Modelling and Data Frameworks for Understanding Infrastructure Systems through a Systems-Of-Systems Lens

Matthew J. Berryman\textsuperscript{a}\textsuperscript{*}, Rohan Wickramasuriya\textsuperscript{a}, Vu Lam Cao\textsuperscript{a}, Qun Chen\textsuperscript{a}, Pascal Perez\textsuperscript{a}

Abstract: Modelling and analysis of large systems of infrastructure systems carries with it a number of challenges, in particular around the volume of data and the requisite complexity (and thus computing resources required) of models. In this paper we present an integrated land use–transportation model of a region in Sydney, and detail how we integrated an agent-based model of location and transport choice with a traffic micro-simulator. We also discuss both some novel architectures for scalability of modelling as well as for fusion and relevant visualisation of large data sets. We have a particular focus on geospatial infrastructure data visualisation.

Key words: Transportation; Land Use; Geospatial Visualisation; Modelling.

I. Introduction

Here we first introduce why we use agent-based modelling, then discuss the types of traffic simulation we considered, then review existing / incomplete modelling platforms for integrated land use – transportation planning, before moving on to a description of our model. There are two key systems we consider: land use (liveability and location choice), and the transportation system; in integrating these we provide a systems-of-systems perspective on the interplay between both.

A. Agent-based modelling

In agent-based modelling, an agent consists of:

1. A list of things that comprise its state. This might include things like age, income, a list of friends / family, etc.
2. Rules for updating the state of an agent, usually in relation to other agents and the agent's general environment.
3. Rules for updating the rules (agent learning / evolution).

Agent-based modelling allows one to effectively capture a very rich set of complex behaviours and interactions, and is therefore highly suited to modelling complex phenomena. It has gained extensive use in the fields of economics\textsuperscript{1}, social science\textsuperscript{2}, ecology\textsuperscript{3}, and biology\textsuperscript{4}, amongst many others.

\textsuperscript{a}SMART Infrastructure Facility, University of Wollongong, Wollongong NSW 2522, Australia
http://dx.doi.org/10.14453/isngi2013.proc_5
Since we are interested in modelling the heterogeneous nature of liveability perceptions and location choice across a population, agent-based modelling provides the best framework for considering this in different land use – transportation scenarios.

**B. Traffic simulation.**

Traffic simulation is a technique wherein vehicle movements through the traffic system are replicated by computer software. The state of the system is then estimated from the aggregated performance of all simulated vehicles. There are two main types of traffic simulation:

- In macro-simulation vehicles are represented as a traffic stream or platoon. The level of congestion on the road and the presence of traffic signals govern vehicle speed. Unlike macro-models, macro-simulation is able to deal with queues and intersection delays, making this better suited for analysis of congested urban streets and arterial roads.
- In microscopic simulation, individual vehicle units are traced through the traffic network. The movements of an individual vehicle are governed by how it interacts with vehicles in its proximity, for example car following models, lane change models, and gap acceptance models, providing granularity in vehicle-to-vehicle interactions not possible in macro simulation.

Since simulation models describe a dynamic process, they can be used to analyse a wide range of traffic problems when:

- Mathematical treatment is infeasible due to its spatial or temporal scale, or the complexity of traffic flow.
- Mathematical formulations represent dynamic traffic control environments as simple, quasi steady state systems.
- There is a need to view vehicle animation displays to gain an understanding of how the system is behaving in order to explain why the results were produced.
- Traffic performance needs to be understood prior to large investments in transportation infrastructure, e.g. what is the optimal geometric design of new entry/exit ramps.

There are two broad categories of traffic simulations:

- Highway: concerned with speed/flow/density relationships; lane utilisation; congestion (‘shockwave’ propagation); and
- Urban: with smaller areas involved and higher vehicle densities, these are more concerned with queue lengths, queue discharge speeds, travel times, delay times.

