

A framework to generate virtual cities as sandboxes for land use-transport interaction models

Rounaq Basu

Massachusetts Institute of Technology

Roberto Ponce-Lopez

Tecnológico de Monterrey

Joseph Ferreira

Massachusetts Institute of Technology

Abstract: One of the major critiques of land use-transport interaction (LUTI) models over the ages has been their over-dependence on individualized software and context. In an effort to address some of these concerns, this study proposes a framework to construct “virtual cities” that can act as sandboxes for testing different features of a LUTI model, as well as provide the capability to compare different LUTI models. We develop an approach to translate any prototypical transportation infrastructure network into a plausible land use zoning plan and synthetic population that are suitable for spatially detailed LUTI microsimulation of the virtual city. Disaggregate units of spatial geometry, like parcels and postcodes, are generated using geospatial techniques applied to the transportation network. Households and jobs are randomly sampled from an actual city, and allocated in the virtual city based on matching density gradients. Students are matched with schools and workers are matched with jobs to complete the calibration of a synthetic population for the virtual city. Following the adjustment of behavioral models to complement the reduced scale of the virtual city, we demonstrate the integration between the land use and transportation simulation components in our LUTI model, *SimMobility*. The benefits of faster convergence times and shorter simulation times are clearly demonstrated through this exercise. We hope that this study, and the open-source releases of the *SimMobility* software with the virtual city database, can accelerate experimentation with LUTI models and aid the transition from individualized LUTI models to a common shared integrated urban modeling platform.

Article History:

Received: June 2, 2020

Accepted: September 4, 2020

Available online: February 28, 2021

1 Background

1.1 A brief history of LUTI models

Transportation networks and land use patterns are known to mutually influence each other, and drive spatial socio-economic processes (such as development and migration) in cities. Accessibility, which forms a bridge between mobility and land use, is a crucial element in shaping patterns of residential location, and urban and regional development (Hansen 1959). Based on Hansen's seminal work, the first generation of land use-transport interaction (LUTI) models came about in the late 1960s, with an aim to better inform transport models in reducing congestion. Although spatial interactions were captured through gravity-based measures, the primary focus of these models was to improve the forecasts of trip origins and destinations. Examples of first-generation LUTI models include the Metropolis model (Lowry 1964), and ITLUP (Putman 1974). Unfortunately, these models left a lot to be desired, with their mechanicalness, theoretical shortcomings, data-hungriness, and complexity (among other limitations) being harshly criticized in a review of large-scale models by Lee (1973).

The emergence of new theoretical frameworks in the fields of econometrics and behavioral economics (such as the random utility theory) in the mid-1970s enabled modeling disaggregate behavior by focusing on the prediction of choices among multiple discrete alternatives. Applications such as the choice of travel mode (Lerman 1976) and residential location (McFadden 1978) provided the impetus for the second generation of LUTI models. The field branched out into two different directions at this time, based on the method by which spatial processes were represented. The first approach was to use regional economic models that focused on a framework to represent trade flows between different economic sectors. Examples include TRANUS (de la Barra 1989), MEPLAN (Echenique *et al.* 1990), and PECAS (Hunt and Abraham 2005). Contrastingly, the other approach sought to improve the scale and method of representing housing transactions and land markets, as can be seen in MUSSA (Martinez 1996) and DELTA (Simmonds 1999).

The third generation of LUTI models was driven by transformative technological advances with significantly greater computational resources and power coupled with more efficient data storage. Cell-based models of temporal urban change and spatial evolution, such as SLEUTH (Clarke *et al.* 1997), came into prominence at this point. In particular, agent-based microsimulation became (and remains) a popular choice for constructing state-of-the-art LUTI models that could represent urban regions at a disaggregate level, wherein individuals, households, firms, and developers could be represented as agents. The early 2000s witnessed the development of three LUTI models - UrbanSim (Waddell *et al.* 2003), ILUTE (Salvini and Miller 2005), and ILUMASS (Strauch *et al.* 2005) - that successfully implemented the agent-based microsimulation approach, and remain active in pursuing improvements to modeling structure and agent representation to the present day. Interested readers may refer to Iacono *et al.* (2008) for a more detailed comparative review of these LUTI models.

1.2 LUTI models for uncertain futures

While technology and computation power have improved by leaps and bounds over the last decade, LUTI models have yet to capitalize on the potential for a more holistic and generalizable modeling approach. In a recent review, Miller (2018) urges the community to go beyond "mere" land use-transport interactions, and aim for a "system of systems" representation of urban regions. Technological and behavioral trends have changed drastically, but none of these new developments have been adequately dealt with by most urban models (Wegener and Spiekermann 2018). The need of the future is to

develop integrated urban models (IUMs) that can remain useful as policy analysis tools in light of potentially disruptive technologies, environmental phenomena (such as climate change and natural disasters), public health crises (such as epidemics and pandemics), and socio-demographic processes (like migration and geriatric trends). Miller (2018) goes on to suggest that IUMs of the future must be fundamental in their behavioral foundations and flexible in their methodological and computational implementations to be “agnostic” modeling environments.

In a recent review of LUTI models with an eye towards their appropriateness for an automated future, Hawkins and Nurul Habib (2019) argue that operational LUTI models were largely developed during a period of relative uniformity in mobility choice sets, and are unenthusiastic about the usefulness of such models in the shared mobility era. More general concerns about the inability of LUTI frameworks to appropriately model emerging trends such as peak car or the impact of information and communication technologies (ICTs) on activity-travel patterns have also been raised (Van Wee 2015). Two major challenges that plague LUTI models are the integration of activity-based travel demand models and appropriate measures of accessibility (Acheampong and Silva 2015). Additionally, a recent review by Lopes *et al.* (2019) highlights how currently operational LUTI models do not adequately recognize all mutual interactions between activities, land use, and transport.

A combined team from MIT and the Singapore-MIT Alliance for Research and Technology has been pursuing the development of a state-of-the-art agent-based microsimulation LUTI model (*SimMobility* - Simulation of Future Urban Mobility) built on underlying activity-based discrete choice models of travel behavior. Considering land-use, transportation and communication interactions, *SimMobility* can be used for a variety of applications, including implementation of intelligent transportation systems, estimating vehicular emissions, evaluation of alternative future scenarios, and generation of innovative policy and investment strategies (Adnan *et al.* 2016). This project was initiated in 2010, and is still under active development, which gave us the opportunity to address some of the aforementioned concerns from the LUTI modeling community. Our recent efforts demonstrate how *SimMobility* can be used to understand the impact of automated mobility on housing-mobility choices (Basu and Ferreira 2020b; Basu and Ferreira 2020a), private vehicle ownership (Basu and Ferreira 2020c), and the future of mass transit (Basu *et al.* 2018a), in addition to highlighting the potential for sustainable mobility futures (Oke *et al.* 2019).

1.3 The need for sandbox virtual cities

Almost all LUTI modeling activity has occurred in the Americas and Europe, e.g., ILUTE in Canada, UrbanSim in the US, MUSSA in Chile, ILUMASS in Germany, and PECAS in the UK, Canada, and California. While being driven partially by funding sources, this phenomenon has also led to LUTI modeling frameworks being overly dependent on particular contexts and softwares. Due to this shortcoming, there are limited opportunities to evaluate the replicability, transferability, and generalizability of LUTI models. The ambitious ISGLUTI project in the 1980s tried to address some of these concerns by comparing the transferability of operational LUTI models at the time, but fell short of encouraging the development of a shared testing platform or study area (Webster and Dasgupta 1991).

Our own experiences in developing *SimMobility* build upon their efforts and echo the sentiments expressed by Miller (2018) that call for a “virtual lab” where researchers can use common tools and shared datasets within an agnostic modeling environment. Although this is a useful step towards the “virtual lab” ecosystem, issues of verification and validation still remain. A common testing ground, or a “virtual city”, might help alleviate these concerns. We recognize that LUTI models often require

very detailed disaggregate data, most of which are often proprietary to maintain privacy. Therefore, creating a virtual city as an open-source sandbox for LUTI models can help accelerate the vision of collaborative and transferable research in this domain.

