A methodology for simulating synthetic populations for the analysis of socio-technical infrastructures

Koen H. van Dam, Gonzalo Bustos-Turu and Nilay Shah

Centre for Process Systems Engineering, Imperial College London, UK
{k.van-dam, gb1612, n.shah}@imperial.ac.uk

Abstract: Modelling socio-technical systems in which a population of heterogeneous agents generates demand for infrastructure services requires a synthetic population of agents consistent with aggregate characteristics and distributions. A synthetic population can be created by generating individual agents with properties and rules based on a scenario definition. Simulation results fine-tune this process by comparing system level behaviour with external data, after which the emergent behaviour can be used for analysis and optimisation of planning and operation. An example of electricity demand profiles is used to illustrate the approach.

Keywords: agent-based model, socio-technical system, synthetic population

1 Introduction

To analyse and understand the operation of socio-technical systems, in which physical systems interact with social networks, simulation models need to include the behaviour of the actors in the model [1]. For example, the emergent behaviour of a group of actors, modelled as agents, could generate the demand for services provided by the socio-technical system. Individual characteristics of agents are necessary to include heterogeneity in the simulation model while remaining consistent with aggregate/average values for the population [2][3]. In this paper a methodology to generate a synthetic population given certain land-use and general population characteristics is proposed. The approach followed uses geographical information system (GIS) data as input which is enriched with land-use data (e.g. population density, floor space of offices) combined with statistics on the population (e.g. employment rates, car ownership). A case study on electricity consumption is used as an example of the proposed methodology and some illustrative results are presented for an area in West London.

2 Methodology

The core idea is that a synthetic population can be generated during the initialisation of a simulation model using a few basic properties for the area considered, key characteristics for the population and distributions of the activities they engage in which

adfa, p. 1, 2011.

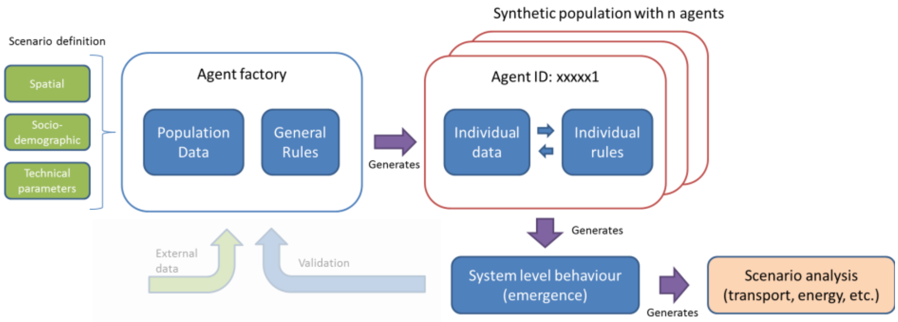

Fig. 1. Methodology of generating a synthetic population based on a scenario definition

together form the scenario definition. An "agent factory" then generates a population of agents to make up the synthetic population, after which *system level behaviour* can be *analysed* or used for *validation* of the generation step. Fig 1 shows a schematic for this workflow from scenario definition to output analysis. The population data and general rules are characterised by probability distributions (e.g. uniform or normal), while the individual data and individual rules at the agent level represent a specific instance (e.g. a chosen departure time). The input of the agent factory includes spatial (i.e. built environment) data, socio-demographic data and technical parameters which can be provided through input files (e.g. shape files, structured documents, or as variable values in the model definition). Examples of typical inputs are given in Table 1.

Table 1. Examples of scenario definition (with some sub properties shown in brackets)

Input category	**General**	**Scenario specific**
Built environment (spatial data)	Roads (type, speed), buildings or areas (land-use, number of floors, total footprint)	Electricity distribution network (capacity and layout)
Socio-demographic	Population density, household size, car ownership level, employment rate, activity schedule (activity, departure, deviation, occurrence)	Electric vehicle type market share, PEV adoption level
Technical parameters	Peak and base electricity demand per household, average speed, etc.	Charging infrastructure access level, charging power rates

From this input data, the agent factory determines how many agents to create and where their activity locations are. The number of agents living in a certain location depends on the population density. Agents may be created in different groups (e.g. workers or non-workers) again taking socio-demographic data into account. The total floor space, based on footprint and number of floors, combined with the land-use (e.g. leisure or commercial) of a building or neighbourhood affects the probability that an agent chooses it as a destination for its activities.

