A GIS-based probabilistic time-activity model for exposure assessment

Beale L, Briggs D, Gulliver J

Abstract

A probabilistic time-activity model has been constructed in a GIS to enhance existing time-activity data or simulate complete time-activity patterns at the individual level, as a basis for exposure modelling. Time-activity surveys typically contain information on the type of location (e.g. home or work) and occupancy times but very rarely do they include detailed spatial data on journey routes. Combining probabilistic techniques with route modelling in GIS provides the opportunity to model individuals' paths through the environment. Journey and occupancy times are modelled on the basis of population distributions from existing time-activity surveys or national statistics. The resulting simulations can be used to assess potential exposures of specific subgroups of people to environmental hazards (e.g. air pollution, noise, traffic accidents) and to compare risks or other indicators under different management or policy scenarios.

Background

Although data on travel behaviour are sparse, the available statistics are remarkably consistent over much of the western world. In the UK, for example, the UK2000 time use survey (National Statistics, 2003) reported an average of about 1.25 h per day travelling (although this rises slightly at weekends). A detailed local time– activity survey in Northampton, UK showed average daily travelling times of 1.2 h for adults, 1.3 h for college students, and 1.0 h for schoolchildren (Briggs et al., 2003) In Germany, average daily travelling times are estimated at 1.7 h for adults and 1.4 h for children (Seifert et al., 2000). In the USA, the National Human Activity Pattern Survey (NHAPS) reported an average of 1.3 h per day spent ''in a vehicle'' (Klepeis et al., 2001).

Most people thus spend a relatively small proportion of their time travelling. Travel environments, however, are often relatively polluted, with the result that journey-time exposures make up a disproportionately large amount of total personal exposures to ambient (and especially traffic-related) air pollution. Journeys are also associated with a range of other potential hazards, including traffic accidents and exposures to infectious diseases. Consequently, they have potentially important implications for health and well-being. By the same token, there are many potential uses for trip modelling for public health and epidemiology, such as:

- the spread of infectious agents throughout a dynamic population,
- the estimation of human exposures to urban air pollution under different emission, meteorological or behavioural scenarios,

- modelling effects of transport policies on behaviours, transport flows and exposures.

Whilst there are increasing numbers of potential sources for time-activity data, the available data tend to be highly aggregated, and to lack spatial and temporal specificity. Trip data, for example, often take the form of population counts showing overall numbers of population movements rather than individual trips or trip chains (e.g. UK census O-D data). Traditionally, data on pedestrian and vehicle movement in the street network have been based on physical counts or time lapse photography. Physical counts involve enumerating people or vehicles that pass a set of points or gates on a network over a set period of time, to give a measure of flow through the network. Time lapse photography extends this approach by recording the movement of all pedestrians or vehicles at a junction in a defined area and drawing trail diagrams to plot their movement. These methods, however, are necessarily limited in terms of their sample size or spatial coverage, and cannot be used to predict future patterns since they do not attempt to explain the processes that cause the movements. Such processes can be considered to be an interaction between a large number of determinants and contingent factors, including the street network, the location and characteristics of particular attractions on that network (e.g. shops, workplaces), and population characteristics such as density, age, gender and socio-economic status. One way of modeling trips is thus to use GIS techniques to link relevant data sets (e.g. on population and local 'attractors'), and use route finding algorithms together with data on travel patterns or preferences to simulate trip behaviours.

Time-activity modeling in GIS

The GIS-based time-activity model outlined here attempts to use this approach to simulate daily activity patterns. The model has been developed as part of the EU-funded HEARTS project, and simulates time-activity patterns as a series of trips, with intervening periods of residence at defined locations. Replication of the modelling over a large number of individuals enables population-level distributions of trip behaviour to be built up. The resulting time activity data can then be intersected with information on environmental hazards, such as air pollution or accident risks, to derive estimates of risks to health and well-being.