Moreover, LUTI models often involve months, if not years, of effort directed towards data cleaning and preparation. In a comparison of LUTI models employed by small and mid-sized urban planning agencies, [Clay \(2010\)](#) reports that the average time spent on model development is around three years with costs going up to as much as \$750,000 per year. An open-source virtual city can help LUTI modelers test their frameworks without having to invest as much time, effort, and monetary resources in model setup.

Drawing from the typology of models of cities for planning proposed by [Batty \(2007\)](#), we postulate that a virtual city should be constructed with a digital iconic representation of a hypothetical city's transportation infrastructure, which should be further augmented using a symbolic representation of people and places resembling an actual city. Similar efforts to construct virtual spaces, which are primarily motivated by privacy concerns, are commonly seen in the domain of infrastructure security. A particular example is Micropolis, a virtual city of 5,000 residents, which was used to simulate the spread of fires on electricity networks ([Bagchi et al. 2009](#)) and water demand networks ([Brumbelow et al. 2007](#)). In the urban planning domain, simulations of virtual urban spaces have been used majorly for educational purposes or as commercial games. Specific examples include Community Land Use Game (CLUG) by [Feldt \(1972\)](#), CityDev ([Semboloni et al. 2004](#)), SimCity ([Adams 1998](#)), and Second Life ([Warburton 2009](#)). However, to the best of our knowledge, there is a gap in the literature related to the formulation of a methodology for construction of virtual cities, as well as case studies of usage of such virtual cities for testing LUTI models.

The major contribution of this paper is a methodological framework for the construction of a virtual city as a sandbox for testing LUTI models. Based on the aforementioned guidelines, we begin with a hypothetical transportation network and construct the virtual city based on plausible judgments about land use constraints and judicious sampling of population and job data from an actual city. It is worth noting that the specific virtual city we showcase is calibrated using data from Singapore, therefore it can serve only to compare different LUTI models (e.g., SimMobility and UrbanSim). However, our methodological approach can be easily used to create several other virtual cities using data from possibly any city in the world. Having access to multiple virtual cities would allow the community to evaluate the transferability of a particular LUTI model, e.g., we could compare the performance of SimMobility across a variety of typological virtual cities that have been created using data from various "real" cities from around the globe.

It is worth mentioning that the literature is quite rich regarding the construction of road networks ([Thomson and Richardson 1995](#)) and transit networks ([Derrible and Kennedy 2011](#)) using graph theoretical and network science approaches, but we could not find an example of constructing a fully populated and functioning city using the transport network. The remainder of the paper is structured as follows. The next section outlines the methodology for the virtual city construction, while the following section provides details about the use of this virtual city in testing our LUTI model (SimMobility) in addition to providing simulation results. The final section of the paper provides concluding remarks and outlines directions for future research efforts.

2 Methodology

We will first describe the data inputs that are required for the virtual city construction, i.e., a detailed transportation network and a land use zoning plan. The following sub-section will outline how disaggregate spatial units, such as parcels and postcodes, are generated based on these data inputs. We will then describe how populations of individuals, households, housing units, and jobs are randomly sampled from an actual city, and appropriately allocated inside the virtual city. Next, the assignment of households to housing units and individuals to jobs is outlined. Finally, the land use and transportation components of the LUTI model are connected in order to determine converged accessibility measures consistent with the initial deployment of residents and jobs across the virtual city. At this point, the LUTI model is ready for simulating urban futures in the virtual city.

2.1 Data requirements

The virtual city construction framework requires two basic data inputs: (a) a detailed transportation network suitable for microscopic transport simulation, and (b) a zoning plan sketching land use characteristics and constraints. In the construction of this virtual city, we take a certain element of creative liberty in designing both requirements in tandem (as they should be), e.g., providing adequate transit stops in residential suburbs and the downtown core. We would like to highlight that this is not the only way to do so, and modelers might even create a virtual city to resemble a scaled-down version of an actual city. It is also plausible to expect that the transport network is exogenously provided by a transport planning agency, and we (as urban planners) would then need to define only a zoning plan that is expected to adequately complement the provided transport network.

Two major approaches to constructing a virtual transport network are discussed in the literature. The first approach is to use latent space models to graph the network topology of cities (Zhou *et al.* 2015), but it requires knowledge about the underlying nodes, or intersections, of the city as a data input. The second approach is arguably more popular and easier to implement because it is based on rules set by the modeler. Sun *et al.* (2002) used a rules-based approach that used buildings, population densities, and disallowed zones as inputs and provided a road network as an output. Vitins and Axhausen (2010) defined ‘network grammars’ that were effective in defining rules-based approaches and evaluating what hierarchies within cities look like. Weber *et al.* (2009) provided a comprehensive rules-based method of generating virtual cities for use in video games and other applications. Their approach can be lauded for being generative and temporal, i.e., the city starts out from a specified central node and permeates outward from there, creating new nodes and edges at a given rate. However, this has been critiqued for requiring significantly large processing times, since the level of detail is extremely disaggregate (e.g., the degree of each intersection). Vanegas *et al.* (2012) proposed an approach that created smaller urban parcels based on land-use, which can be effective for filling a grid of street networks into areas where their requirement is known ahead of time, e.g., in the sub-division of a neighborhood.

Despite the presence of fairly substantial literature in this domain, the lack of a generalizable methodology that can construct both road and transit networks based on land use is apparent. Similarly, there is a gap in the literature with respect to the construction of an appropriate zoning plan based on given transport networks. We adopt the latter as one of the motivating research questions for this study. We develop a rule-based approach to adding plausible land use patterns and residential/employment distributions to exogenously specified transportation infrastructure. This is done in a manner that retains certain spatial and demographic characteristics of an actual city, such that the

virtual city can be a useful sandbox for experimenting with different LUTI models that use behavioral models calibrated for a real city.

A hypothetical road network, which spans 290 kilometers, is shown in Figure 1(a). The road network data includes both node information (such as number of turns, number of edges, etc.) and edge information (such as number of lanes, maximum speed, capacity, etc.), which can enable its use in a detailed traffic micro-simulation. Based on the road design, we also construct a hypothetical bus network comprising bus routes with 72 distinct stops (see Figure 1(b)), and a hypothetical MRT/subway/metro network comprising eight stations (see Figure 1(c)). The transport infrastructure motivates the creation of 24 Traffic Analysis Zones (TAZs), each of which covers an area of 4.1 square kilometers on average, as shown in Figure 1(d). The virtual city is thus designed to span a total area of about 100 square kilometers. Overall, these specifications can be assumed to have been provided by a hypothetical transport planning agency.