In this work, the general activity schedule AS_k for each group k of agents is defined with a list of 4-tuples:

$$AS_k = \{(ACT_j, MDT_j, SD_j, PD_j)\} \qquad \text{(Eq.1)}$$

For each activity ACT_j, the departure times are modelled as a stochastic variable following a normal distribution with a mean departure time MDT_j representing the peak hour of that period and considering a standard deviation SD_j to account for variability in the departure time among agents. A departure probability PD_j is included in the model to simulate that not every agent in that group undertakes all the activities.

Three types of analysis can be done then by making changes to the scenario input:

- Changes in land use or spatial planning can be explored by adjusting the spatial configuration input files. This can be used to explore the impact of different masterplans and proposed developments on the same population.
- Different population characteristics (i.e. socio-demographic data) can be provided to see how they impact the way people use the available city infrastructures. This way one could experiment how changes in the population living in an area could be reflected on the demand for infrastructure services.
- The user can experiment with the impact of different behavioural rules. This could be used to test the consequence of incentives or policies that change how people travel, work and engage in activities, by changing the occurrence and the timing.

In next section an example is given following this methodology, including case-specific data and behavioural rules for the agents.

3 Case study: simulating the electricity consumption in an urban neighbourhood, including electric vehicles

The example application used in this paper is the charging of plug-in electric vehicles (PEVs) in an urban area, following the description in [4]. Driving reduces the state of charge (SOC) of the battery which is recharged when the car is plugged into a charging unit. Simulation of the spatial and temporal distribution of the heterogeneous PEV owner activities generates PEV journeys and individual demand for charging, giving detailed insights about where and when there is the potential to recharge the PEV batteries. To assess the impact of this "mobile" load, local distribution network conditions (including local demand for electricity from "static" loads leading to constraints on the distribution network) have to be determined. The spatially- and temporally-explicit static load is also simulated using the area’s occupancy that is estimated using the same population and their activities and transport demand, leading to electricity consumption profiles in the different areas of the simulated city.

For this example, the number of PEVs simulated for each area is calculated based on the number of cars (based on the population size, household size and car ownership levels per area[1]) and level of PEV diffusion. Finally, the population for each area is based on the density[1] and footprint area[2] (see Fig. 2). A combination of general and scenario specific data is thus used (see Table 1) to set up the scenario.

---

[1] extracted from Office for National Statistics data (http://neighbourhood.statistics.gov.uk/)
[2] from OrdnanceSurvey MasterMap data (https://www.ordnancesurvey.co.uk/)

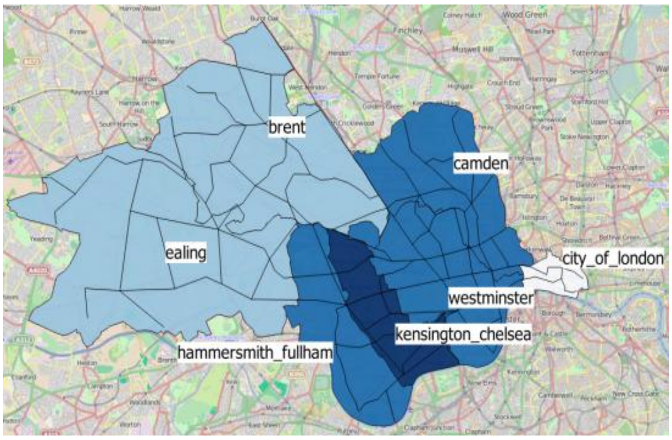

Fig. 2. City layout used for the simulation. Colours represent the density for each borough.

