Geosimulation, activity microsimulation, and safety analysis: commonalities and the potential gain in merging analytical methods

Konstadinos G. Goulias
GeoTrans Laboratory and Department of Geography
University of California Santa Barbara
goulias@geog.ucsb.edu

Abstract

Highway safety analysis and transportation system simulation share a wide spectrum of spatial and temporal resolutions. At the very fine level of space, forensic engineering techniques for accident reconstruction are applied to individual vehicles and their components, the analysis of conflicts among vehicles and traffic streams operates at the level of roadway intersections and segments of highways, risk assessments as hot spots operate at the level of a city or even higher geographical subdivisions. Similarly along the time dimension we have analyses of second by second movements, time of day of accident occurrence, weekly and monthly accident analysis, and the accident involvement and severity in the life span of individuals. Safety analysis also considers social determinants of accident occurrence and risk as well as perception and attitudes towards risk. Geosimulation and activity microsimulation (activity analysis herein) are also developed at similar levels of spatio-temporal multiplicity to study interactions among persons and to develop synoptic measures of quality of life, transportation system performance, and policy impact analysis. As one would expect there are many potential opportunities for synergy between these two fields of research and practice. On the one hand, activity analysis offers a tremendous amount of data about determinants of accident risk and background information where and when accidents occur. Safety analysis can also gain from a modeling movement to finer social, spatial, and temporal resolution for assessing accident risk in a more detailed, comprehensive, and informative way. On the other hand, perception of risk and attitudes toward risk may inform activity microsimulation in new ways never attempted in the past when theory development and data collection are done with the dual objective of informing safety and activity analysis. In this paper an overview of the methods in the two fields is provided first. This is followed by examples of potential analysis that can be done when merging analytical methods from safety and activity microsimulation. The paper concludes with data collection examples for this type of analysis and a few ideas of next steps in research.

Paper Submitted for Publication in the International Journal of Shipping and Transport

March 2011
1. Introduction

Transportation safety analysis aims at surveying events that can cause death, injury, and property damage. It also aims at understanding the causes and contributing factors to accident occurrence, severity of accident occurrence, risk and exposure to risk, and the actions required to mitigate all these. Geosimulation and activity microsimulation create virtual worlds of transportation systems and the agents living within these systems to assess land use and transportation policies at a very fine spatial and temporal resolutions (named activity analysis herein). These relatively new methods can provide a variety of data for more informed safety analysis. This paper examines briefly surface transportation safety analysis and possible synergies with activity analysis. The aim is to illustrate the possibility of using similar analytical methods, sharing data to enhance the methods used, and ultimately build planning tools that enhance quality of life and minimize accident occurrence jointly.

As one would expect surface transportation safety analysis examines a variety of “agents” moving within the transportation system. These agents are individual vehicles (e.g., passenger cars or commercial vehicles) operated by drivers of a wide range of training and experience in a multitude of environments and individual psychophysical conditions. Moreover, operational and policy analyses of the behavior and performance of these agents and the performance of the system within which they act requires the use of a wide spectrum of spatial and temporal analytical fidelity and resolutions (Huang and Abdel-Aty, 2010). At the very fine level of space, forensic engineering techniques for accident reconstruction are applied to individual accidents with key objective the understanding of the physical and human factor causes of accident occurrence (Brach and Brach, 2005). At a somewhat coarser spatial resolution of a traffic system component, risk assessment uses a variety of tools to examine roadway intersections (e.g., analysis of conflicts among the different movements) and traffic and highway design principles to minimize accident occurrence at intersections and at different elements of roadways (see http://safety.fhwa.dot.gov/intersection/signalized/presentations/nchrp500_v12/ - accessed March 2011).

At higher levels of spatial abstraction such as an urban environment we also aim at identifying high risk locations in a city (Nitz et al., 1995, Schneider et al., 2004), and even perform spatial analysis of much larger geographical regions such as counties of a US state or
regional departments in a country (Aguero-Valverde and Jovanis, 2006, Chappellon and Lassare, 2010). Based on these analyses preventive designs and policies can be developed that change by jurisdiction to match local circumstances and the resident population (e.g., the resident mix of very old and very young drivers). Similarly along the time dimension, we have analysis of time of day of accident occurrence that shows there are specific times during a day during which most accidents occur (Folkard, 1997), we also find weekly and monthly accident analysis showing seasonality (Langlois et al., 1985), and studies of accident involvement and severity in the life span of individuals (e.g., the well known high fatality rate of the very young and the very old – see Abdel-Aty et al., 1998) but also the differential rates of changing/improving accident occurrence among different age groups (Cheung and McCartt, 2011). We also have other “time” dimensions such as time-to-task, which is the duration of driving before an accident occurs, and the combination of temporal on and off driving patterns (Kaneko and Jovanis, 1992). Many macro-temporal and daily rhythms are correlated and they also contain a variety of spatial and social context informants (Nitz et al., 1995, Radun and Radun, 2006). In fact, safety analysis also considers social determinants of accident occurrence and risk as well as perception and attitudes towards risk with many exogenous and endogenous determinants (Milia et al, 2011). All this motivates the increasing use of analytical methods that are more sophisticated, complex, but also more informative than simple descriptive statistics or basic regression techniques (see the review in Huang and Abdel-Aty, 2010, and the injury severity model review by Christoforou et al., 2010). The conference for which this paper was written is the ultimate demonstration of the complexity of safety analysis and the need for approaches that attach different issues within this field with multiple analytical tools. A similar increase in sophistication happens in urban transportation system simulation and focus here on one quantitative type of analysis.