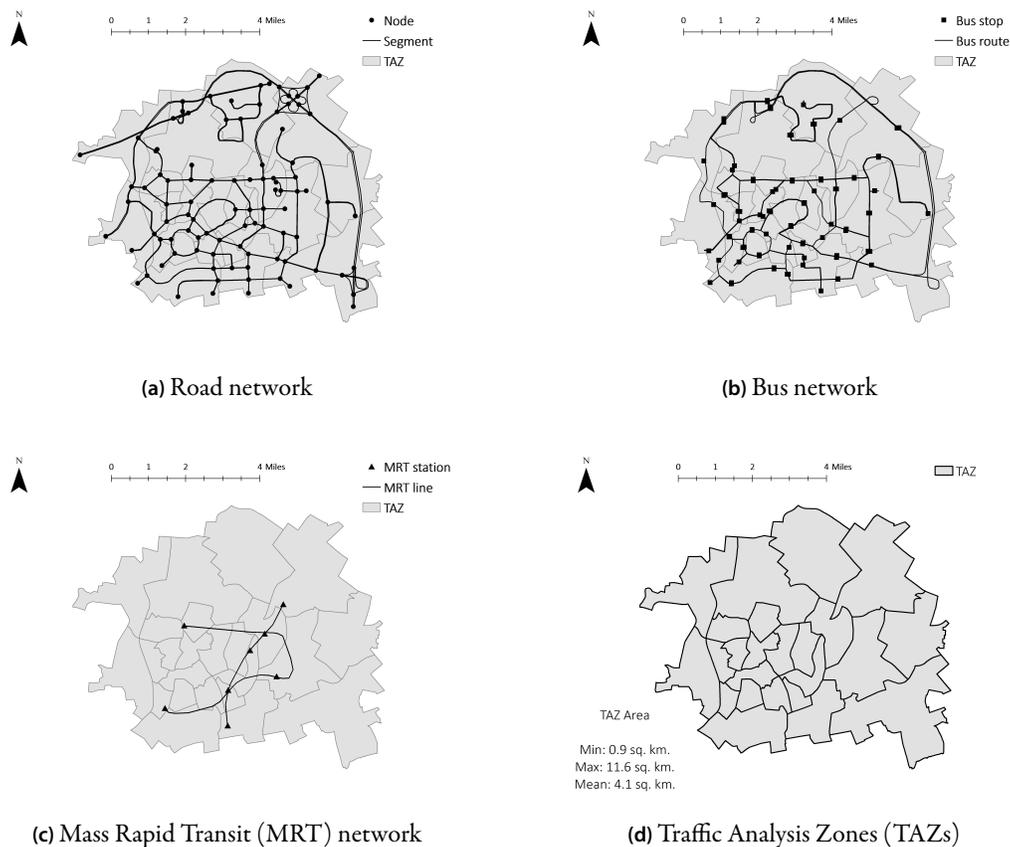


Figure 1: Transportation network in the virtual city

For the transport infrastructure provided in the virtual city, we assume that the region has a monocentric urban morphology. The land use zoning plan, which portrays the spatial restrictions for allocating residential and work locations, is presented in Figure 2. Based on the monocentric city design, we place commercial land use in the center of the city. The south-west of the city is designated as industrial land use, and a few zones between downtown and the south-west are marked as mixed commercial and industrial to provide an appropriate transition. The east of the city is mostly residential,

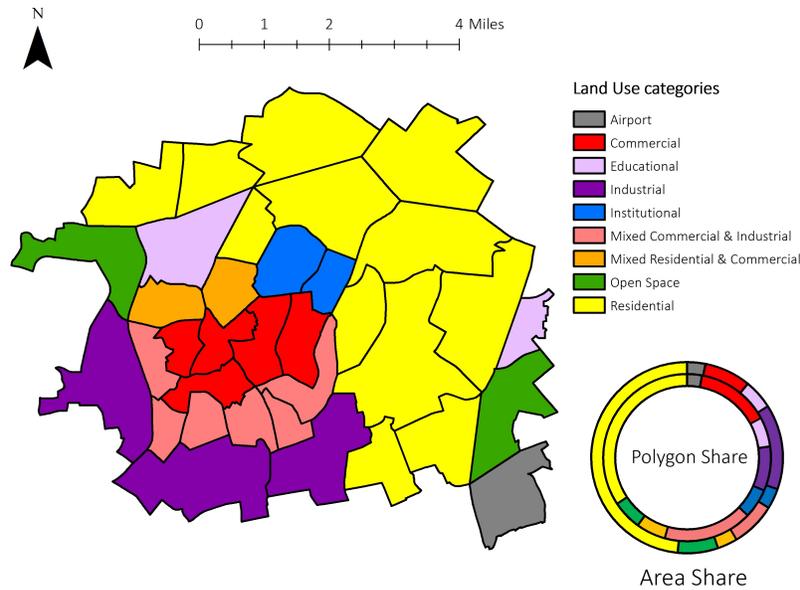


Figure 2: Land use zoning plan for the virtual city

with a couple of zones marked as institutional and educational each to represent government offices and school/university campuses. A couple of zones on the edges of the city are marked as open spaces (which leaves room for additional development in the future), along with an airport in the south-east of the city. In a few cases, large TAZs in the suburbs have been broken into multiple land use polygons having different land use codes. Some of these constraints are useful in spatially segregating different job types (such as manufacturing and industrial vs. office workers), while others provide a spatial distribution of households and jobs that is plausible for urban micro-simulation.

2.2 Generation of disaggregate spatial geometry

The two data inputs considered thus far are usually provided by different sources, i.e., the transport planning agency provides the transport network, while the the land use zoning plan is under the purview of the urban planning (or regional development) agency. Even if well-coordinated across agencies, the TAZs and land use polygons are too macro-scale to enable vehicle-level simulation of urban mobility. To take full advantage of state-of-the-art micro-simulation capabilities, it is desirable to generate spatial zones that are fine-grained enough to be at the block-level so that the synthetic population can be dispersed with sufficient granularity. The locations where people live and perform activities are distributed among buildings and land areas, whose locations are generally approximated by address points and zip codes (or postcodes, in the case of Singapore). Individuals move among these zones using the transport network to carry out daily urban activities. The population and building data for a city are generally tagged with such postcodes, and each postcode is associated with a building or address point location at or near the center of its zonal area. With this in mind, we propose the construction of parcels and postcodes as disaggregate units of spatial geometry.

Although people live and perform activities in buildings, they enter the transport network at various “entry points” when they need to travel between different buildings or zones. If the locations of

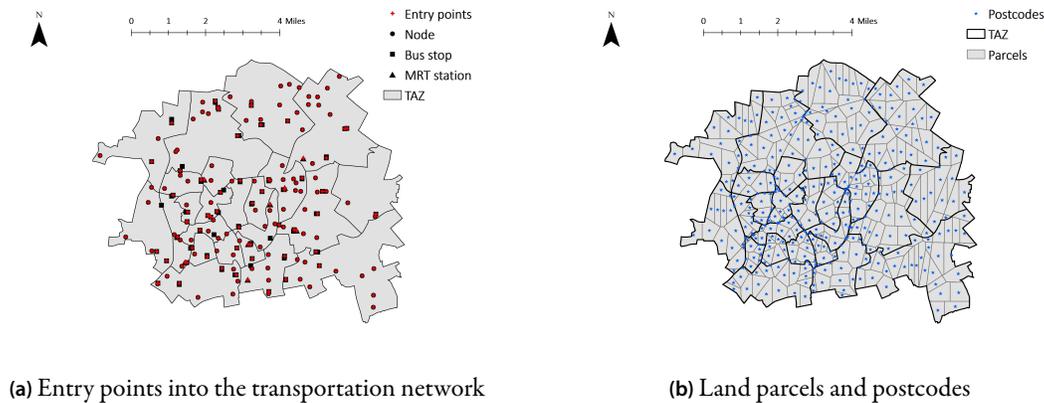


Figure 3: Generating disaggregate spatial geometry for the virtual city

buildings were known ahead of time, we could assign the nearest nodes of the road network as the entry points. However, this is not the case, and we need to work it out in reverse. Therefore, we consider all nodes of the road network, all bus stops, and all MRT stations as possible “entry points” into the transport network, which are shown in Figure 3(a). We then treat the entry points as seed generators for Voronoi cells or Thiessen polygons, which are generated using Fortune’s algorithm (Fortune 1987). This approach has been used extensively in several geospatial applications, such as measuring rainfall in an area (Thiessen 1911), and analyzing the source of epidemic outbreaks (Johnson 2006). These Thiessen polygons can be interpreted as areas that are most likely to be served by each entry point of the transport network. However, some of these areas were adjusted further for three main reasons. First, they should not cross TAZ boundaries, therefore we cut areas that did so into multiple smaller regions to maintain geometric consistency with the TAZ boundaries. Second, we sub-divided large areas of open space and undeveloped land so that future development could be considered for manageable chunks of land, as is often seen in the real estate market. Finally, we removed slivers that arose from these processes of sub-division as they most likely stem from our basemaps (i.e., TAZs, land use polygons, and Thiessen polygons) being drawn at different scales and levels of accuracy.

After going through the multiple steps mentioned above, the “cleaned” Thiessen polygons that we are left with are designated as land parcels. Accordingly, we define postcodes as the centroids of these resulting land parcels. In this case, the land parcels are at the scale of Census blocks in American cities. A map of the 175 land parcels and postcodes against a backdrop of TAZs is presented in Figure 3(b).

