An Overview of GeoSimulation for Smart City Planning, Design, and Operations

Konstadinos G. Goulias
Department of Geography and GeoTrans Laboratory
University of California Santa Barbara
1832 Ellison Hall, UCSB
Santa Barbara, CA 93106-4060
USA
goulias@geog.ucsb.edu

Abstract—Smart city propositions emerge from a private firm need to develop new markets for advanced technology and public motivation to manage resources efficiently to mitigate the negative impacts of activities in the city. These initiatives also need to support planning medium- and long-term goals in comprehensive ways that satisfy the three-dimensional sustainability mission of environment-economic-society and support provision of indispensable public services. In this paper we review these services, describe a few currently accepted views about smart cities, and explore the components of a tool in activity-based microsimulation to forecast travel demand. This provides a short description of the state of the art in GeoSimulation and allows the identification of opportunities for smart city modeling and simulation together with gaps in information, immediate next steps, and future directions to create a simulator for smart cities.

Index Terms—GeoSimulation, Comprehensive Planning, Smart Cities, Sustainability, Behavior.

I. INTRODUCTION

Private industry motivation for smart cities emerges from a need to develop new markets for advanced technology and in particular information and communication technology (ICT) that is integrated with the services of a city. Examples include the IBM smart planet initiative and its smart cities outgrowth (http://people4smartercities.com/editorial/city) and the CISCO "internet of everything" and its own collaboration such as the "City Protocol" (http://cityprotocol.org/collaborative-platform). These initiatives are the beginning of a wider movement to increase a city's ability in optimizing services but also facilitate the flow of material. This is a heavily "operations" oriented emphasis and motivated by current needs. Public motivation for advancing ideas about smart cities, however, also emerge from a need to manage resources efficiently and mitigate the negative impacts of activities in the city (see the example in Barcelona - http://smartcity.bcn.cat/en) today, and in the medium- and longer-term future. The usual three-dimensional sustainability mission of environment-economy-society is fundamental in these considerations and smart city definitions have accounted for this. One of the most popular visions shows smart cities developing along six dimensions that are mutually strengthening and include smart economy to advance competitiveness, smart people to enhance social and human capital, smart governance to ensure participation by all, smart mobility to improve transport and accessibility, smart environment to protect natural resources and sustainability, and smart living to advance quality of life [1,2]. Advancement of smart city policies along these six dimensions is reminiscent of comprehensive planning [3] and provides a framework to develop metrics as performance indicators [2, 4, 5]. Creating and monitoring suitable indicators and standards is one of the hardest problems in planning. Using these indicators cities can use ranking to benchmark and identify gaps and projects to close these gaps; it is also used as a way to showcase cities that are able to streamline smart city planning and development [6, 7, 8, 9, 10]. Considerable funding from central agencies in Europe and strategic plans to move towards implementation of these ideas are also at a mature stage with proof of concept in real life applications [11]. Urban planners are also discovering the opportunities of smart city initiatives, but, also to the need of building bridges with more traditional city planning [12]. This is particularly important in the metrics used to judge success and their monitoring.

