an original activity
Class 2: A competing activity being added to the schedule partially overlaps one or two original activities
Class 3: A competing activity completely overlaps one or more shorter original activities

Figure 3. Various classes of schedule conflict are shown, each of which are addressed by TASHA’s conflict-resolution algorithms (Source: Roorda, 2005)

3. Data Sources and Preparation

Broadly-speaking, TASHA is based on two classes of data as inputs. The first category of data is used in TASHA’s scheduling algorithms to predict how people make decisions in cases of a scheduling conflict (see Section 2). For instance, in the activity generation stage a person may be assigned to be at a work activity and a shopping activity concurrently. TASHA has a set of deterministic rules for how people would deal with such conflicts, which have been validated with observations in the CHASE (Computerized Household Activity Scheduling Elicitor) dataset. CHASE was a small-scale survey (n=264 households with 423 persons, undertaken in three waves in panel form from 2002 – 2004) in which respondents were asked to first indicate their planned activities during a forthcoming week, and then to update their activity-travel dairy as their week unfolds (Roorda and Miller 2005). CHASE respondents’ patterns of how they accommodated scheduling conflicts as they arose are consistent with the rules for scheduling activities within TASHA (Roorda 2005).

The CHASE survey procedures and instruments were a bespoke design for use in developing TASHA’s scheduling conflict resolution rules. Data of this type, however, is not widely-available for different spatial and societal contexts, in particular for contemporary London. A decision was taken to transfer the scheduling conflict resolution algorithms sourced from the CHASE survey (undertaken in Toronto in the early 2000s) into the TASHA London application. It is recognised that we therefore introduce an implicit assumption that the ways in which Londoners made decisions to resolve personal scheduling conflicts are broadly similar to the observed strategies of CHASE respondents.

The second class of data required for the London TASHA application is more standard in travel demand analyses and hence more widely-available. We therefore employ datasets specific to London, with the exception of the first item as noted below. This class of data includes:

- Frequency distributions of activity duration, start time and frequency: In TASHA, these distributions are generated from the observed data from the Toronto travel survey (TTS). The results we report herein employ this dataset; preparing these datasets using travel survey data from London is the focus of ongoing research.
• **Disaggregate population and employment data**: We use data from the 2001 edition of the *London Area Travel Survey* (LATS), a large-sample household travel survey (n=29,973 households with 67,252 persons) of respondents living in Greater London and environs. LATS included a household interview component and a 24-hour travel diary.

• **A spatial zoning system**: We employ the zonal system developed for the *London Transportation Studies* (LTS) series of travel demand models. This system includes 1285 zones, of which 1016 are within the LATS study area.

• **Zone-to-zone travel time estimates**: We incorporate zone-to-zone travel time estimates for the year 2001 from the LTS model.

TASHA takes as input the frequency distributions of activity duration, start time and frequency for 262 different combinations of activity purpose (work, shopping, other, etc.) and personal characteristics such as age, gender, employment/student status, and employment classification. In principle, these distributions can be generated from observed patterns in the LATS survey data, however at the time of writing we employ these distributions from the Toronto TASHA application, which is based on the TTS dataset. As such, in the preliminary results which we report here there is an implicit assumption of congruence between these distributions as collected from TTS respondents and those of Londoners in 2001.

Of the various datasets listed here, aligning the LATS disaggregate data with TASHA’s data input requirements was the most complex data integration task. In particular it involved mapping between categories (e.g. activity purpose, employment classification, etc.) which were qualitatively different in LATS than Toronto’s TTS survey.

For instance, respondents in the TTS survey who indicated that they are employed or in school were asked the address of their ‘usual place’ of work or school, respectively. However, LATS respondents were not asked similar questions. We therefore developed a set of procedures to impute comparable information from a combination of the LATS travel diary and household interview data, though it would appear that this qualitative dissimilarity in the underlying travel survey datasets resulted in fairly significant inaccuracies in the results of our early London TASHA model runs as discussed in the following section.

4. Preliminary Results

The results which we report in this section are based on an initial run of the TASHA London application. This initial run followed directly from conversion of the TASHA source code to accommodate the spatial zoning system we employ for London and preparation of input datasets from London consistent with TASHA’s required data formats as described in Section 3. The results presented herein should therefore be interpreted as diagnostic in nature only; at the time of writing there are significant discrepancies between the TASHA London outputs and patterns observed in the LATS survey data. The results we present here should not be used for behavioural or policy analysis.