2.3 Population allocation through density-matching

After the generation of appropriately disaggregate spatial geometry, the next step is to populate the virtual city with households, jobs, and buildings. While it is plausible to arbitrarily create a synthetic population of such elements, the random spatial distribution of the population might be subject to critique. There is strong historical evidence that the long-term evolution of cities is based on transport network evolution, wherein population density follows transport networks, and consequently, accessibility. Examples of transit-oriented development are found aplenty in Copenhagen (Knowles 2012) and other European metro areas (Papa and Bertolini 2015), while the post-WWII strategy of American development to prioritize building highways is often considered as one of the primary causes of suburbanization and sprawl (Baum-Snow 2007). Based on these findings, we proceed to randomly

sample the population of households, jobs, and buildings from an actual city (Singapore, in this case) and allocate them to the virtual city based on transport network density. It is worth highlighting that there is no reason to favor Singapore over other cities or metro areas at this stage, and the only reason for our choice is the availability of relevant data at the appropriate level of detail.

Transport network density can be measured in a variety of ways. For example, we could compute the linear road length within each zone. However, the roads tend to be the boundaries of many of our zones, which is why we choose to count entry points, comprising nodes from the road network and public transit stops, close to each postcode (which is where buildings are located) as our measure of transport network density. We assume the service area of each postcode to be 500 meters, because this represents how far people are willing to walk from urban activity centers at the postcodes to the entry points of the transport network (Basu *et al.* 2018b).

The population sampling technique follows a three-step process. First, we compute the number of entry points that lie within a 500-meter Euclidian buffer of each postcode in the virtual city, and assign the transport network density for all 175 postcodes. The population and job densities are computed in a similar manner for all postcodes in the actual city (i.e., Singapore) using the synthetic population data. Second, the postcodes in Singapore (which represent buildings in dense areas, and blocks in less urbanized areas) are ranked based on the density measure, and categorized into 175 groups to correspond with the virtual city postcodes. Third, we select an appropriate sampling rate (10% in the case of Singapore), based on which households, jobs, and buildings are randomly sampled from the synthetic population based on the group in which the postcode lies. For example, if the postcodes are ordered in increasing order of density, 10% of households and jobs will be randomly sampled from the 175th ordered category of Singapore postcodes (i.e., the ones with the highest population densities) and assigned to the 175th virtual city postcode, which has the highest transport network density.

This process allows us to guarantee that a household residing in a densely populated area in Singapore is allocated to a virtual city postcode that is located in an area with high density of transport network “entry points”. The same can be said for jobs as well. Based on this density-matching technique, we also oversample buildings (housing units) and firms (jobs) to make sure that the virtual city has a suitable vacancy rate, in keeping with the rate observed in the actual city from which we draw the population. It is worth noting that a realistic vacancy rate is essential for housing market models to function properly.

The final element in this stage is to allocate educational institutions to the virtual city. We define three types of institutions: (a) elementary (i.e., primary or secondary) schools, (b) high schools, and (c) universities. We use the land use zoning plan to postulate that high schools and universities can be placed only in areas that are designated as educational land use due to requirements of large spaces. However, elementary schools can be placed in both educational and residential land use polygons, as they are more likely to be located near residential clusters. We determined the number of each type of educational institution based on the same sampling technique used for the population, but using the density of households with children this time. Using this information and the aforementioned zoning constraints, we randomly selected a subset of postcodes from the eligible zones and assigned appropriate educational institutions to them accordingly. Finally, we end up with 30 elementary schools, four high schools, and one university that is spread out over 10 postcodes (which are essentially buildings in this case). Figure 4 presents the spatial distribution of the educational institutions overlaid on the land use zoning plan. The completed version of the virtual city is shown in Figure 5.

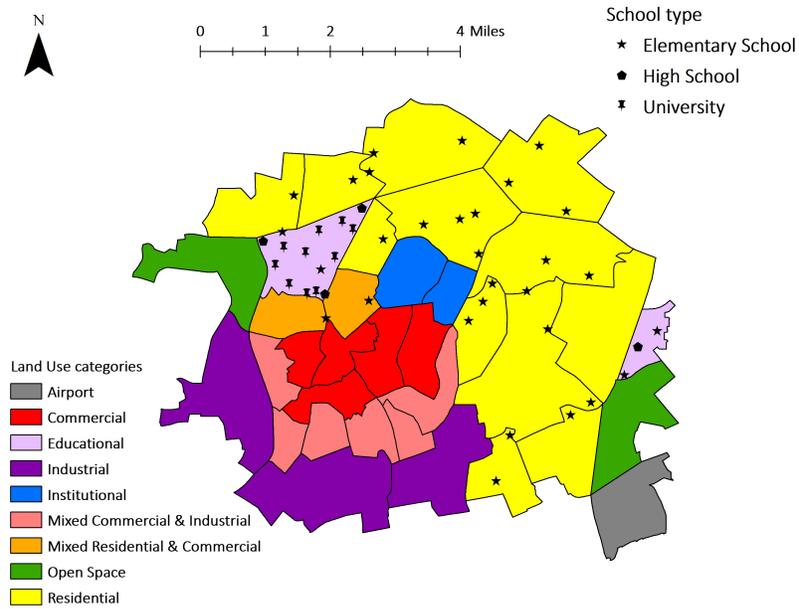


Figure 4: Educational institutions in the virtual city

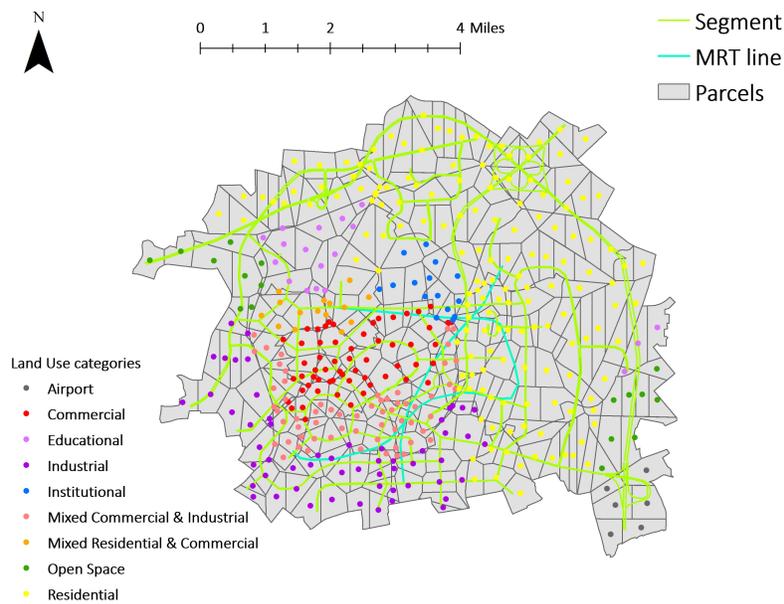


Figure 5: An overview of the completed virtual city

2.4 Assigning workers to jobs and students to schools

While we have successfully created a synthetic population for the virtual city, we have yet to assign workers to specific jobs and students to specific schools. This “job assignment” process is essential

for simulating daily commuting in the transport simulator component of LUTI models. Since there is only one university, it is relatively trivial to assign every student of age 18 years or above to one of the postcodes in the university. Children, i.e., students below 18 years of age, are assigned to the school that is closest to their residential location, provided there are enough empty spots. Appropriate capacities are placed on the schools based on their category, i.e., high schools have a larger capacity than elementary schools, so it is possible that several students might be going to a school that is second- or third-nearest to their home. In such instances, our algorithm tried to minimize the total travel time for students, while respecting school capacity constraints.

Unfortunately, assigning workers to jobs in the virtual city is not as trivial. A discrete choice model is used for job assignment, wherein accessibility measures form a subset of the explanatory independent variables. The alternatives presented by this model are at the TAZ-level, because the transport simulator usually estimates travel demand and provides origin-destination (OD) matrices at the TAZ-level. Hence, the attractiveness of a TAZ as a workplace location is assumed to be proportional to a weighted sum of the number of jobs available in the TAZ and the accessibility that a worker would experience if they lived in their current housing and had a job located in that TAZ. Interactions of the accessibility measure with socio-demographic characteristics of the worker and job industry sectors further enrich the heterogeneity captured by the discrete choice model.

