The DIM2SEA Agent-Based Modeling Framework

A. Yair Grinberger, Daniel Felsenstein, Peleg Samuels

WP 10/17

Increasing Urban Resilience to Large Scale Disasters: The Development of a Dynamic Integrated Model for Disaster Management and Socio-Economic Analysis (DIM2SEA)

funded by JAPAN Science and Technology Agency (JST) and Ministry of Science, Technology and Space, Israel (MOST)

1 Introduction

This paper presents the conceptual and structural framework behind the simulation tool developed within the DIM2SEA project. The simulation tool builds upon previous work (. Grinberger & Felsenstein, 2014, 2016; Grinberger et al., 2015, 2017a) and articulates a comprehensive agent-based model (ABM) which estimates the long-term consequences of a large-scale urban disaster.

Previous treatment of urban disasters in the agent-based (AB) literature have considered issues such as flooding (Dawson et al., 2011), fires (Chen & Zhan, 2008), earthquakes (Crooks & Wise, 2013), terrorism (Park et al., 2012), and industrial accidents (Salze et al., 2014). Most of these studies have devoted their attention to short-term evacuation and recovery processes (Zimmerman et al., 2010; Chen et al., 2012). While these studies investigate important aspects of disaster management and mitigation, the shocks caused to an urban system by a disaster often lead to the emergence of long-term effects, which can have a significant impact on residents and the urban system. In contrast to the short-term view, a long-term perspective on this issue cannot consider only the physical and material dimensions of recovery, vulnerability, and resilience. I t needs to adopt a more comprehensive approach to the simulation of the urban system. Outside the disaster-oriented literature, ABMs have been shown to be useful in integrating different tangible and intangible aspects of a computable system, such as supply and demand, thus representing full working of markets and other equilibria generating systems (Hepenstall et al., 2005; Ettema, 2011; Magliocca et al., 2011, 2015; Filatova, 2015: Olner et al., 2015).

Every ABM naturally relies on the identification and formalization of ‘agents.' Within ABM’s, agent’s behavior can be simulated with great sophistication and complexity. Transitioning from abstract and theoretical environments to real-life systems calls for representation of real world populations. While the individual-level data needed for realistically characterizing agents is rarely accessible, it can be generated synthetically. The literature offers several approaches including population gridding (Linard et al., 2011), areal interpolation (Reibel & Bufalino, 2005), and dasymetric representation (Eicher & Brewer, 2001; Mennis 2003). The approach we use relies on distributing data discretely, calling for combining different sources structured and non-structured data. Such approach has been shown to be appropriate for integration into a simulation framework (Grinberger et al., 2015, 2017a; Zhu & Ferreira, 2015).

In this paper, we offer three new contributions to a previously developed modeling framework that adheres to the above outline. First, we add the consideration of joint distributions of variables to the algorithm that generates the synthetic population (i.e. allocation algorithm). Additionally, this algorithm is enhanced by a theoretically-based spatial allocation approach. Second, we integrate a labor market sub-model into the simulation framework incorporating both the supply and demand side of the market. Finally, we migrate both the allocation algorithm and the model into a Matlab computing environment. This offers many advantages for the current research. The next section introduces the general modeling framework. It is followed by three sections, each discussing one of the contributions outlined above.

2 Modeling Framework

Following the discussion above, the modeling framework developed within the DIM2SEA includes three stages (see Fig. 1):

- *Data disaggregation and socio-economic profiling*. This includes the implementation of a disaggregation algorithm that creates a synthetic representation of each individual and household within the database, based on aggregate census area and population counts. Households are first characterized in terms of size and age distribution, and then individuals are created and characterized along several socio-economic dimensions. The allocation is iterative in nature; the allocation of one characteristic, as related to other attributes, does not begin until the previous category has been completely allocated (meaning the control total for a certain category, derived from population counts data, is achieved). Upon the completion of the socio-economic allocation procedure, individuals and households are allocated to specific locations, using a spatial database of buildings and assets.
- *Agent-based simulation modeling.* This includes the development of an ABM which accounts for the different sub-systems that act within an urban area. Individuals and households are defined as the 'agents' – the atomic units of the system. In accordance, the synthetic database is imported into the model and transformed into a dataset containing all agents. Agents behavior is defined along three dimensions: residential location, workplace location, and daily activities. In each dimension, the choice of movement destinations (whether it be relocation or daily mobility) is probabilistically set in accordance with the agent's preference structure and constraints (e.g. social tolerance in residential environments and income for residential location decisions). The environment is also modeled as a dynamic entity, based on the real-life information. Data regarding the location and characteristics of census tracts, buildings, and assets are imported into the model and transformed into various datasets. A trickle-down process is defined along the three dimensions of the residential market, labor market, and land-use system, where the total of agents behavior alters the characteristics of census tracts, which affects building-level and asset-level attributes. In such a way, changes to the land-use system, demand for labor, and housing values are simulated.
- *Generating a multi-dimensional database.* The dynamics of the simulation model are recorded via changes to the distribution and characteristics of individuals, wages, and economic activity. This results in a multi-dimensional database that accurately describes the spatio-temporal dynamics of multiple urban sub-systems. As such, it is the main output of the framework, promoting both analysis and dynamic visualization of processes of change in the aftermath of a disaster.