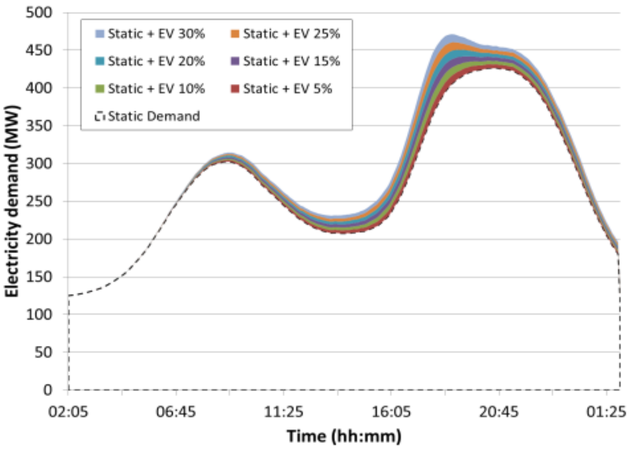

Fig. 3. Output of simulated electricity demand for 5%-30% PEV adoption.

During the initialisation of the model the agents are created based on the number of PEVs in each area. Then, each agent is linked to a home, office and typical place for shopping and leisure activities (based on the land-use) and its charging infrastructure access level is defined. Next, this agent is linked with a PEV with an initial SOC. Finally, the activity schedule for each agent is created based on the general activity schedule of the agent's group (see Eq. 1). If desired this process can be deterministic to enable replication of scenarios or stochastic to explore parameter space.

Once the population is created the simulation is run for a working day and the residential and PEV electricity demands are generated. Different scenarios of PEV adoption (from 5% to 30%) are considered to analyse the impact of PEV on the distribution network in different areas of the city (see Fig. 3). Although the simulation generates individual agent data, the outputs can be aggregated spatially, or according to different group attributes of agents (workers/non-workers, etc.).

4 Conclusions

Social simulation is essential for realistic analysis of urban infrastructure systems in which human activities drive demand [5]. If these demands are not part of the simulation but provided as input (e.g. average values obtained from surveys) it means the model cannot provide decision support for the impact of incentives and behavioural change, or explore how such demands may change in the future. By including the actors and their behaviour, the bottom-up generation of demands means feedback loops can be explored by seeing what the response of the population is on infrastructure changes. This way the social, the physical or both systems can be varied to study the impact on overall system behaviour. For the case presented this means that one can to experiment with detailed socio-economic data, looking at the impact of social factors on PEV adoption levels or smart charging operation. Separating generation of the synthetic population from the rest of the simulation means this model component is re-usable and flexible in multiple case studies, but also makes it transparent how the agents are generated based on selected input data.

References

[1] Van Dam, K.H., Nikolic, I. & Lukszo, Z. (2013) Agent-based modelling of socio-technical systems, Springer, ISBN 978-94-007-4932-0

[2] Harland, K., Heppenstall, A., Smith, D. & Birkin, M. (2012) Creating Realistic Synthetic Populations at Varying Spatial Scales: A Comparative Critique of Population Synthesis Techniques in Journal of Artificial Societies and Social Simulation 15 (1) 1

[3] Moeckel, R., Spiekermann, K., & Wegener, M. (2003) Creating a synthetic population. Proceedings of International Conference on Computers in Urban Planning and Urban Management (CUPUM), Sendai, Japan, May 2003

[4] Bustos-Turu, G., van Dam, K.H., Acha, S. & Shah, N. (2014) Estimating plug-in electric vehicle demand flexibility through an agent-based simulation model. 5th IEEE PES Innovative Smart Grid Technologies (ISGT), 12-15 October, 2014, Istanbul, Turkey

[5] Sovacool, B.K., Ryan, S.E., Stern, P.C., Janda, K., Rochlin, G., Spreng, D., Pasqualetti, M.J., Wilhite, H. & Lutzenhiser, L. (2015) Integrating social science in energy research, Energy Research & Social Science, Vol 6, Mar 2015, pp 95-99