Ideally, we'd like this "fixed home-varying workplace" accessibility to be an activity-based accessibility (ABA) measure provided by the transport simulator, as it captures the expected maximum utility derived by the individual from all possible activity patterns carried out throughout the day (Dong *et al.* 2006). However, it is not possible to compute these ABAs the first time around, because the initial (or seed) workplaces of workers are not known. Therefore, we use travel skims (i.e., OD matrices of travel time and cost) pertaining to the transport network as the seed accessibility values to simulate one full day in the transport simulator and complete one round of job assignment. Now that all workers are assigned to a workplace (albeit temporarily), we can then leverage the capability of our transport simulator to compute ABA measures for all workers, and use these values in the job assignment model. Finally, we repeat the job assignment process based on the updated ABA values, and update the job location for each worker accordingly.

3 Testing LUTI models for the virtual city

This section provides details on how the virtual city can be explored as a viable study region for micro-simulation of land use and transport interactions. First, we describe our agent-based LUTI micro-simulator, *SimMobility*, and its components. Second, we provide a framework to assess the accessibility integration between the housing market simulator and the daily transport simulator using the virtual city. Finally, we provide empirical results from the simulations and discuss the usefulness of having a virtual city as a sandbox.

3.1 SimMobility

SimMobility is a multi-scale agent-based microsimulation platform that incorporates time-scale dependent behavioral modeling through activity-based frameworks (Adnan *et al.* 2016). Through the consideration of interactions between transportation and land use, *SimMobility* can be used for a variety of applications ranging from implementation of intelligent transportation systems to evaluation of alternative future scenarios. *SimMobility* encompasses three major components:

- **Long-Term (LT):** This detailed land use-transport simulator involves the creation of a synthetic population of individuals, households, firms and establishments (Zhu and Ferreira 2014). This is followed by household-level residential location and vehicle availability choices, and individual-level employment or education location choices. The temporal scale of this component ranges from days to years.
- **Medium-Term (MT):** This component contains a mesoscopic supply simulator coupled with a microscopic demand (daily activity) simulator (Basu et al. 2018a). Daily travel decisions like mode choice, route choice, activity-travel patterns, and incident-sensitive (re)scheduling are considered at the temporal scale of minutes to hours, up to a single day.
- **Short-Term (ST):** This microscopic traffic simulator involves lane-changing, gap acceptance, route choice, and acceleration-braking behavior at the temporal scale of seconds to minutes (Azevedo et al. 2017).

The *housing market model*, implemented through a daily bid-auction approach, forms the crux of the LT modeling framework. Households are awakened using pre-computed probabilities (conditional on current tenure status and age of household head) through the *awakening model*. Next, the *screening model* creates behaviorally consistent and computationally feasible choice sets of housing units that each household could potentially bid on. The *hedonic price model* provides the estimated market price for all housing units in the study region. Through the *willingness-to-pay model*, households then evaluate the expected consumer surplus for each unit in their choice set and compare against their current residential unit. Finally, households choose to bid on the unit that provides the maximum expected consumer surplus, provided it is higher than that provided by their current unit. After completing a possible residential relocation, households then re-evaluate their private mobility holdings through the *vehicle availability model*. Additionally, individual members of such relocated households re-consider their job location choices through the *job assignment model*. While this is the current implementation of the LT modeling framework in SimMobility, alternative long-term choice models could be easily swapped in and readily applied to the virtual city constructed from the transportation infrastructure using our technique.

The LT and MT components are connected through activity-based accessibility (ABA) measures that are disaggregate utility-based measures of the value of alternative daily activity patterns. Simulation of alternative scenarios in MT enables measurement of individual-level ABAs that would result from participation in those scenarios through logsums. For example, consider an individual living in the i th Traffic Analysis Zone (TAZ) and working in the j th TAZ. It is crucial to evaluate hypothetical scenarios such as (a) *fixed home-variable work*, where the individual can choose from all possible TAZs for their work location while keeping the residential location (TAZ i) fixed, and (b) *variable home-fixed work*, where the work location (TAZ j) is fixed but the residential location is allowed to vary. For every possible combination of home and work TAZs, MT provides a logsum value for each individual in the synthetic population. These ABAs are used as explanatory independent variables in LT models, in the form of individual-specific logsums in the job location choice model, and household-specific logsums in the residential location choice and vehicle availability choice models.

3.2 Assessing the LT-MT integration

A “full loop” simulation of the LUTI model (see Figure 6) is required to assess the LT-MT integration, wherein LT models are used to obtain a stable equilibrium of housing and job locations, followed by a stable equilibrium of daily activity patterns in MT. The feedback loop occurs through the synthetic

population and accessibility measures. LT passes a synthetic population of individuals (whose housing and job locations, and private mobility holdings are known) to MT, which then simulates one whole day of travel, based on which accessibilities calculated using the updated transport network costs and times are passed back to LT. These accessibilities are then used to recalibrate the synthetic population by adjusting housing and job locations and private mobility holdings.

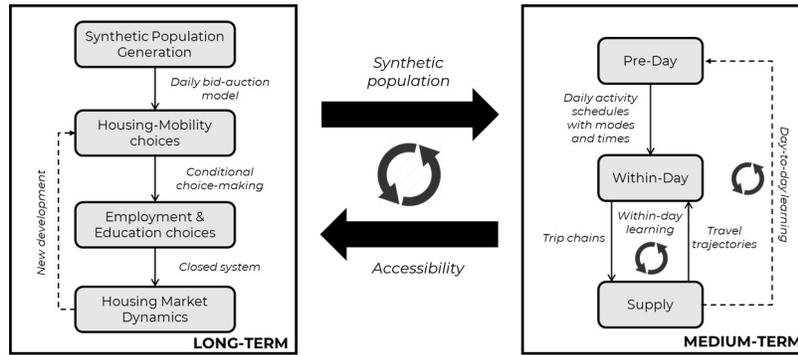


Figure 6: Integration of the Long-Term and Medium-Term modules in SimMobility

However, using only one MT simulation might be fallacious because the seed network cost and time tables come from travel skims provided by the transport planning agency. It would be more prudent to conduct multiple iterations of a “full loop” of day-to-day learning in MT, wherein multiple days are simulated with each day’s realized travel patterns informing the next day’s network costs and times. The day-to-day learning can be interpreted as providing more accurate network information to travelers, who can then make more informed decisions in scheduling their daily activities, not unlike trip planning tools. We could also include within-day learning, wherein one day is simulated multiple times keeping the demand fixed and updating the network costs. This can be interpreted as a quasi-equilibrium of a particular daily activity pattern, which bears some similarity with real-time navigation systems. However, it is worth bearing in mind that including within-day learning would increase the simulation time.

In order to evaluate whether equilibrium has been reached, we select two performance measures. The first is the normalized root mean squared error (RMSN) of the link travel time. Since this measure is calculated over every link in the road network, minimizing the RMSN can be thought of as a “local” optimization of network performance. The root mean squared error (RMSE) at the i th iteration is given by:

$$RMSE_{i,i-1} = \sqrt{\frac{\sum_{k=1}^L (\hat{t}_{k,i} - \hat{t}_{k,i-1})^2}{L}} \quad (1)$$

where L is the number of links in the transport network; and $\hat{t}_{k,i}$ is the realized travel time on the k th link during the i th iteration. Normalizing this RMSE by the travel time during the i th iteration yields:

$$RMSN_{i,i-1} = \frac{RMSE_{i,i-1}}{\bar{t}_i} \quad (2)$$

where \bar{t}_i is the mean realized link travel time during the i th iteration. The second performance measure is the weighted root mean squared error (wRMSE) of zone-to-zone (i.e., TAZ-to-TAZ) car travel times. Since we aim to minimize the wRMSE for all zones in the study area (i.e., the virtual city), this procedure can be thought of as a “global” optimization of network performance.