Many of the indicators that are in the repertory of smart city assessments are already in the planning-modeling-simulation-monitoring framework of regional transportation plans provided by other planning and operations activities at different geographical scales but with the same sustainability objectives [13, 14, 15]. Monitoring and assessing plans and scenarios of policies within them is done using a considerably large data procurement program and extensive modeling and simulation technologies. These tools are mainly designed to assess the impact of policies on travel behavior and predict travel demand changes but also to address problems of
housing and urban development. In California, which is leading the United States in sustainability legislation, this research area is heavily motivated by a need to assess policies quantitatively. These model systems include demographic microsimulation of individuals and their households as well as simulation of employment shifts spatially. This is done under an umbrella of comprehensive transportation planning and sustainability assessments that include the estimation of vehicle kilometers traveled and CO2 emission reduction. Key motivator to do this are policies to increase land use density and diversity with the hope that people will use more public transportation, walk, and bike, while, reducing the use of private cars. Other policies such as pricing of transportation services and demographic and employment trends are also examined. These policies influence many facets of daily behavior such as activity participation and destination choice, but, they also influence residential, work, and school location choices as well as car ownership and car type purchasing. Ultimately, modeling of a region, metropolis, or urban environment aims at quantifying material flows (e.g., passengers and goods) under different policy scenarios. This has similarities with Urban Metabolism (UM), the accounting of flow of energy and materials as well as sinks in a city, which is becoming an important tool in achieving longer-term optimality of city functioning. UM is a somewhat more comprehensive approach than transportation planning because it includes the entirety of energy consumption and flow of material from all human activities but less detailed in the policies examined [16]. In strategic planning for regions that include many cities, large scale simulation of policy portfolio scenarios are used for regional transportation plans and UM assessments. Quantification of flows of material and people using simulation software is also aided by "big data" to fill gaps left by incomplete data collection initiatives. All this points to: (a) convergence of policy analysis needs (e.g., pricing of services); (b) simulation methods and modeling paradigms (e.g., simulation of the life of people and the facilities they use); and (c) development of indicators that can be shared among different analysis projects (e.g., emission of CO2 by mobile and stationary sources). Smart city initiatives can take advantage of this convergence and use many indicators that are produced for other planning initiatives. Transportation planning initiatives can also benefit from smart cities actions in the entire spectrum from policy definition to traffic operations. For example, smart cities emphasize e-services that should be included and integrated within the policy objectives of transportation (e.g., examine the relationship between travel and e-commerce) and joint methods to address policies, enhance the analytical background and solutions, and ultimately become the integrators of many policy initiatives. To illustrate all this with an example, I will first describe briefly a new application method that was developed by academic institutions and currently refined and used in practice to improve GeoSimulation of a large region (18 million residents in 119 cities). At the end, I will sketch a few options for possible developments to create simulators for smart cities. I will also point out to gaps found in the smart city literature.

II. GEO SIMULATION

The simulation paradigm discussed here is called SimAGENT (Simulator of activities, greenhouse emissions, networks, and travel). This tool is an activity-based microsimulation model system that was developed at UCSB in collaboration with UT Austin and Arizona State University [17] for the Southern California Association of Governments (SCAG). SCAG is the largest metropolitan planning organization in United States, representing six counties, 191 cities and more than 18 million residents. SimAGENT includes population synthesis that recreates the entire resident population in this region, provides locations for residences, workplaces, and schools for each simulated person and household, estimates car ownership and type, and includes an array of other personal and household characteristics. Then, a synthetic schedule generator recreates for each resident person in the simulated region a schedule of activities and travel that reflects intra-household activity coordination for a day [18]. These synthetic activity and travel daily schedules are then converted into vehicular travel and assigned to the network. The assignment output is then used to produce estimates of fuel consumed and pollutants emitted (including CO2) by different classes of vehicles. In the background we also have a detailed map of accessibility indicators that are computed using an inventory of business establishments (e.g., each building classified by its land use, number of employees, and a variety of other data).

The simulation process starts with PopGen in which the entire resident population is synthetically generated/recreated person-by-person and household-by-household based on the method described in Pendyala et al. [19]. The input to this software is the spatial organization of the simulated area in the form of zone-specific univariate distributions of resident person and household characteristics provided by the US Census for a baseline year. As the population is recreated on a person-by-person and household-by-household basis, these distributions are used as the control totals for each spatial unit of analysis (approximately 4,000 geographic subdivisions called Traffic Analysis Zones in this version of the model system but easily expandable) in an iterative algorithm that starts from a multivariate set of relationships (in essence a cross-tabulation) among the person and household variables used as seed information. For future years, these distributions are forecasts provided externally and can be modified at will. The multivariate set of relationships can be kept constant (assuming a steady state of demographic relationships) or can be changed to capture the impact of changing population composition and associated relationships including but not limited to age, birth rates, and household size.