We first analysed the temporal pattern of the start times of people’s out-of-home (OH) activities. Figure 4 shows the temporal distribution of start times for people’s various activities over the 24-hour representative weekday (TASHA’s analysis period). In reviewing this visualisation of the TASHA London output data, we noted several regularities:

1) Observable peaks in the distribution at the top of each hour (particularly during the period 10:00 AM to 14:00 PM)

2) A large mass (of approximately 9%) of people’s out-of-home activities were observed to start in the 15-minute time slot between 08:45 and 08:59 AM.
Peaks in the activity start time distribution at each hour are to be expected with scheduling models estimated from self-reported datasets; researchers have previously noted (Battelle, 1997) that people completing activity-travel diaries tend to disproportionately report times in round number increments (e.g. 5-minute, 15-minute, or 1-hour increments). With the data sources available on this project (LATS and TTS), however, the extent to which this is an artefact of data collection methodologies is unknowable. The other broad possibility for explaining such patterns is that people’s activities perhaps may disproportionately be scheduled to start in round time increments (e.g. a 9:00 – 5:00 workday). The relative salience of these two explanations, however, is not identifiable with the available data sources.

With regard to the second observation, the 9% mass in activity start times in the 08:45 – 08:59 AM time band, upon further review this peak was found to be associated with education activities (i.e. attendance at school or university by a student). Figure 5 shows the temporal distribution of start times for only education activities. Approximately 47% of education activities were observed to start in the 15-minute time slot between 08:45 and 08:59 AM. When analysing only education activities, a second, smaller mass in the distribution (3% of education activities) emerged, corresponding to the time slot from 12:15 – 12:29 PM.
Investigation of these two masses in the temporal distribution of education activities uncovered a significant data integration issue. The TTS survey datasets on which TASHA’s distribution patterns are based includes travel diary data only from respondents aged 11 years and over. In order to simulate activity patterns for younger schoolchildren, the TASHA algorithm deterministically allocates schoolchildren aged 6 – 10 an education activity starting at 08:45 with a duration of six hours and 45 minutes. Kindergarteners (defined in TASHA as schoolchildren aged 5 or under) are randomly allocated a half-day education activity starting at 08:45 AM or 12:15 PM lasting for three hours and 15 minutes. The logic for these rules is derived from the school-day schedules of a sample of Toronto-area public and private schools.

For the TASHA London application, however, the available data (LATS) includes travel diary information for all respondents aged 5+. Therefore, one is in principle able to develop education activity start time distributions for schoolchildren aged 5 – 10 for the London TASHA application, whereas this was not feasible in Toronto using the TTS dataset. Evidence from LATS shows the distribution of education activity start times to be less strongly-peaked in any particular 15-minute time band than the distribution predicted by TASHA, and further there is no indication of a secondary mid-day peak (see Figure 6). On the basis of this evidence, and anecdotal knowledge of the British education system, we believe there may well be structural differences between the context of London and TASHA’s algorithm (based as it is on data collected from a North American region) in how the school day is temporally organised. Hence revising this portion of the TASHA algorithm and distribution input information with data from LATS is a planned refinement as this research proceeds.

Figure 6. Temporal distribution of start times for education activities (observed in LATS survey data)

Figure 7 shows the same data structure as Figure 5, but for all activity purposes except education. With the strongly-peaked education purpose removed, the patterns of the remaining activity purposes can be seen more clearly. A large proportion of work activities begin in the early morning hours, which is intuitively logical. Away-from-usual-workplace business activities are distributed roughly evenly across the working day. Shopping and activities classified as ‘other’ (a rather broad category in TTS and therefore the TASHA algorithms) are distributed more widely, including a significant proportion in the early evening hours. Activities classified as ‘other’ in which two or more adults from the same household participate (i.e. ‘joint other’ activities) are peaked in the early evening, an intuitive result.

Interestingly, the peaks in activities starting at the top of each hour are quite large in the hours from 09:00 AM to 14:00 PM, but weak at other times of the day. The interpretation of this pattern is unclear; we anticipate verifying whether it is replicated when London-specific
start-time distribution data has been prepared from LATS. This result may be an artefact of the TTS survey design.

Figure 7. Temporal distribution of start times for all activities except education purpose (predicted by TASHA)

In addition to this preliminary review of temporal patterns of activity participation, we evaluated the rate of journeys per day implied by the activity participation pattern predicted by TASHA. Several noteworthy patterns emerge when comparing TASHA’s predictions with the observed data in the LATS survey.

Table 1 shows the cross-tabulation of Londoners by the number of journeys per day observed in LATS and predicted by TASHA. ‘Correct’ predictions on the diagonal of the matrix are highlighted. In both the observed and predicted datasets, a plurality of people make two journeys in the 24-hour day, and these proportions are of similar magnitude (42% observed versus 47% predicted).