$$wRMSE_{i,i-1} = \sqrt{\sum_{k=1}^N w_k \cdot (\hat{T}_{k,i} - \hat{T}_{k,i-1})^2} \quad (3)$$

where N is the number of origin-destination (OD) pairs at the zonal level in the transport network; $\hat{T}_{k,i}$ is the mean realized car travel time for the k th OD pair during the i th iteration; and w_k is the normalized weight of the k th OD pair, which is measured as the proportion of trips taken for the k th OD pair among all N OD pairs. The zone-to-zone travel times are updated between iterations for both within-day and day-to-day learning loops using the following update rule. While completely replacing the value might seem tempting, such an approach might cause instability across successive iterations and require a longer time for convergence. Instead, we use a parametric formula to calculate the weighted average of the most recent travel time and the second-to-most-recent travel time realized from the simulation.

$$T_{k,i+1}^0 = \alpha \cdot \hat{T}_{k,i} + (1 - \alpha) \cdot \hat{T}_{k,i-1} \quad (4)$$

where $\alpha \in [0, 1]$ is the update weight that can be chosen by the modeler; $\hat{T}_{k,i}$ is the mean realized car travel time for the k th OD pair during the i th iteration; $T_{k,i}^0$ is the seed zone-to-zone travel time for the k th OD pair during the i th iteration; and $T_{k,1}^0$ is obtained from travel skims provided by the transportation planning agency.

Thus far, we have allocated a synthetic population with behaviorally consistent residential and employment locations to the virtual city, and outlined a framework for obtaining converged accessibility measures that can act as stable feedback mechanisms between the LT and MT components of a LUTI model. A final item of consideration before conducting simulations is to adjust the behavioral models that include spatial measures as explanatory variables. Behavioral models are usually calibrated using household travel surveys or real estate transaction data from an actual city. Since the scale of the virtual city is comparatively smaller, the coefficients for spatial measures such as distances to local amenities, trip lengths, or population density require adjustment by the modeler to maintain consistency of the estimated accessibility measures. Likewise, housing prices and other costs must be expressed in “virtual city” dollars, not Singapore dollars. Following the implementation of this adjustment process, we are ready to test our LUTI model for the virtual city.

3.3 Empirical results

We ran 30 iterations of the “full loop” in MT, where both the demand and supply simulators were updated at the end of each iteration without within-day learning. When within-day learning was enabled, we ran 30 iterations of the “full loop”, where the supply simulator was run five times for each daily activity pattern generated by the demand simulator. Recall that within-day learning can significantly increase simulation time, but can also result in faster convergence due to more supply iterations being simulated. Therefore, empirical results are examined with an eye towards understanding the effect of the update weight parameter (α) and the usefulness of enabling within-day learning on the rate of convergence.

Three values of α are selected, i.e., $\alpha \in [0.2, 0.3, 0.5]$, where a higher value of α indicates a greater weight being placed on the most recent update rather than the seed value. The effects on the RMSN of link travel times are shown in Figure 7. We find that a lower value of α results in a smoother error function that converges faster, while a higher value of α causes the error function to jump around. This is in line with our intuition, and, incidentally, using $\alpha = 0.5$ (which is arguably a moderate weight) leads to the RMSN function failing to converge within 30 iterations. Using within-day learning speeds up the convergence rate, compared to when it is disabled. More importantly, convergence occurs almost exponentially faster in terms of simulated days (but not simulation time). The sub-figures show each value of RMSN at the beginning of each day as a large dot, all of which are connected by straight lines. The within-day RMSN values are represented by smaller dots, and only exist in the “5 within-day iterations” case. When $\alpha = 0.2$, we see that it takes only three days for RMSN to converge when within-day learning is enabled, in lieu of 14 days without within-day learning. However, it is worth noting that these three days actually entail three demand iterations and ($5 * 3 =$) 15 supply iterations, while the alternative includes 14 iterations each of demand and supply. Therefore, it is possible to obtain significant time savings by using the within-day learning approach, provided convergence occurs fast enough.

A similar exercise is carried out for the “global” performance metric, i.e., the wRMSE of TAZ-to-TAZ travel times, as shown in Figure 8. Our observations are mostly similar to those noted earlier. A lower value of α provides a faster convergence rate, which is somewhat accelerated by enabling within-day learning. However, the gains are not as significant in this case. For example, when $\alpha = 0.5$, it takes 11 iterations to converge without within-day learning, which can be reduced to 9 iterations by enabling it. It is also worth noting that the PM peak travel times are noisier than the AM peak counterparts, owing to the PM peak being spread out over a wider period of time and experiencing higher volumes of total traffic.

The empirical results suggest that the LT-MT integration in our LUTI model, SimMobility, works successfully and performs reasonably well for the virtual city. If we were to test this using an actual city like Singapore, the same exercise would have taken us months, even with the help of multi-threaded high-performance computing clusters. Apart from model testing, the results also highlight the importance of using within-day learning to obtain a faster rate of convergence, which can result in potential savings in simulation time. Moreover, the update weight parameter should be set appropriately so that network costs and times are updated gradually instead of being replaced completely with every iteration. These insights are useful for modelers who might want to use similar features in their LUTI models on large-scale urban regions.

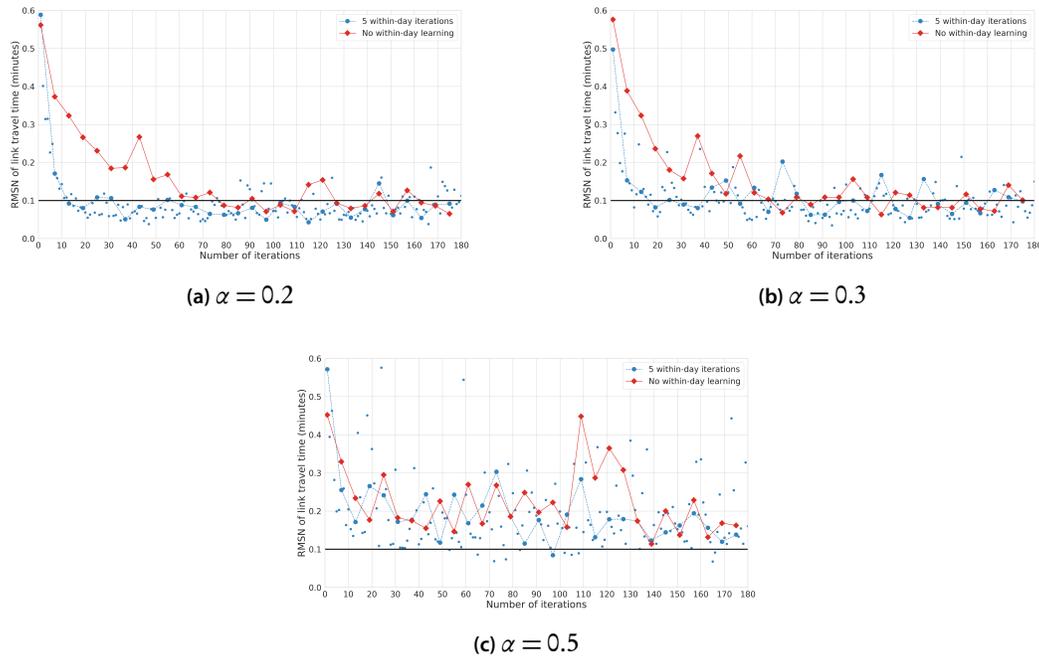


Figure 7: Evaluating the effect of the update weight (α) on convergence of link travel times

4 Conclusion

Despite the numerous critiques of LUTI models, they can be useful tools in testing and analysing policy programs for uncertain future scenarios. However, currently operational LUTI models seem to be overly dependent on individual software, proprietary datasets, and specific contexts. In a call to the LUTI modeling community, Miller (2018) urges for a “virtual lab” ecosystem, where researchers can work on a single integrated open-source software with shared data across contexts. Such an ecosystem can provide numerous benefits, primary of which might be to address the concerns of model generalizability, transferability, and validation. This can even accelerate the evolution of LUTI models to integrated urban models (IUMs) by including important socio-economic and spatial phenomena in their consideration of urban regions.

