# Trace Data for Activity-based Travel Demand Forecasting Models

Kostas Goulias
Department of Geography
University of California, Santa Barbara

A position statement for the FHWA-sponsored Peer Exchange and CSISS Specialist Meeting

**GPS and Time-Geography Applications for Activity Modeling and Microsimulation**

Santa Barbara, CA
October 10-11, 2005

In travel demand analysis and forecasting we finally see progress in advancing a variety of ideas that were originally proposed in the 1960s and 1970s. One field of research and practice that dominates proposals for new techniques is the activity-based approach with three basic ingredients:

- Explicit treatment of every day life as a time allocation exercise. One expression of this is to treat travel as derived demand, i.e., participation in activities such as work, shop, and leisure motivate travel but travel could also be an activity as well (e.g., taking a drive). Time allocation is viewed as management of episodes (events defined by their starting time, duration, and ending time) and they are arranged in a sequence forming a pattern of behavior that can be distinguished from other patterns (a sequence of activities arranged in a chain/tour of episodes). These episodes and their chains are not independent and their interdependency is accounted for in the theoretical framework(s). For example, one can identify episodes that belong to a "project" such as prepare an evening meal for friends. This project includes a trip to a wholesale store and stay at the store, a trip to a gourmet grocery store and stay at the store, a variety of meal preparation tasks at home, arrival of the friends at home and so forth. More complex projects may focus on children education, career advancement, caring for older parents, and so forth.

- The household is considered to be the fundamental social unit and treated as a decision making unit. Interactions among household members are explicitly modeled to capture task allocation and roles within the household. The relationships and change in these relationships as households move along their life cycles are also modeled. In this way, individual's motivations, commitments, constraints, and distributed intelligence are explicitly modeled and they are simulated. This allows to model the behavioral process underlying scheduling of activities to identify potential for change under varying policy scenarios;

- Explicit consideration of constraints by the spatial, temporal, and social dimensions of the environment is also given a primary focus. Consistent with and inspired by time-geography these constraints can be explicit models of time-space prisms or reflections of these constraints in the form of model parameters and/or rules in a production system format.

Operational activity-based model systems follow four theoretical and computational traditions that are combinations of:

- Microeconomic models based on Becker's time allocation and McFadden's conditional Logit formulations that develop maximum utility models to simulate policies using appropriate variables within each model;
- Production system/computational process models following Newel's and Simon's rule-based approaches simulating behavior using statements of the type - if X= A then Y=B type of rules;
- Cellular automata models following von Neumann's ideas that also contain rule-based models; and
- Statistical pattern recognition and transition probability approaches to create "data-driven" models.

The input to these models are data of social, economic, and demographic information of potential travelers and land use information to create schedules followed by people in their everyday life. The output are detailed lists of activities pursued, times spent in each activity, and travel information from activity to activity (including departure time, travel time, mode used, and so forth). This output is very much like a "day-timer" for each person in a given region and contains as a subset output usually produced by four-step applications. Unlike the four-step model that produces summaries of coarse behavioral description at the level of small geographic regions (called traffic analysis zones), activity-based approaches produce simulated activities for individuals called *synthetic schedules*. Moreover, given the statistical/econometric maturity of the field we also begin to see outputs that in addition to the averages produced are also accompanied by measures of variability.

To produce these outputs all activity-based approaches use some kind of computational evolutionary engine that is called *microsimulation*. Each model system operates in a somewhat different way but all aim at recreating a virtual microcosm in which individuals and their households are the focus. Their environment is represented by parcels of land where they live and they visit and stay. Their movement is represented by vehicles on highways. Time progression depends on the application but one aim is to develop second by second microsimulations to match other model systems developed for traffic operations.

Many recent model reviews find that applications use hybrid paradigms that are combinations of statistical/econometric models and computational process models to represent behavior. Other use statistical models embedded into microsimulation frameworks to evolve either individuals and/or households over time. Modeling and (micro)simulation appear to be concentrating at two poles. They are either designed for long horizons such as 25 years in the future with yearly cycles aiming at long range forecasting or for shorter periods aiming at reproducing within a day activity and travel patterns. Attempts to reconcile and coordinate different time scales are also starting to appear in the literature and emerging theoretical frameworks.