To represent employment opportunities and the spatio-temporal distribution of activity participation locations, SimAGENT uses opportunity-based accessibility indicators at the level of the US Census block (203,191 US Census blocks cover the entire study area). In this way, we represent the ease (or difficulty) of reaching 15 different types of industries (representing the opportunities for activity participation) from each of these blocks within 10, 20, and 50 minutes of roadway travel buffers from each of the 203,000 pegs [20, 21].
types of industries included in SimAGENT are: (a) Agriculture, forestry, fishing and hunting and mining; (b) Construction; (c) Manufacturing; (d) Wholesale trade; (e) Retail trade; (f) Transportation and warehousing and utilities; (g) Information; (h) Finance, insurance, real estate and rental and leasing; (i) Professional, scientific, management, administrative, and waste management services; (j) Educational; (k) Health; (i) Arts, entertainment, recreation, accommodation and food services; (m) Public administration; and (p) Other services (except public administration). Different accessibility values are obtained for the morning peak period (6 to 9 AM), midday (9 AM to 3 PM), evening peak period (3 to 7 PM), and at night (7 PM to 6 AM) capturing not only the different roadway conditions, but also the patterns of opening and closing of businesses during the day by allowing within each period above to also have different opening and closing hours of each industry type. The resident population with its detailed characteristics and a selection of indicators of the accessibility they enjoy are the inputs for the next block. Accessibility indicators are used in many of the behavioral models of the baseline year. Then, for subsequent simulation years they are modified based on models to capture changes in the spatial distribution of economic activities but also accounts for time-of-day variation in the level of service of simulated networks.

SimAGENT also includes migration, educational achievement, employment and industry type in which individuals work, formation and dissolution of households, birth and death of individuals, and an array of spatial choices they make in their lifetime including residential location, work location and school locations. In this way each person and household created in the population synthesis, and located in each zone of the study region, is given additional characteristics. For example, when we examine persons in college a model is used to assign a college location, which is also a hierarchical function of accessibility. Workers are identified using a labor force participation model that is a function of age, gender, education, and presence of children in the household. Employed persons are then assigned (in a probabilistic way) to their type of industry, work location (which is also a function of accessibility), weekly work duration, and work flexibility. Each individual is also assigned a driver’s license depending on age, gender, and ethnic group. Using these characteristics, household income is computed as a function of race, presence of elderly individuals, education level of members of households, and employment industry of workers in the household. Then, a residential tenure model (own or rent) and a housing type model are used to assign each household to a single-family detached, single-family attached, apartment, and mobile home type of residence.

An important model in this simulation system is car ownership and type. This type of model in essence determines the predicted non-commercial regional vehicle fleet mix that is used as input to emission estimation software. This is also particularly important for California because of the expected market penetration of electric cars and the incentive programs created at the state and federal levels in the US to promote this type of technology. This model system can be used to assess different incentive structures promoting environmentally friendly technologies in cars. The model handles multiple vehicle holdings, body types, fuel types, age, and use (miles) simultaneously [22]. This model includes 55 alternatives for body type/vintage (9 body types – Subcompact Sedan, Compact Sedan, Mid-size Sedan, Large Sedan, Coupe, Cross-utility car, SUV, Van, Pickup) and 5 vintage categories (New to 1 year, 2-3 years, 4-5 years, 6-9 years, 10-12 years, >12 years), 47 alternatives for vehicle make, and several hundreds of models across the many body type/make combination categories. Explanatory variables include household composition indicators and residential accessibility. After this step, each of the household vehicles is allocated to a driver in each household based on a probabilistic model using as predictors variables such as gender, education, and employment. The outcome at this point of simulation is a spatial distribution of all the residents by different social and demographic levels as well as employment and school locations assigned to each person. In addition, each household is assigned to a housing type. This resembles a complete Census of the resident population and can be done at any level of spatial aggregation. The next set of model simulates the life of persons in a day.