This study provides a step forward in that direction by proposing a framework to create “virtual cities” as sandboxes for testing LUTI models, which can help alleviate concerns over context dependency. Our approach allows any hypothetical transport infrastructure to be combined with plausible land use constraints and then translated into a virtual city by distributing residents and jobs in a manner that reflects observed patterns in an actual city. Such a virtual city can be a useful sandbox for LUTI explorations since it enables the performance of any arbitrary transportation infrastructure to be simulated using population and employment distributions, and related behavioral models, that are sampled from real cities and spatially distributed to match observed density gradients.

Geospatial techniques are employed to create disaggregate spatial geometry such as parcels and postcodes, with additional adjustments keeping future development and consistent spatial boundaries across layers in mind. In line with literature showing how population density follows transportation infrastructure density, we allocate households and firms that are randomly sampled from an actual city (Singapore) through a density-matching technique. Finally, students are matched to educational

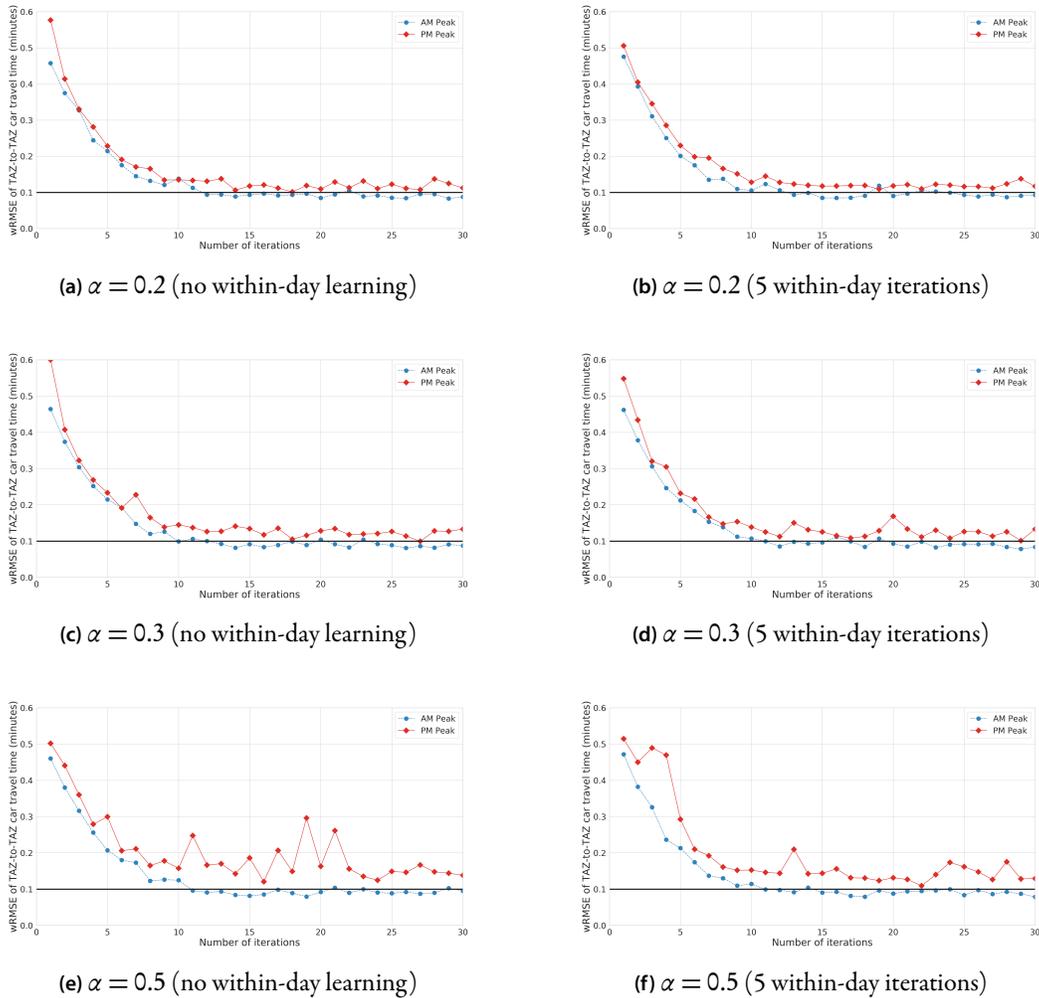


Figure 8: Evaluating the effect of the update weight (α) on convergence of TAZ-to-TAZ travel times

institutions and workers are matched with jobs to construct a calibrated synthetic population for the virtual city.

We also demonstrate the integration between the land use and transportation simulators in our LUTI model, SimMobility, using the virtual city. The benefits of using the virtual city to test this integration are manifold, including faster convergence rates and shorter simulation times owing to the virtual city being at a much smaller scale than an actual city. We aim to release an open-source version of the SimMobility software, along with the virtual city database, by the end of our project. We also hope that the methodology suggested in this paper will allow such virtual cities to be more easily tuned to match the behavior and characteristics of populations in different “real” cities. In this case, it will become easier and more practical to explore differences in LUTI implementations by experimenting with prototypical virtual cities. In closing, we believe that this study and the open-source releases have the potential to facilitate greater collaboration and cross-contextual use of a shared research framework within the LUTI modeling community.

Acknowledgements

This research was funded in part by the Singapore National Research Foundation through the Future Urban Mobility group at the Singapore-MIT Alliance for Research and Technology Center. We particularly appreciate the contributions of Nishant Kumar, Bat-hen Nahmias Biran, Simon Oh, and Chetan Rogbeer, among other past and present members of the SimMobility research team.

References

- Acheampong, R. A. and E. A. Silva. 2015. Land use–transport interaction modeling: A review of the literature and future research directions. *Journal of Transport and Land Use*, 8(3):11–38.
- Adams, P. C. 1998. Teaching and learning with SimCity 2000. *Journal of Geography*, 97(2):47–55.
- Adnan, M., F. C. Pereira, C. M. L. Azevedo, K. Basak, M. Lovric, S. Raveau, Y. Zhu, J. Ferreira, C. Zegras, and M. Ben-Akiva. 2016. SimMobility: A multi-scale integrated agent-based simulation platform. In *95th Annual Meeting of the Transportation Research Board*.
- Azevedo, C. L., N. M. Deshmukh, B. Marimuthu, S. Oh, K. Marczuk, H. Soh, K. Basak, T. Toledo, L.-S. Peh, and M. E. Ben-Akiva. 2017. SimMobility short-term: An integrated microscopic mobility simulator. *Transportation Research Record*, 2622(1):13–23.
- Bagchi, A., A. Sprintson, and C. Singh. 2009. Modeling the impact of fire spread on the electrical distribution network of a virtual city. In *41st North American Power Symposium*, pp. 1–6. IEEE.
- Basu, R., A. Araldo, A. P. Akkinapally, B. H. N. Biran, K. Basak, R. Seshadri, N. Deshmukh, N. Kumar, C. L. Azevedo, and M. Ben-Akiva. 2018a. Automated mobility-on-demand vs. mass transit: A multi-modal activity-driven agent-based simulation approach. *Transportation Research Record*. doi: [10.1177/0361198118758630](https://doi.org/10.1177/0361198118758630).
- Basu, R. and J. Ferreira. 2020a. Can increased accessibility from emerging mobility services create a car-lite future? Evidence from Singapore using LUTI microsimulation. *Transportation Letters*. doi: [10.1080/19427867.2020.1731993](https://doi.org/10.1080/19427867.2020.1731993).
- Basu, R. and J. Ferreira. 2020b. A LUTI microsimulation framework to evaluate the impact of autonomous mobility on the choice of housing-mobility bundles. *Environment and Planning B: Urban Analytics and City Science*. doi: [10.1177/2399808320925278](https://doi.org/10.1177/2399808320925278).
- Basu, R. and J. Ferreira. 2020c. Planning car-lite neighborhoods: Examining long-term impacts of accessibility boosts on vehicle ownership. *Transportation Research Part D: Transport and Environment*. doi: <https://doi.org/10.1016/j.trd.2020.102394>.
- Basu, R., A. Jana, and R. Bardhan. 2018b. A health care facility allocation model for expanding cities in developing nations: Strategizing urban health policy implementation. *Applied Spatial Analysis and Policy*, 11(1):21–36. doi: <https://doi.org/10.1007/s12061-016-9208-0>.
- Batty, M. 2007. Model cities. *Town Planning Review*, 78(2):125–151.
- Baum-Snow, N. 2007. Did highways cause suburbanization? *The Quarterly Journal of Economics*, 122(2):775–805.
- Brumbelow, K., J. Torres, S. Guikema, E. Bristow, and L. Kanta. 2007. Virtual cities for water distribution and infrastructure system research. In *World Environmental and Water Resources Congress 2007: Restoring our natural habitat*, pp. 1–7.
- Clarke, K. C., S. Hoppen, and L. Gaydos. 1997. A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and planning B: Planning and Design*, 24(2):247–261.
- Clay, M. J. 2010. Developing an integrated land-use/transportation model for small to medium-sized cities: Case study of Montgomery, Alabama. *Transportation Planning and Technology*, 33(8):679–693.
- de la Barra, T. 1989. Integrated transport and land use modeling: Decision chains and hierarchies.
- Derrible, S. and C. Kennedy. 2011. Applications of graph theory and network science to transit network design. *Transport Reviews*, 31(4):495–519.