Transportation system simulation, particularly in regional travel model building, aims at the development of a planning support system that allows planning agencies to build scenarios of change. These scenarios are used in regional transportation plans to draw policy pathways of possible actions and to assess the potential success of these policies (e.g., the introduction of electric vehicles, change in land use to increase density and mix diversity, introduction of pricing schemes for roadway and city centers). The use of this type of models can be traced back to the 1950s when the freeway beltways were planned for Chicago, Detroit, and Pittsburgh in the US. At that time simulation models were mostly based on spatial interaction regional science theories.
aiming at linking land use to transportation system performance. Over time we experienced a gradual movement away from coarse level analyses and increasing acceptance of behavioral theories and analytical methods motivated by policy questions moving away from the construction of major projects and closer to policies that require understanding of markets.

The use of these simulation models is increasing among practitioners and it is reaching unprecedented sophistication in terms of theories, data collection, and modeling and simulation techniques. A detailed review of the models used in terms of the wide spectrum of spatial, temporal, and social dimensions is provided for integrated land use transportation models by Hunt et al. (2005) and for activity and travel behavior models by Henson et al., (2009). The plethora of advances reviewed in those papers includes models and experiments to create computerized virtual worlds and synthetic schedules at the most elementary level of decision making unit using microsimulation and computational process models; data collection methods and new methods to collect extreme details about behavior and to estimate, validate, and verify models using advanced hardware, software, and data analysis techniques; and integration of models from different domains to reflect additional interdependencies among land use, transportation and telecommunications. Most important for the discussion here is to realize that recent developments of model building and regional simulation, in practice and in research laboratories, have created methods that simulate the life of people in a variety of contexts from the very long term to represent choices individuals make to form households and purchase homes to the extremely short term such as the route each person follows from an origin to a destination (see the examples in Salvini and Miller, 2005, Bradley et al. 2010). In this way one can perform activity and travel demand analysis at any level of spatial and temporal aggregation.

Very important is also the use of synthetic population generation to recreate all residents and travelers of a region and simulate their behavior. In parallel, these same models are used in forecasting and they are integrated with land use and demographic models to create scenarios of possible change in space and time (called Geosimulation herein). In this way future populations are created for a given region and through simulation each person is assigned an activity and travel schedule under different scenarios of policies. In the next section an illustration of a typical regional simulation model system is offered. This is followed by a section on possible synergies between the two analytical fields.
2. Geosimulation and Activity Microsimulation

Policy analysis is often done using software that creates scenarios of change to assess the impact of proposed policies on the environment. These are tools that are broadly defined as Planning Support Systems (see Geertman and Stillwell, 2004, for a wide ranging review). In modeling and simulation for travel behavior analysis these tools have a sixty year old history of development and they span a multiplicity of disciplines (see the review by Henson et al., 2009). Most important, however, for the review here is a modeling and simulation trend to build “integrated” models. Integration in this case is intended as the connection of models that were designed for different purposes but as they are applied for different time periods one model needs to provide input for another. To be more precise these are not just models but groups of models intended to perform major functions. For simplicity of presentation let us consider a case study of a large hypothetical region of many millions of residents. Typical simulation for this region will contain a backbone string of models that aims at recreating in a somewhat abstract form what happens in the region today and then create scenarios of future development that includes case studies of continuing current policies and scenarios of introducing major policy changes.