Given the above differences, and our focus on small urban regions, micro-simulation provides the best framework for our model, as well as providing an easy alignment between the agent-based model and micro-simulation, since there is a one-to-one mapping between individuals and cars in a household and entries in the TRANSIMS inputs, in addition to public transport vehicles.
C. Review of existing integrated land use – transportation models

Here we review a number of existing models / attempts at an integrated land use – transportation model, and where they exist we discuss how our model differs:

- UrbanSim\(^5\)
- SimTRAVEL\(^6\)
- ILUMASS\(^7\)
- ILUTE\(^8-10\)

Essentially though, the key novelty of our research is to combine an agent-based land use model with a micro-simulator model of traffic.

*UrbanSim* does not have a transport simulator component, rather it provides only a framework for coupling to a transportation model, and this has been done with transportation models like *EMME*\(^11\). *EMME*, however, is a macro-simulator, thus is not suited for our purposes.

It is not an agent-based model, so while it has descriptive power and can (within the usual constraints and limitations) make predictions, it has very little power to explain why different demographics act the way they do in particular scenarios (other than to re-iterate the generic features of behaviour that went into describing the rate equations). More recent coupling attempts include MATSim\(^12\).

In *SimTRAVEL*, access to transportation and results from the transport simulator (like congestion and travel time) are the only drivers of land use demand. Although it does have a traffic simulator, this is a mesoscopic simulator, Malta\(^13\), and therefore will not capture the dynamics at a sufficiently low level of scale to be able to assess different light rail alignments and their impact on traffic and land use patterns. Besides, SimTRAVEL uses a software, called PopGen, to create a synthetic population, based on an Iterative Proportional Updating (IPU) approach\(^14\). While some attempts have been made to adapt PoPGen to the Australian context, we found that its heavily US context-driven data management and approach were not optimal in a data-rich Australian context. We preferred to use a Synthetic Reconstruction approach\(^15-16\).

There are several projects that do combine, or did/will try to combine, an agent-based model with a micro-simulator, the *ILUMASS* project and the *ILUTE* project.

The *ILUMASS* project sought to bring together a land use model and a transportation mode, however unfortunately failed to get off the ground, due to having two sub-groups, one from a land use background, the other from a transportation background\(^7\). Our team had the advantage of a single team from a modelling background, drawing on expertise on both the land use and transportation backgrounds. The other reason that ILUMASS failed was in trading off scientific realism for modelling complexity, something that our team struggled with at times. In the end, the main one was around transportation complexity. Although micro-simulation is very detailed, that detail brings with it time complexity of the software (the time taken to simulate all of the traffic movements).

The *ILUTE* (Integrated Land Use and Transportation) project has been progressing well, but has not focussed on individual decision making around housing much, and in particular omits some of the key drivers of and a module around liveability, which is a key requirement
of our project and broader research.

D. Background to the SMART integrated land use – transportation model

The SMART Infrastructure Facility has developed an agent-based model of the South Randwick area that demonstrates the complexities of urban renewal and transit-orientated development. The agent-based modelling approach is adopted as it has the capacity of simulating the actions and interactions of autonomous agents (that is individuals or collective entities such as households or institutions) with a view to assessing their heterogeneous effects on the system as a whole. We have built a simulation model that can re-create in silico the observed complexity of urban systems, and generate—often unexpected—emerging patterns of social responses to changes in public policies or infrastructure assets. The model incorporates the following components: street network, public transport lines and timetables, traffic flow, land rates, population growth, individual travel routines, liveability factors, and the link between urban development and transportation options. The individually perceived liveability which forms the basis for the agents’ decision making of movement comprises of factors such as housing costs, population density, socio-cultural diversity, available amenities and transportation options. The modelled agents include individuals and households living in a given area, land-use and transport planners, land developers, and transport operators.

The model simulation starts by building a ‘realistic’ population of around 110,000 agents for the City of Randwick and Green Square precincts (baseline population). This population then evolves over a (typical) 20-year simulated period, with models built to simulate birth, death and marriage/de facto relationship formation and breakdown. We also consider changes to employment status and, importantly for our model, changes to the number of vehicles per household and changes to travel diaries, driven by this evolution. To be realistic, the baseline population adequately matches the distribution of individuals and households living in a given area as per the demographics information provided by the Australian Census.