- Dong, X., M. E. Ben-Akiva, J. L. Bowman, and J. L. Walker. 2006. Moving from trip-based to activity-based measures of accessibility. *Transportation Research Part A: Policy and Practice*, 40(2):163–180.
- Echenique, M. H., A. D. Flowerdew, J. D. Hunt, T. R. Mayo, I. J. Skidmore, and D. C. Simmonds. 1990. The MEPLAN models of Bilbao, Leeds and Dortmund. *Transport Reviews*, 10(4):309–322.
- Feldt, A. G. 1972. Clug; community land use game. player's manual with selected readings.
- Fortune, S. 1987. A sweepline algorithm for Voronoi diagrams. *Algorithmica*, 2(1-4):153.
- Hansen, W. G. 1959. How accessibility shapes land use. *Journal of the American Institute of Planners*, 25(2):73–76.
- Hawkins, J. and K. Nurul Habib. 2019. Integrated models of land use and transportation for the autonomous vehicle revolution. *Transport Reviews*, 39(1):66–83.
- Hunt, J. D. and J. E. Abraham. 2005. Design and implementation of PECAS: A generalised system for allocating economic production, exchange and consumption quantities. In *Integrated land-use and transportation models: Behavioural foundations*, pp. 253–273. Emerald Group Publishing Limited.
- Iacono, M., D. Levinson, and A. El-Geneidy. 2008. Models of transportation and land use change: A guide to the territory. *Journal of Planning Literature*, 22(4):323–340.
- Johnson, S. 2006. *The ghost map: The story of London's most terrifying epidemic—and how it changed science, cities, and the modern world*. Penguin.
- Knowles, R. D. 2012. Transit oriented development in Copenhagen, Denmark: From the finger plan to Ørestad. *Journal of Transport Geography*, 22:251–261.
- Lee, D. B. 1973. Requiem for large-scale models. *Journal of the American Institute of planners*, 39(3):163–178.
- Lerman, S. R. 1976. Location, housing, automobile ownership, and mode to work: A joint choice model. *Transportation Research Record*, 610:6–11.
- Lopes, A. S., C. F. G. Loureiro, and B. Van Wee. 2019. LUTI operational models review based on the proposition of an a priori ALUTI conceptual model. *Transport reviews*, 39(2):204–225.
- Lowry, I. S. 1964. A model of metropolis. Technical report, RAND CORP SANTA MONICA CALIF.
- Martinez, F. 1996. MUSSA: land use model for Santiago City. *Transportation Research Record*, 1552(1):126–134.
- McFadden, D. 1978. Modeling the choice of residential location. *Transportation Research Record*, (673).
- Miller, E. J. 2018. Integrated urban modeling: Past, present, and future. *Journal of Transport and Land Use*, 11(1):387–399. doi: <https://doi.org/10.5198/jtlu.2018.1273>.
- Oke, O., Y. M. Aboutaleb, A. P. Akkinpally, C. L. Azevedo, Y. Han, P. C. Zegras, J. Ferreira, and M. Ben-Akiva. 2019. A novel global urban typology framework for sustainable mobility futures. *Environmental Research Letters*.
- Papa, E. and L. Bertolini. 2015. Accessibility and transit-oriented development in European metropolitan areas. *Journal of Transport Geography*, 47:70–83.
- Putman, S. H. 1974. Preliminary results from an integrated transportation and land use models package. *Transportation*, 3(3):193–224.
- Salvini, P. and E. J. Miller. 2005. ILUTE: An operational prototype of a comprehensive microsimulation model of urban systems. *Networks and Spatial Economics*, 5(2):217–234.
- Semoloni, F., J. Assfalg, S. Armeni, R. Gianassi, and F. Marsoni. 2004. CityDev, an interactive multi-agents urban model on the web. *Computers, Environment and Urban Systems*, 28(1-2):45–64.

- Simmonds, D. C. 1999. The design of the DELTA land-use modelling package. *Environment and Planning B: Planning and Design*, 26(5):665–684.
- Strauch, D., R. Moeckel, M. Wegener, J. Gräfe, H. Mühlhans, G. Rindsfuser, and K.-J. Beckmann. 2005. Linking transport and land use planning: The microscopic dynamic simulation model ILUMASS. *Geodynamics*, pp. 295–311.
- Sun, J., X. Yu, G. Baciú, and M. Green. 2002. Template-based generation of road networks for virtual city modeling. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, pp. 33–40. ACM.
- Thiessen, A. H. 1911. Precipitation averages for large areas. *Monthly Weather Review*, 39(7):1082–1089.
- Thomson, R. C. and D. E. Richardson. 1995. A graph theory approach to road network generalisation. In *Proceeding of the 17th International Cartographic Conference*, pp. 1871–1880.
- Van Wee, B. 2015. Toward a new generation of land use transport interaction models. *Journal of Transport and Land Use*, 8(3).
- Vanegas, C. A., T. Kelly, B. Weber, J. Halatsch, D. G. Aliaga, and P. Müller. 2012. Procedural generation of parcels in urban modeling. In *Computer Graphics Forum*, volume 31, pp. 681–690. Wiley Online Library.
- Vitins, B. J. and K. W. Axhausen. 2010. Patterns and grammars for transport network generation. In *Proceedings of 14th Swiss Transportation Research Conference*.
- Waddell, P., A. Borning, M. Noth, N. Freier, M. Becke, and G. Ulfarsson. 2003. Microsimulation of urban development and location choices: Design and implementation of UrbanSim. *Networks and Spatial Economics*, 3(1):43–67.
- Warburton, S. 2009. Second Life in higher education: Assessing the potential for and the barriers to deploying virtual worlds in learning and teaching. *British Journal of Educational Technology*, 40(3):414–426.
- Weber, B., P. Müller, P. Wonka, and M. Gross. 2009. Interactive geometric simulation of 4d cities. In *Computer Graphics Forum*, volume 28, pp. 481–492. Wiley Online Library.
- Webster, F. V. and M. Dasgupta. 1991. Land use and transport interactions: Report of the ISGLUTI study. Technical Report CR295; ISSN: 0266-7045, Transport and Road Research Laboratory.
- Wegener, M. and K. Spiekermann. 2018. Multi-level urban models: Integration across space, time and policies. *Journal of Transport and Land Use*, 11(1). doi: <https://doi.org/10.5198/jtlu.2018.1185>.
- Zhou, E., A. McGlaughlin, and D. Turan. 2015. Generating synthetic road networks from various reduced dimension representations. URL http://snap.stanford.edu/class/cs224w-2015/Generating_Synthetic_Road_Networks_from_Various_Reduced_Deimension_Representations.pdf.
- Zhu, Y. and J. Ferreira. 2014. Synthetic population generation at disaggregated spatial scales for land use and transportation microsimulation. *Transportation Research Record*, 2429(1):168–177.