Table 1. Cross-tabulation of Londoners by number of journeys per day, observed in LATS survey versus predicted by TASHA

<table>
<thead>
<tr>
<th>LATS observed journeys per day</th>
<th>TASHA predicted journeys per day</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>----</td>
<td>---</td>
</tr>
<tr>
<td>0</td>
<td>9%</td>
</tr>
<tr>
<td>1</td>
<td>0%</td>
</tr>
<tr>
<td>2</td>
<td>4%</td>
</tr>
<tr>
<td>3</td>
<td>1%</td>
</tr>
<tr>
<td>4</td>
<td>1%</td>
</tr>
<tr>
<td>5+</td>
<td>1%</td>
</tr>
<tr>
<td>Sum</td>
<td>15%</td>
</tr>
</tbody>
</table>

A small proportion of people (1%) are observed in LATS making a single journey; that is, they performed precisely two activities at different places, one of which may have been – but is not necessarily – their home. By the nature of the LATS dataset, the start time of such a person’s first activity, and the end time of their second activity, is not known, thus neither activity is fully-specified. The LATS diary dataset is censored by the beginning and end of the 24-hour diary period at 04:00 AM and 03:59 AM the following day, respectively. The TASHA algorithm, however, simulates entire uncensored OH activities, from which the demand for mobility is derived. By this nature, no person would be predicted by the TASHA algorithm to make precisely one journey, which is in fact what we observe.

The ‘tail’ of the distribution of journeys per day is substantially larger in the observed dataset than the TASHA prediction (17% versus 5% making five or more journeys, respectively). This result is plausible, as Londoners exhibit a larger proportion of walking journeys than
Torontonians (29% in LATS versus 6% in TTS), and one may hypothesise that this might, at least in part, explain the substantial proportion of people in London who reported making five or more journeys in a single day. In support of this hypothesis, only 9% of TTS respondents reported making five or more journeys in their diary day, versus the 17% so observed in LATS.

This result merits further study, as in addition to perhaps indicating systematic differences in Londoners’ and Torontonians’ mobility patterns associated with the different spatial, network, and socio-economic contexts, it may also [in part] plausibly be an artefact of different survey designs in TTS and LATS, or in some way due to the structure of the TASHA algorithms. Respondents to the TTS survey, for instance, were not asked to record walking trips unless they were to/from work or school. As with the temporal distribution of activities, we anticipate observing whether these patterns are replicated when London-specific activity frequency distribution data has been prepared from LATS.

Another finding from Table 1 is notable – the relatively large proportion of people predicted in TASHA not to travel (15% observed in LATS versus 33% predicted by TASHA). In part, we believe this is due to a structural difference in the design of the TTS and LATS surveys as noted in Section 3. TTS respondents who were employed and/or in education were asked the address of their usual place of work and/or school, which TASHA uses as input data points. LATS respondents were not asked a similar question in the household interview, however; LATS respondents were only asked whether they are employed or in education, not the address of their usual place of work/school. The procedure we developed to impute such a person’s place of work/school also incorporated the LATS travel diary data. Hence we have no knowledge of a LATS respondent’s ‘usual’ place of work or school if they did not travel to it on travel diary day.

Table 2 shows this structural difference in the TTS and LATS datasets. Whereas 49% of TTS respondents are recorded as workers (and thus recorded with a usual place of work), only 33% of Londoners can be so defined using the imputation methods we devised. The proportion of London’s population observed in LATS to be employed in education is marginally lower than Toronto’s as captured in TTS. This may be related to a further minor difference in survey design; TTS respondents may be classified as employed and/or in education, whilst LATS respondents are classified as either employed or in education. We may expect that this results in a marginally lower proportion of Londoners being classified as employed or in education, because any respondents engaged in both employment and education would only be classified in the dataset as one or the other. However, from TTS and LATS alone, the degree to which these differences are artefacts of different survey designs, or rather systematic differences in the populations of London and Toronto, is not identifiable. As we do not have access to the TTS disaggregate data (for reasons of respondent privacy), we cannot observe the proportion of Torontonians who are classified as both employed and in education, which would be useful in identifying the magnitude of this effect.

| Table 2. Comparison of employment and education data between TTS and LATS |
|--------------------------------------------------|------------------|------------------|
| Proportion of population employed | TTS survey data | LATS survey data |
| Proportion of population employed and with an observed usual place of work | 49% | 45% |
| Proportion of population in education | 25% | 22% |
| Proportion of population in education and with an observed usual place of school | 15% |

The implications of this on the mobility patterns predicted by TASHA are not fully known at the time of writing, but we believe they may be large and may at least partly explain the discrepancy between the proportion of Londoners observed in LATS to not travel on the random survey day & the proportion predicted by TASHA. For instance, in the initial run of TASHA 24% of children between the ages of 5 and 10 (inclusive) were predicted to not travel, whereas only 8% of children in this age range were observed in LATS as not
travelling. In TASHA’s activity frequency distribution, roughly 5% of children aged 11-15 (the youngest age bracket for which diary data were collected) were observed to not have attended school on their diary day, which seems an intuitively plausible proportion given that the representative day which TASHA predicts is during school term time.