Schematically, Figure 1 shows a typical overall model structure that includes a few major components in the way they are practiced in a few leading US planning organizations. The entire modeling process starts with population synthesis that attempts to recreate the resident households and their household members. This is a person-by-person and household-by-household recreation based on a basic method called iterative proportional fitting algorithm and enhanced to account for data imperfections as in Beckman et al. (1996) for the use in TRANSIMS, Guo and Bhat (2007), Auld et al. (2008), and Ye et al. (2009). The input to this software and block of methods is the spatial organization of the simulated area in the form of geographic zone-specific univariate distributions of person and household characteristics provided by the US Census and other demographic projections provided by other agencies. For future years these distributions are forecasts based on either externally provided data and/or internally simulated data using population evolution models that modify households as they progress in time.
The outputs from this block of models are persons and households and a short list of their characteristics. These need to be enriched by adding each person’s education level, driver’s license holding, and employment and their relevant longer term choices such as location of schools and jobs. This is the task of the block labeled long term choices. In the first year of the simulation workers are identified using a variety of regression models and lookup tables that allow us to assign these attributes in a probabilistic manner. Using these characteristics, household income is computed as a function of its major determinants (e.g., race, presence of elderly individuals, education level of members of households, and employment industry of workers in the household). In some applications we also find residential tenure models (i.e., to own or rent a dwelling unit) and housing type models (i.e., assigning households to specific dwelling units type such as single-family detached, single-family attached, apartment, and mobile home or trailer). In this portion of the simulation an important model is car ownership and type for each household. This eventually in the simulation is used to predict the composition
of non-commercial regional vehicle fleet mix that is used as input to the emission estimation software later in the sequence schema. In one such application for California, models predict for each household vehicle holdings, body types, fuel types, age, and use (miles) simultaneously. Predicted household vehicles are then allocated to drivers in each household based on a suitably designed probabilistic model.

In this way the simulation cascade on the left hand side of Figure 1 produces the spatial distribution of all the residents in a region with as many social and demographic characteristics as desired and located at any geographic level needed, provided suitable methods to allocate residents to each level are created. This in essence yields a virtual Census of microdata that can be “mined” for information and spatial analysis and it can also be used to draw samples for in-depth analysis. It is also possible to focus (and zoom into) a specific subarea (e.g., a city, neighborhood, or even a roadway segment) to perform more detailed analysis and modeling or to develop synoptic measures at different spatial or temporal summaries. All this is then used as input for models that simulate the activity and schedule of persons and their households for multi-day periods (Auld and Mohammadian, 2009) or just a single day depending on the data available and planning applications (http://www.cc.utexas.edu/prof/bhat/CEMDAP.htm).

For a given day the next set of models of Figure 1 (block labeled daily schedules and choices) simulates the life of persons by recreating their daily activities and travel by obeying relationships of interdependence within a household (e.g., assigning parents to give rides to children going to school, allocating time to household grocery shopping, and so forth). The models in this way create a complete description of the movement of each individual over space and time that is congruent with the movements of the rest of the household in which each person belongs. In this way, for each person, we have information about the type of activity, when, where, how long, with whom, in what sequence, and interrelationships with other persons and locations in the activity-travel engagement pattern. The end result resembles the data from a complete household travel diary.

Then, the output from this block is used in different ways. The data are analyzed to identify different behaviors or they can be converted into inputs for a routing and traffic assignment model to produce predicted traffic volumes on highway links. This can be done using the detailed time of day scheduling of activities. For example, in an application for the Southern California Association of Government, which is the largest US planning agency (see
we have developed policy scenarios and studied policy impacts on timing decisions of individuals (e.g., advancing or postponing the starting time of their trips) and changes in destination choices throughout a day. Figure 2 provides an example of this output from Santa Monica, California, which is a community along the Pacific Ocean. We also made assessments of pricing policies on a variety of social and demographic segments to illustrate the asymmetric impact of these policies. In addition, we coupled this output with the more traditional four-step model routines to perform traffic assignment and emission estimation. Moreover, we are also advancing along the path of using detailed routing algorithms that can track in the simulation individual vehicles and eventually compute emissions at fine spatial and temporal resolution. Subregions with all the detailed network and traffic data can also be extracted to use in traffic simulation.

The models of the middle panel of Figure 1 take the spatial distribution of residents and businesses in year t=0 and evolve them to the next year using again probabilistic models and a variety of urban economy simulation techniques (Hunt et al., 2005). The right hand side panel of Figure 1 is a repetition of the first year repeatedly applied for many years (there are currently running scenarios to the year 2035) at different levels of resolution depending on the analytical needs of the policy one wants to study.
Number of persons at each place by activity type and traffic flow from 7:00 AM to 8:00 AM

Number of persons at each place by activity type and traffic flow from 11:00 AM to 12:00 Noon