Beyond its statistical validity, this artificial society also displays decisional and behavioural patterns, based on individual perceptions of liveability in consistency, with empirical evidence from a survey and the literature. Traditional transport models consider individual attributes such as age, education or income, as well as household-level characteristics to account for the diversity of transport. The novelty of the new SMART model lies in combing a travel model with a model of land use, in an agent-based framework, allowing agents to incorporate information about present and future availability of housing stock type, parking facilities, recreational amenities and shopping conveniences. This new flexibility allows for a large range of future scenarios to be tested, e.g. social responses to specific land-use or transport policies.

II. Liveability

Following Fernandez and colleagues (2005), we drive our residential mobility model through a dynamical model of perceived liveability. The conceptual structure of our liveability model is synthesized in the diagram below. From a subjective perspective on liveability, individuals tend to shape their preferences according to six factors describing various aspects of living conditions: (1) home, (2) neighbourhood, (3) transport, (4) entertainment, (5) services and (6) work. Each factor can be described through a series of attributes. The mix of attributes and their associated valence (e.g., an attribute can be perceived negatively or positively) depend on individual perceptions.
In order to implement a decisional process we propose to adapt the conceptual model proposed by Lindberg and colleagues for residential (re)location. The model assumes that a preference is established or a choice is made based on evaluations of the attribute level. For each factor, attributes are given even weights and they contribute equally to the overall valence of the factor. The factor level can be interpreted as a value/belief structure in which factors can be ranked and given different weights. According to Lindberg et al., for a given individual, factor ranking and attribute evaluation processes depend on his/her life cycle stage, current location and peer influence. This was confirmed through empirical work. Figure 1 shows the dimensions of liveability that our model considers. Our work on liveability provides details of liveability perceptions across demographic profiles, which is used as input to the agent-based model, in conjunction with data we have on the areas in which agents can live.

![Diagram of liveability dimensions](image)

**Figure 1:** The six categories of liveability, and the aspects thereof, that our model captures.

We have developed our model primarily to explore individual decision-making around liveability, and this depends on a good understanding of an individual’s transportation options in an area. For this reason we need a traffic micro-simulator, as we discuss in the next section.

### III. Integration

#### A. Architecture

Figure 2 shows the high level architecture. There are documents we also hold for some of the lower level components that we’ve written as well as some documents around
specification and testing. Note that there are a number of smaller Java libraries not shown. The architecture uses and has a loose coupling, with open standards where possible, between model components to allow for necessary evolution and future expansion of the software components.

The core of the model uses the Repast Simphony (sic) agent-based modelling platform, which was chosen as a good platform for building general-purpose agent-based models. We use it to load and schedule our agents (in random order, with a seed that can be fixed), as well as scheduling the running of TRANSIMS to simulate the transportation of agents. All of the above is packaged in a virtual machine (VM) running (Ubuntu) Linux, for the following reasons:

- There are a large number of different discrete software packages that need to be installed and maintained. Having them in a VM allows us to maintain a single, well-understood environment (such as the version of PostGIS), within which our software sits.
- Having a VM image allows easy deployment of a development environment to new developers.
- Having a VM image allows us to easily deploy to a private cloud for the purposes of Monte Carlo and/or multiple scenario simulation.

![Figure 2: Architecture, showing the end user, the private compute cloud running the model virtual machine(s) as well as a virtual machine running a temporary database for the configuration of scenarios and holding output data. Also shown is our data staging system for transforming this data along with other sources of data into processed data ready to be stored in a database and displayed on our dashboard web server. The CloudStack allows for simple web-based control of the model virtual machines, which run on top of a hypervisor (VMware ESX in our case, due to its support for large numbers of nodes).](image)

We use a central database server to store input data, and in particular configuration data for the different scenarios (including any random number seeds for the replicates), intermediate data that we want to keep a record of, and model outputs. Having a centralised database server aids us in quickly spooling up multiple instances of the model from the same
virtual machine image—each one in turn will load a different scenario as managed through the input database. The central database then captures all of the output data (allowing each VM instance to be small), ready for collection in our dashboard data visualisation and analytic system.