The model is probabilistic in concept, and can be run using either individual-level or aggregate (population) level data from existing time-activity surveys. It is also designed to cope with varying levels of data completeness. Data requirements are shown in Table 1. A complete time-activity data set to run the model would include individual data on the start and end location of each trip (X/Y), the start and end times of each trip, trip mode, travel speed, route and activity type (i.e. activities undertaken at the destination). When this information is complete, the model simply reproduces the time activity sequence. If any of these parameters are missing from the time-activity database, however, they can be imputed probabilistically on the basis of available statistical information. For example, if the destination is not defined, it can be estimated by sampling from all the available destinations that match the specified activity type and travel time. Equally, if the route is not specified, it can be simulated as the shortest distance or travel-time between the start and end locations using PATH function in GIS. If no individual level data are available, the model imputes the entire time activity sequence for a virtual individual by sampling the statistical data. These statistical data comprise generalised distributions for the study population (Table 1). Time-activity modelling can thus be performed using a mixture of individual level and statistical (aggregated) data from time-activity surveys.

Table 1. Data requirements for time activity modelling

Information requirement	Individual level	Statistical/aggregate
Location	x,y co-ordinates of origin and destination	Density distribution (e.g. of population, workplaces) or land cover map
Activity type	Classification of each location (e.g. home, school, work, shops)	Transition matrix by time of day (population transfers between different activity types)
Start time	Departure from origin	Transition matrix (plus target times for specific activities – e.g. start/end of school day)
End time	Arrival at destination	
Travel mode	Mode of travel (e.g. car, walk, bicycle, bus)	Modal distribution by travel time
Route	Digitised route	Network data (e.g. roads, bus-routes, pedestrian paths)
Speed	Actual travel speed (or time and distance)	Average speeds by mode (and network link, if available)

Model description

The model has been programmed in ArcView GIS. It is menu driven, with the selected input data defining the correct modules that must be called to model any unknown data, whilst utilising all known data. The modules are outlined in Table 2. To facilitate integration of the various modules, a fixed format of input data is required. This specifies all the required fields for trip-modelling, with data that must be generated by the model entered as empty fields. On start-up, the input data files are automatically checked, any required matching data verified (e.g. network speeds and modes) and the relevant statistical data then selected.

The input data also includes a unique person-ID, which identifies each individual being modelled. This enables a distinction to be made between trip-chains - i.e.

a series of linked journeys that an individual has taken throughout the day – and a set of independent journeys undertaken by a number of different individuals.

Table 2. Modules used in trip modeling.

MODULE	DESCRIPTION	INPUT DATA
HOME	Computes home location by sampling population density (within a zone or across a study area)	Population density
TRANSIT	Selects destination type for current hour by sampling a transition matrix	Hourly transition matrix
MODE	Selects mode by finding travel mode along the specified route that best matches travel time	Origin Destination Route (by mode) Travel time
ROUTE1	Selects route and destination by finding route (by predefined mode) to available destination types that best matches the specified travel time	Origin Destination type Mode Travel time
ROUTE2	Computes mode, route and travel time between predefined origin and destination by sampling all available mode-route options, using a least-cost algorithm	Origin Destination
ROUTE3	Computes mode and route between predefined origin and destination by selecting the mode-route combination that best matches specified travel time	Origin Destination Travel time
ROUTE4	Selects route (for the predefined mode) from origin to destination on the basis of a least-cost algorithm	Origin Destination Mode
DESTIN	Defines a set of search radii for each mode, by sampling statistical data on trip length by destination type and mode, and then maps the distribution of available destinations of that type within these radii	Mode Destination type
TRIP	Computes mode, route and destination on the basis of a least-cost algorithm by sampling all available combinations (from the predefined origin to all destinations of a predefined type within a specified set of search radii)	Origin Destination type Map of available destinations
TIME	Computes start and end time by randomly selecting times within the available travel hour, based on the specified travel time	Origin Destination Route Mode Travel time

Model operation

In order to calculate any route the minimum requirement is clearly an origin and destination. If these are not specified in the input data, they must be derived at the outset of modelling, commencing with the origin. In modelling the initial origin for any time-activity sequence, the location type is always assumed to be an individual's home. Modelling also starts, by default, at midnight. Where the place of residence is not specified in the input data, a population-weighted point or shapefile coverage is randomly sampled to define the home. For journeys that constitute a trip chain, only the first origin is sampled in this way, with all subsequent origins being automatically defined as the destination from the preceding trip.