Figure 2 Output of the Activity Analysis Simulator SimAGENT
3. Synergies

The brief safety analysis review in the introduction and the example of the previous section hint to a few major elements that safety analysis and activity analysis have in common. Ideally, both behavioral models and safety models should be estimated using longitudinal data and the specialized literature has a few examples that also use multilevel and multiperiod regression models. In addition, the Transportation Research Board’s Strategic Highway Research Program 2 is currently funding a longitudinal naturalistic driving data collection study in many different driving environments to provide this type of data (SHRP, 2010). Similarly, for activity-travel behavior model building some of the most advanced methods are based on Geographic Positioning Systems combined with trip maker interviews to understand activity and travel scheduling as well as interactions among travelers spanning relatively longer periods (Axhausen et al., 2002, Auld and Mohammadian, 2009). This activity of longitudinal data collection complements the more traditional panel surveys that interview repeatedly households and their members over time to establish the determinants of their propensities to change behavior among years (Golob et al., 1997). Equally important is the “width” of data collection in which the subjects are observed in different spatial contexts and in fact this is a dimension recognized as extremely important in most industrialized countries with data collection that is spread throughout their territory. As research and practice progress to more sophisticated methods we also realize that a third dimension in data collection, “depth”, is also required to provide context and inform about additional determinants of behavior and safety related events. In safety we know that social and situational circumstances impact a driver’s performance. Similarly, in activity and travel behavior situational factors are very important in understanding behavioral paradoxes. In fact, we see in safety analysis and travel behavior increasing acceptance that values and attitudes should be identified and relevant data collected to explain otherwise unexplainable occurrences. One can also envision household surveys designed to collect data for safety and behavior jointly to achieve economies of scale and to provide a common platform for cross-fertilization between the two fields of study. Data analysis as one would expect for safety analysis and travel behavior analysis employ the same multivariate regression models and in typical transportation statistics courses are taught with examples from both fields. In fact we are moving rapidly to statistical modeling methods that enable analysis of multiple causes and
multiple effects recognizing the multiple scales in data structures and these somewhat more advanced techniques are also becoming practice. In addition, mapping and spatial statistics techniques that are used to identify locations of special interest (e.g., hot spots vs high accessibility locations) or to explain phenomena that display spatial and temporal correlation and heterogeneity are also used in both fields.

Activity analysis, however, offers an additional unique opportunity for safety analysis because it provides added background to the different events (accidents, fatalities, injuries) with detailed statistics about population, infrastructure characteristics, and changes in the environment. Therefore, a few of the easiest to accomplish tasks are: (1) Develop exposure statistics that are based on the amount of traveling (vehicle kilometers, trips, or even congestion levels) and data on accident occurrence; (2) Perform risk assessments in space, time, and social groups and any combination of the above using the space-time evolution of geosimulation and any observed data on accident occurrence; (3) Develop background maps when studying risk based on population composition at specific localities and its change over time, including car ownership and type as well as fleet mix and its spatiotemporal evolution; and (4) Study changes through data matching and secondary data analysis in population values, norms, and attitudes and their correlation to specific behaviors as well as accident occurrence.

In addition to these data analytic research tasks there are other potential areas of synergy that can benefit both research areas. The first is in the development of a comprehensive conceptual framework that combines safety and behavior objectives (e.g., ecological socioeconomic models of change). Bronfenbrenner’s Bioecological model of change is one candidate. The model is based on a developmental theory with its core the zone of proximal development: Human development in the life span is a journey through increasingly more complex reciprocal interaction between a human organism and other organisms, objects, and symbols in its environment (Vygotsky, 1978). This model is known as the Person-Process-Context-Time (PPCT) model (Bronfenbrenner, 2005). Person is defined as person factors representing individual differences in physiological and psychological states, tempo, and biological intensity of reactions. Process is the stream of psychological acts that are called proximal processes and considered to be the primary engines of development. Context is the physical, socio-emotional, and mental setting in which behavior takes place. Time is ontogenetic (person development) time, cohort time, and historical time. Proximal and distal in terms of relationships in
microsystems, mesosystems, exosystems, and macrosystems. I used this model to illustrate the need to think about car ownership and mobility from a developmental perspective and illustrated the continuity of behavioral processes in the life span of mobility decisions. Of course this is not the only conceptual framework that may apply to both settings but it is worth examining further.

With or without a common theoretical framework we can also envision a data collection activity/project that aims at satisfying objectives in a general inventory building effort that requires a large sample to represent the entire population. This type of data collection can happen immediately after major Census campaigns with the objective of describing the behavior in an area, state, or the entire country. Attached to this major “main” survey one can imagine the creation of “satellite” surveys with a much smaller sample that collect in-depth information about a focused topic (e.g., attitudes toward driving). The satellite design is such that enables expanding the answers to the entire population extracting a same sample from the sampling frame of the main survey. The combination of a set of safety related satellite surveys, activity and travel surveys, and a main household survey yields data of unprecedented coverage of diverse topics and may offer unique opportunities for research that bridges across different fields of inquiry.

Acknowledgment

Sincere thank you goes to Professor Amalia Polydoropoulou and Maria Kamargianni who organized the superb symposium on “Road Safety. European Reality and Perspectives” at the Homerium Cultural Centre of Chios Municipality (September 24-25, 2010) for which this paper was prepared. Dr. Yali Chen is also acknowledged for preparing the maps of Figure 2. Funding for this paper was provided by the University of California Transportation Center and the University of California Office of the President. This paper does not constitute a policy or regulation of any Local, State, or Federal agency.
References