For each synthetically generated household and person within each household, daily activity and travel patterns are created in two blocks of models: (a) the generation step in which work and school activity participation and timing decisions are created, children’s travel needs are estimated and an allocation of escort responsibilities to parents takes place, and independent and joint activity participation decisions are modeled; and (b) application of the scheduling of activities that produces the sequence of activities, with the departure and arrival times, activity durations, mode for each trip, and determination of the location of each activity. 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 engagement pattern. The majority of these behavioral facets are modeled using econometric models (discrete choice models, hazard models, regression equations, and cross-tabulations) that use as explanatory variables household composition and characteristics (e.g., household size, number of children, number of vehicles owned), individual characteristics (e.g., gender, age, race/ethnicity, job type), location characteristics (e.g., opportunity-based accessibility indicators, infrastructure available, transit availability, population density), and tour/mode/stops alternative attributes (e.g., in-vehicle travel time, distances to destinations). The end result resembles a complete diary of all activities and travel by all persons in a household including trips, stops, activity types, activity start and ends times, and modes for all trips. SimAGENT simulation includes 17,317,284 persons in 5,721,914 households. Simulation of policies cause changes in destinations, activities alone and with others, durations, and modes creating “realistic” behavioral changes. In the initial model designed pilot tests of policy scenarios were performed to examine timing decisions of individuals (e.g., advancing or postponing the starting of
trips). At this point of the simulation we have movements of people among activity places represented as points at the center of polygons. These are defined by geographic subdivisions that are used by agencies to simplify the complexity of the built environment called Traffic Analysis Zones (TAZ). We also have trips that go from one TAZ to another classified by each car simulated above and by each of the modes included in the simulation (e.g., car sharing, public transportation and so forth). The next step takes these trips and assigns them to a network of roadways and public transportation. The microscopic scale of SimAGENT allows employment of a variety of specialized software at different scales. The most detailed in space and time is the Transportation Analysis SIMulation System (TRANSIMS) developed by the Los Alamos National Laboratories [23]. The software is used to first spatially allocate all households to a detailed network and then identify routes from each origin to each destination and subsequently simulate vehicles on the network. In SimAGENT, approximately 64.5 million activity records and 156.5 million travel plans are routed in the network using TRANSIMS. The output contains the duration of each activity, location where it was completed, mode used to arrive to each activity location, route followed by each simulated vehicle, other data about interaction in household of activity and travel, and the vehicle used. It also converts externally provided origins and destinations of trucks into travel plans. Using the inventory of the daily paths of all the members of households and the use of their vehicles, the vehicle trajectories, and a second-by-second emissions estimation model called Comprehensive Modal Emissions Model (CMEM), is also used to provide emissions per network link [24]. Additional byproducts using different combinations of data from this simulator also include activity assignment from the TAZ geography to business establishments [25], spatial allocation of households and their carbon footprint at specific buildings [26], and development of policy scenarios and vehicle type market penetration based on household vehicle fleets and their utilization [27, 28].