In cases where the journey destination information is missing, a transition matrix is sampled to select a destination (activity) type. Transition matrices can be derived from national data (e.g. census or household surveys), purpose-designed surveys, or indirectly from information on time-activity patterns (e.g. hours spent at school or work). Transition matrices are constructed to show the cumulative percentage of people moving from any one source activity (origin) to any other target activity (destination), by hour of the day (Table 3): twenty-four matrices must, therefore, be provided for any single day. Separate matrices may also be specified for specific population subgroups (e.g. children, commuters). Using the relevant field for the current origin, sampling is carried out by randomly generating a number (between 0 and 100). This value is then used to select the relevant cell, and thus destination type. For example, for a trip starting at home, a random number of 37 would give a destination type 'work' (between 20% and 50%) in Table 3.

Table 3. Cumulative percentages of time-activity transitions for population (all ages) at time 08.00-09.00

Time period 08.00-09.00		Origin (at 08.00)				
Destination (by 09.00)		Home	Work	School	Shops	Leisure
	Home	20	10	0	10	20
	Work	50	50	0	40	40
	School	90	85	100	70	60
	Shops	95	95	100	95	70
	Leisure	100	100	100	100	100

N.B: table is read downwards, according to origin. All the columns in the matrices must sum to 100.

The actual destination location must then be derived. If this is not specified in the input data, it is imputed probabilistically, using data on the density distribution (or actual location) of relevant activity centres. Thus, for the journey from home-work, the location of 'work' must be specified. A module 'destin' maps the distribution of available destinations for the specified type, by:

a) setting a search radius for each available mode from the origin, by sampling distributions of trip distances by mode and trip type (example below)
b) mapping the distribution of all available locations within this radius. Locations of destinations, by type, are specified *a priori* in a point coverage or shapefile.

Where not specified in the input data, travel mode, route and journey length (distance and time) are then modelled. These are clearly interdependent, since not all routes are available for any specific mode and the journey length depends on the route and mode used. Modal preferences, in turn, vary depending on the length (i.e. duration) of the journey.

Where the travel time is known, the route and mode are selected by modelling the travel time to the destination by all available routes and modes, and selecting the one which most closely matches the specified travel time. In the event of more than one eligible option, the route is selected randomly from the tied set.

Where the travel time is not specified in advance, the mode of travel and route is selected probabilistically by randomly selecting from all the available route-mode combinations on the basis of travel time. Statistical data on modal preferences by distance (e.g. based on national travel surveys) are used for this purpose. Table 4 presents an example: preferences (for n number of modes for each distance class) are presented as cumulative percentages. To enable different destination-route-mode combinations to be compared, combinations are weighted as follows:

$$W_{ijk} = [(1-(T_{ijk} / \Sigma T_{ijk})) * (P_{ij} / \Sigma P_{ij})] \; / \; \Sigma \, [(T_{ijk} / \Sigma T_{ijk}) * (P_{ij} / \Sigma P_{ij})]$$

Where : W_{ijk} is the weight assigned to route *i* by mode *j* to destination *k*
T_{ijk} is the travel time taken to complete route *i* by mode *j* to destination *k*
P_{ij} is the preference for mode *j*, over the computed travel time for route *i*.
Values of *Pij* are assigned on the basis of trip time e.g. using Table 3

[Eq. 1]

The resulting weighted preferences are then rescaled to 100, and a mode-route combination chosen by drawing a random number between 1 and 100.