In order to deal with this qualitative difference in survey design, we plan to extend TASHA’s destination choice models for non-work/non-school activity purposes to also predict activity location for workers and students for whom we cannot identify their usual place of work/school from the LATS diary data.

5. Summary and Further Work

In this paper, we present preliminary results in the development of an activity-based travel demand modelling system for London. TASHA, an activity-travel model developed by Miller and Roorda (2003) at the University of Toronto, is the tool selected for this implementation. The ultimate objective of this research is to develop a comprehensive agent-based model of an urban area, including land use and transport processes, with the specific aim of understanding and optimising energy demand patterns across the different domains. A secondary objective of this research is to analyse spatial transferability of the models, which is of particular importance to the UES project as we seek to develop modelling systems that can be deployed in different urban areas with a high degree of reliability.

In this preliminary stage of our research, we implement TASHA for London under the strong assumption that all sub-models are transferable from the context of Toronto (at various points between 1996 and 2004) to London in 2001. In particular, we use the frequency distributions for the activities from the Toronto travel survey data, the location choice models estimated for Toronto, and the scheduling algorithm designed using the Toronto CHASE data. We believe that the first of these assumptions (that of congruence between the activity distributions in the TTS and those of Londoners in 2001) is likely to be less robust than the assumption of similarity between the strategies for addressing scheduling conflicts in TTS respondents and Londoners. This appears likely as the former is outside of the realm of purely cognitive processes (prioritising amongst different categories of activities), and instead associated with observed traces of activity-travel behaviour. We believe it is more plausible for such observations to be dependent on spatial idiosyncrasies and social context, and hence are less confident in the validity of this assumption.

The lessons to be learnt from the preliminary results are two-fold. First, we note the relatively tight coupling between survey design, the characteristics of the resulting empirical datasets, and the procedures employed in the TASHA activity-scheduling algorithms. As the research proceeds, caution is called for in preparing datasets specific to London to which the TASHA procedures will be applied. One must systematically review outputs for consistency with the local empirical data, and be prepared to iteratively revise the work plan as discrepancies arise which may signal that modifications to the work plan are required. For instance, the lack of a data point in the LATS survey dataset regarding people’s ‘usual’ place of work/school has led to our decision to impute such people’s work/school locations using econometric estimation procedures.

The second lesson to be drawn from our experiences to date is distinct from the data integration issue noted above, though related. As TASHA is based on datasets collected in a North American region with particular social and economic structures and a spatial-transport system with certain characteristics, it is perhaps to be expected that there may be differences between the ways Londoners organise their daily lives and the prescriptions in the TASHA algorithms, and that these differences may lead to structural differences between observed and simulated patterns of activity participation and travel. The apparent differences in the temporal organisation of a schoolchild’s day between Toronto & London are one example; the challenge for researchers is to identify which discrepancies are indicative of substantive differences between London and Toronto, and which are artefacts of subtly different data collection methodologies which underpin the two datasets. The distinctions between the two possibilities are not necessarily identifiable from comparing the two data sources; in these cases the application of judgement and alternative techniques must be considered.
We are currently in the process of relaxing the assumption regarding the frequency distributions, and are hence constructing appropriate distributions using the LATS data. This is anticipated to overcome several issues identified in section 4, such as the longer tail in the distribution of journeys per day in London. At the same time we are re-estimating the activity location choice models using London data, and freshly estimating work and school location choice models. The next model run is therefore expected to present a clearer picture of the true transferability of the TASHA scheduling algorithm. Following that, we propose to update TASHA along several lines, such as the incorporation of models specifying the nature of in-home time use, the incorporation of models that capture the use of technology in activity participation and ultimately models that predict energy demand related to activity participation and travel.

6. References

Acknowledgements
We would like to thank the many people who have supported, and continue to support, us in this research. Eric Miller and the entire TASHA team at the University of Toronto; our colleagues at Transport for London; and BP, the sponsors of the UES project.