Within the model virtual machines, the Hibernate API effectively provides a layer of abstraction and better handling of the database queries of reading, manipulating and storing data. JDBC (and the accompanying PostgreSQL JDBC driver) provide the underlying database connectivity. The PostgreSQL database was selected due its support of SQL standards, ease of configuration, licensing (BSD-style), and importantly its support of geospatial information system (GIS) extensions, through PostGIS.

B. TRANSIMS

TRANSIMS was chosen as the traffic micro-simulator as, in its current iteration, it is a clean, efficient, C++-based (including good use of STL) platform that supports an individual (person and vehicle) level of modelling, and supports detailed microsimulation of traffic to support the requirements of our software, including but not limited to:

- road-by-road and minute-by-minute analysis of traffic patterns; and
- details of what individuals are going where on public transport, and analysis of usage (raw, and percentage utilisation).

We acknowledge the choice is not perfect, in particular we only use one run (with some limited annealing) of traffic simulation, as discussed below however we’re only interested in a fairly coarse level of behaviour, and other concerns such as support (via Argonne National Labs) and our familiarity with it were also factors. We have a set of scripts (bash, and batch when running on Windows) that drive the execution of TRANSIMS, and we pass data to and from it via the filesystem (sets of text-based input and output files, plus the network files derived from the shape files for the road and bus networks).

The integration has been conducted by suspending the rest of the agent-based simulation, while execution of TRANSIMS takes place. Normally one would use a process analogous to simulated annealing to arrive at the solution; running the router to establish initial routes, then finding when vehicles jam, and either redirecting them off the street temporarily into a park (if the numbers are sufficiently low) or by then re-routing them using the router and then running the simulation until numbers jammed are sufficiently low. Given the typical travel volumes (around 100,000 commuters), and our desire to simulate a 20-year period, we are forced to run only one typical weekday and weekend in simulation per year, and run only one iteration of the router. We have compared this with test runs of multiple iterations of router and the core microsimulator of vehicle movements, and found that travel times are within 5%; this we consider sufficient for our purposes.

Agents in our model have the following properties:

- Age
- Gender
- Weekly income
- Membership in a household (including single person households)
- Occupation
Highest education finished, and
Critically, agents have a travel diary, synthesised from household travel survey (HTS) data, and with work journeys also drawn from journey to work data.

The method we are proposing for such an assignment comprises two major stages. The first stage deterministically searches in HTS data for households that best match the household type, number of children under 15 years old, and number of adults of a synthetic population household. The deterministic search carried out in those steps gradually relaxes the constraints on exact matching of the number of residents and exact matching of the number of children younger than 15 years old so that the search always returns at least one HTS household.

The second stage randomly selects a HTS household from the list of households from stage one and assigns travel diary of individuals in this HTS household to those in the synthetic population household. The random selection in this stage follows a uniform distribution.

The travel diaries are selected from HTS data at household level. This is because we wish to reserve the inter-dependencies (in terms of the sequence, travel times and purpose) of daily trips of individuals in the same household.

TRANSIMS by itself is a tool to simulate the travel behaviour of each agent throughout an entire 24-hour period. Therefore, in order to integrate with TRANSIMS, a synthetic population with agents’ activities, which are their travel diary in another word, is essentially required.

In terms of travel diary, this is a set of all trips that an agent travels during a day. As a result, for each trip agent will need to provide for TRANSIMS their ID, the household ID that they belong to, the purposes of the trip (for instance, go to home, go to work, go to school, go shopping, go for social recreation or other purposes), the travel mode of the trip (for instance car, bus, train, bicycle, walk, or using carpool as a car passenger), the start time and expected arrival time of the trip, the origin and destination location of the trip. If agent travels by car, our software additionally provides to TRANSIMS which car in the house they are using (for instance the second car in the house) and where they park that car as well.