Table 4. Cumulative percentages of journeys by mode and time

Time	Car	Bus	Cycle	Walking
0-0.5	5	15	25	100
0.5-1.0	10	30	40	100
1.0-2.0	25	45	65	100
2.0-3.0	45	65	80	100
4.0-5.0	65	80	90	100
6.0-10.0	80	95	100	100
10.0-20.0	90	100	100	100
>20.0	95	100	100	100

N.B: table is read cross-ways, according to journey length. The values of W_{ij} should sum to 1.0.

If the route forms part of a trip chain, the preferences are rescaled after each trip (by halving the percentages for all other modes and attributing the difference to the previously used mode) in order to increase the likelihood of travelling by the same mode.

The exact start time of any activity or trip within that hour is selected randomly. The one exception is when the previous journey exceeds an hour; in that event the modelled destination time is set as the earliest possible start time for the next trip.

The speed along each route varies both spatially (i.e. from link to link) and over time (e.g. depending on traffic congestion). The relevant speed for each link in the route shapefile is found from the input network coverage. The relevant time period is defined by referencing the journey start time.

In order to derive time-activity patterns for a specified (sub-)population, simulations are carried out for a large number (e.g. several hundred) of individuals and the results pooled to give a distribution. Exposures during these activities can also be assessed, by intersecting the time-activity profiles for each individual with spatio-temporal data on relevant hazards (for example, hourly air pollution maps or traffic accident risks). Aggregation of the resulting exposures can thus be used to derive distributions for the whole population or sub-group, and these can be linked to dose-response functions to estimate potential impacts on health.

Conclusion

The methodology described here enables time-activity patterns to be modelled on the basis of variable levels of data. The model is probabilistic, and designed to extract the maximum amount of information possible from the available data by using the advanced functionality of GIS. Applications are diverse. They include studies of population dynamics, accessibility, environmental exposures and health risks, transport and land use planning. Nevertheless, except where modelling is carried out on the basis of reliable individual-level population data, the results should not be regarded as models of actual behaviour or exposures. Instead, they represent scenarios, representing specific population groups and activity patterns. One of the major applications of the model is thus to analyse potential impacts of different policies or interventions – such as changes in traffic management or air pollution control strategies - by comparing an indicative group of virtual individuals under different conditions. At an individual level there can be no substitute for data collection; at a population level, however, this method provides a powerful tool both for research and policy.

References

Briggs D.J., Denman A.R., Gulliver J., Marley R.F., Kennedy C.A., Philips P.S., Field K. and Crockett R.M., Time–activity modelling of domestic exposures to radon, *J. Environmental Management* **67** (2003), pp. 107–120

Klepeis N.E., Nelson W.C., Ott W.R.,. Robinson J, Tsang A.M., Switzer P., Bhar J.V., Hern S.C. and Engelmann W.H., The National Human Activity Pattern Survey (NHAPS) a resource for assessing exposure to environmental pollutants, *J. Exposure Anal. Environmental Epidemiology* **11** (2001) (3), pp. 231–252.

National Statistics, UK 2000 Time Use Survey, Office for National Statistics, London (2003).

Seifert B., Becker K., Hoffmann K., Krause C. and Schulz C., The German Environmental Survey 1990/92 (GerES II) a representative population study, *J. Exp. Anal. Environmental Epidemiology* **10** (2000) (2), pp. 103–114.

Author Information

Dr Linda Beale
Research Associate (GIS)
Department of Epidemiology and Public Health
Imperial College London
Upper Ground Floor, Praed Street Wing,
St Mary's Campus, Norfolk Place,
Paddington, London, W2 1PG
Email: l.beale@imperial.ac.uk

Prof David Briggs
Chair in Environmental and Health Sciences
Department of Epidemiology and Public Health
Imperial College London
Upper Ground Floor, Praed Street Wing,
St Mary's Campus, Norfolk Place,
Paddington, London, W2 1PG

Dr John Gulliver
Research Associate
Department of Epidemiology and Public Health
Imperial College London
Upper Ground Floor, Praed Street Wing,
St Mary's Campus, Norfolk Place,
Paddington, London, W2 1PG