In this context, traces (i.e., time sequenced paths in space) of individuals and/or vehicles using geospatial technologies can serve a variety of purposes. These traces can be produced by any type of technology and do not necessarily need to be produced by GPS devices alone. The ultimate objective of activity-based approaches is second by second and parcel by parcel simulation of individuals and their social units (households, firms, professional groups, and so forth) on existing highways and pathways. Assuming provision of data will be at that level of spatial and temporal resolutions, there are at least six beneficial roles for GPS data in this context:

1. Use as one source of data to build household time-space prisms.

Travel demand forecasting models use the Hagerstrand-type time-space prism idea to model constraints for individuals. Ultimately we would like to model interactions with other persons via coupling constraints but also the dynamics of everyday communication and task allocation within social groups. These scheduling dynamics, however, may change dramatically shaping each individual's time space rapidly and making the prism itself a much more elastic entity than originally envisioned by Hagerstrand (and all subsequent formulations - see Janelle's position paper) much less restrictive in its influence on travel behavior. Richer longitudinal (many subsequent days) GPS data of all the persons in a household and activity diaries matched to the data may provide the first source of information about household patterns and their change from one day to the next. This information can be used to identify household-based time-space prisms and the rapidly changing action spaces implied by these prisms.

2. Use as a cognitive verification machine about perception of time and distance

   Earlier contributions to travel behavior research in 1970s demonstrate that traveler behavior is a function of knowledge about spatial opportunities and time. The striking majority of models use space and time as measured and observed by analysts instead of the persons traveling. GPS and related data offer a unique opportunity to compare these two sources of information in terms of distances (measured and perceived), time durations (measured and perceived), and location sets (observed and considered by the traveler).

3. Serve as external validation/verification devices for synthetic schedule generators

   The synoptic measures described by Janele are strategically important devices for synthetic schedule validation when we simulate a geographic area. They provide a richer database than traffic counts because they contain a longitudinal record of movement for a substantial number of persons. One can envision a variety of validation exercises using different aggregation schemes to perform validation/verification of models ranging from individuals to entire cities.

4. Serve as route choice base estimation data

   Very few activity-based models address the "assignment" of people and vehicles on the transportation system. As we learn more about these new approaches, route choice emerges as the next important step in modeling and simulation. One way to build route choice models is to use GPS traces that are complemented by other variables as the fundamental model estimation database.

5. Main source of data for the study of habit persistence and anchor points

   Limited past research shows people repeat a few basic patterns of behavior over time. It also shows there are anchor points (locations) around which activities and interactions with other persons take place (e.g., home, work, school). Availability of traces offers a unique opportunity to study cycles in behavior and the circumstances under which cycles and habit are broken and replaced with new patterns.

6. Complement travel activity data to produce more accurate data and statistics

   A more mundane task in activity data collection is imputation/augmentation of missing data to identify short forgotten trips and missing from activity/travel diaries and hidden behavioral aspects the diaries fail to capture (e.g., walk from parking lot to the office).

Although there are other data collection and data capture needs for activity-based approaches, GPS traces and related data should become a standard tool to complement activity/travel diary surveys. Vehicle-mounted instruments are the first step in procuring the data needed to inform modeling and simulation but we need to move toward wearable devices that can collect spatio-temporal traces covering entire households and other social groups for all their movements in a day.

**References**

Pribyl O. and K.G. Goulias (2005) Simulation of daily activity patterns. Chapter 3 in *Progress in Activity-Based Analysis* (Ed. Harry Timmermans), Elsevier, pp. 43-65.

Goulias K.G. (2004) Transportation planning and modeling, Chapter 29. In *Handbook of Transportation Engineering* (Editor M. Kutz), McGraw-Hill, New York, NY. Pp. 29.3-29-25