As a result, by using these input data and the road and transit network, TRANSIMS will simulate the travel behaviour of each agent throughout 24 hours of a day in order to conduct analyses of agents’ interactions so that agents are traced for every second of the day. Therefore, the location of each agent, car or transit vehicle is known for every second of the day as well as the traffic flow and congestion are provided.

Based on these output data, the Sydney model collects the travel time of each trip, using them to calculate the travel cost of the trip by using the current travel mode and other travel modes. Agents, based on these costs, will make their own decision about their travel mode for their trips in the next time step. Our model also utilises the congestion statistics from TRANSIMS output to calculate the satisfaction for agents to make a decision of relocation (staying or moving out the study area).

C. Output dashboard analytic and visualisation software

The dashboard system comprises:
• The centralised database that holds raw data outputs from the different model runs across the model VMs; and
• ETL (extract, transform, and load) functions on the output database server
• A server running the YellowFin data analysis and visualisation platform, including the web front end.

All of these are outlined further below.

We have a database server that holds a variety of infrastructure databases, processed model output databases, and a metadata database that holds all of the metadata associated with all of the other databases as well as descriptions, input data, and configuration details of all of the model runs.

Raw data from the model outputs is transformed using custom ETL functions, which as the name suggests is where the data is extracted from the raw data database, transformed into a form suitable for use in the dashboard. In particular this means use of star schemas for performance and other reasons as below, as well as performing any necessary statistical/summary transformations, which also leads to good performance of the dashboard front end. We use a schema following the star schema design method for the transformed data, for the following reasons:

• Star schemas are are simple to understand
• They provide excellent query performance. In particular, using star schemas avoids costly joins, sorting aggregation, etc.
• Hierarchies, levels and attributes exist in one place; and
• Conformed dimensions can be reused.

YellowFin is a business intelligence package that supports rapid development of graphs and other analytics around a set of defined outputs. One key aspect of YellowFin that makes it stand out from the rest of the business intelligence software packages is its excellent capability to handle geospatial datasets including vector (point, lines and polygons) and raster data. We benefit from this capability by using mainly YellowFin’s mapping to generate highly interactive map-based reports that are used as stand-alone reports or as embedded reports in dashboards. YellowFin supports Web Mapping Services (WMS) from OpenStreetMap tile servers to enrich those map-based reports. We use GeoServer to serve various ancillary spatial data as WMS layers that are then accessed by YellowFin map-based reports.

Different web dashboards, and reports (viewable from the web site, or sent via email) can be set up to allow point and click analysis of model data by end users. These can be re-used across different model scenarios. Figures 3 and 4 show a variety of outputs from our model, as displayed on our dashboard web site, with the output data from an early validation run of our model.
Figure 3. (left) The web-based dashboard for a road, showing traffic density along the route, by map and by road segment in northbound and southbound directions. (right) The web-based dashboard for a travel zone, showing a breakdown of household types, satisfaction with the area by age, and the distribution of dwelling capacities (1, 2, 3, and 4+ bedroom dwellings), for the area of Sydney shown in the map.

Figure 4. Summary statistics for a whole study zone for 2007, featuring (a) population density, (b) transport mode share across all purposes (mode share for particular purposes can be selected), (c) trips by time of day.
IV. Discussion and Conclusions

We have successfully integrated a traffic microsimulator (TRANSIMS) with an agent-based model of liveability and location choice, however some further work to do on validation is required. To date our model of population evolution, starting from 2006 ABS (Australian Bureau of Statistics) Census data, has been validated against 2011 ABS Census data, and traffic has been validated against both traffic count data and congestion data with some measure of success on main roads, though future validation will need to be done after some corrections on more local streets.

The integration of the YellowFin BI layer provides a useful and flexible means for visualising outputs from both the broader ABM as well as from TRANSIMS.

Our model successfully reproduces some key dynamics of the interplay between transportation and land use, and in future work we will present the completed validation and results from different scenarios.

Acknowledgements

We would like to acknowledge the contributions of Jie Yang, Johan Barthelemy, Jun Ma, and Nam Huynh.

References