III. SMART CITY INDICATORS

Unavoidably when performance indicators are designed to make comparisons among different cities an emphasis is given to breadth (coverage of multiple services) at the cost of depth (a more complete description of behavior). The ISO 37120:2014 [29] includes indicators about the economy, education, energy, environment, finance, fire and emergency response, governance, health, recreation, safety, shelter, solid waste, telecommunications and innovation, transportation, urban planning, wastewater, water and sanitation. This is comprehensive and aims at describing an entire city. However, it is envisioned as a static and aggregate snapshot and does not address spatial heterogeneity of these standards. In contrast, urban planning indicators and standards provide an equally ambitious enumeration of areas covered but emphasize micro design indicators and standards. Planning for smart cities most likely needs to use a spatiotemporally multiscale vision of monitoring to plan, design, and evaluate smartness of smart cities and allow analysis to move across different scales. The SimAGENT example shows that one can use modeling and simulation to develop indicators that can be upscaled or downscaled to match the smart city planning needs. However, this requires to also identify the scale at which smart city planning should take place and this again depends on a variety of factors not included explicitly in smart city studies. One can envision scale to be derived in an inductive way (bottom up from particular instances) and deductive way (top down from general principles). As discussed in other papers in this symposium there are ways to use citizens as sensors to shape the context and desired environment smart cities should target. Moreover, smart city applications also show the importance of what planners call public involvement and stakeholder feedback. An added key question in all this enterprise is about the optimal key performance indicators to guide the design. If in fact we aim at improving quality of life we need to consider jointly objectively measured indicators (by analysts using the best data available such as accessibility to physical and virtual services) and subjectively measured indicators (perception of availability and quality of services via user surveys with strategically chosen questions). SimAGENT shows we can do very well in the objectively measured indicators by fusing data from multiple sources and recreating realistic worlds computationally to describe spatial heterogeneity the in quality of service offered in a city. We could envision being able to do the same using spatial perception surveys [29, 30] and subjective well-being [31], but only lately we begin to understand the data collection needs for this complex relationship between perception and behavior [32]. Moreover, SimAGENT and the examples offered in this paper are heavily focused on transportation and its expanded scope that touches many other areas such as the economy, environment, energy, and housing and can be supported by modeling and simulation data and software representing the evolution of urban economy in comprehensive ways [33]. Key city services such as water provision and waste can also be added in an urban metabolism simulation framework that produces key indicators even for very large scale environments [34]. ICT and "big data" are used in transportation to understand interactions with travel behavior [35] and to complement modeling and simulation [36]. These activities, however, lack a true integration of ICT in modeling and simulation to account for the impact of ICTs in simulating travel. This can be remedied by smart cities planning efforts but first requires institutional support to emerge as a major policy area. In this way we will be able to count on an institutional framework for data collection and indicator creation congruent with the scope of smart cities. Although EU investment was made to do exactly this, the lack of "threats" like high pollutant concentrations and climate change does not provide enough motivation for policy enforcement in the way it is done in California, which is able to mandate specific policy actions (e.g., sustainable community strategies and mandatory...
decrease in CO2 emissions to meet a target). An opportunity to do exactly this comes from the emerging policy aiming at maintaining and improving quality of life (objective and subjective). At the core of quality of life is human interaction and the design of smart cities should bring at the forefront a goal of increasing human interaction and avoid isolation. My proposition for the next generation smart city plan-design-operate framework and indicators is to: (a) develop a new layer of indicators that derive a union of measures from Urban Metabolism and Comprehensive Planning to produce more complete ecological indicators; (b) increase the depth of the e-service indicators using lessons learned in smart cities worldwide; (c) add indicators about the use of time and the human interactions in using time; (d) commence a program of defining subjective and perception that include subjective-well-being indicators. Key to all this, however, is also the creation of a measurement and reporting program that allows to measure the variation of these indicators in space and in time.

REFERENCES

Konstadinos G. Goulias (Ph.D. in Engineering University of California-Davis, 1991, MS Engineering University of Michigan-Ann Arbor, 1987, Laurea (5 years and a thesis) in Engineering University of Calabria, Italy, 1986) is a professor of transportation at the University of California Santa Barbara Department of Geography and director of the GeoTrans laboratory. He served as professor of transportation in the Civil and Environmental Engineering Department of the Pennsylvania State University from 1991 to 2004 where he also directed transportation research centers and programs. He chaired the Travel Behavior and Values Committee and the Task Force on Moving Activity-based Approaches to Practice for the Transportation Research Board (TRB) and served in many other organizations and task forces including the Institute of Transportation Engineers and the American Society of Civil Engineers. Goulias edited two books (Transportation Systems Planning: Methods and Applications published by CRC Press and Transport Science and Technology published by Elsevier) and published more than 270 research reports and papers. He is the co-founder and co-editor in-chief of the journal Transportation Letters, he is also member of the Editorial Advisory Board of Transportation Research Part B and the Journal of Intelligent Transportation Systems. Goulias worked in Australia, Germany, Greece, Italy, Japan, the Netherlands, Portugal, and the United States developing new household survey methods and other data collection techniques as well as statistical and spatial modeling techniques, simulation frameworks, and expert reviews of technologies and engineering practice and policies.