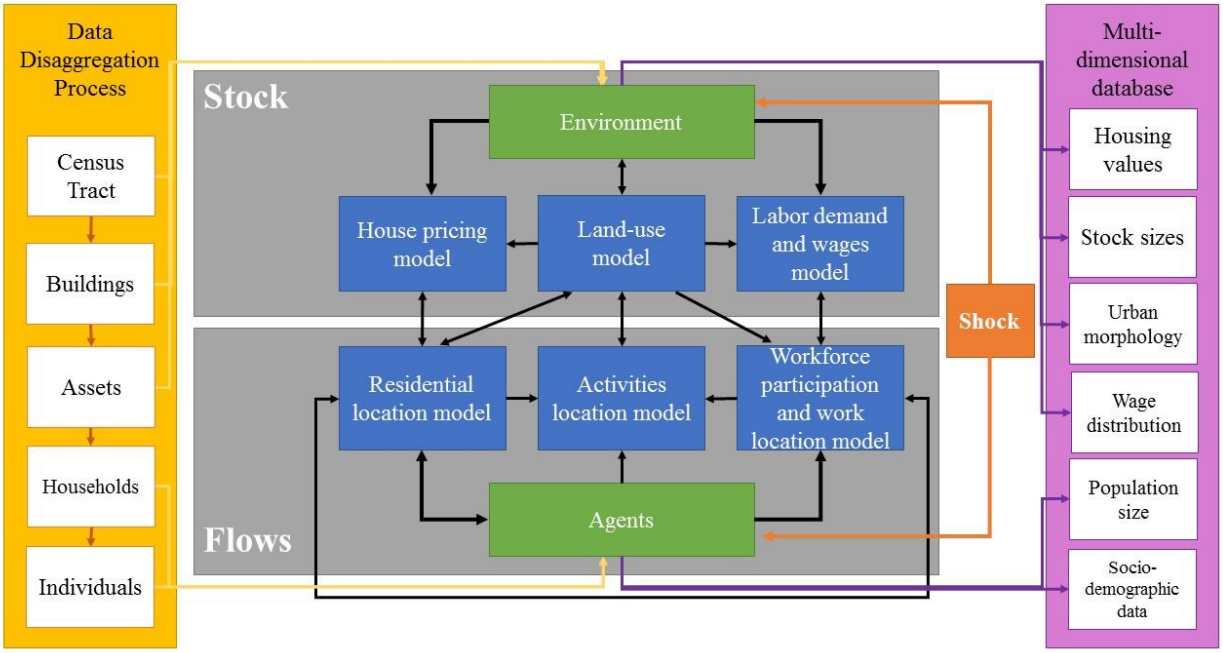

Figure 1. Modeling framework.

3 Allocation Algorithm: Joint Distributions and Spatial Allocation

Previous implementations of the allocation algorithm (see Grinberger et al., 2015, 2017a) have generated synthetic populations from census-tract-level data via an IPF (iterative proportional fitting) procedure, where categorical attributes are allocated to individuals according to population counts, one category at a time. While this procedure considers the correlation between different attributes (such as education and income) within the allocation, it does not capture entirely the joint distributions which these correlations produce. This is evident when the resulting joint distributions in the synthetic data are inspected. We observe that not only do the joint distributions not follow any intuitive form, in many cases they seem to not exist at all, essentially creating independent variables, from phenomena that are clearly intertwined. Figure 2 shows, for illustration, the resulting correlation between the number of cars owned by a household and its earnings. Estimated correlation coefficients for the study area are extremely low, and their distribution is centered on zero.

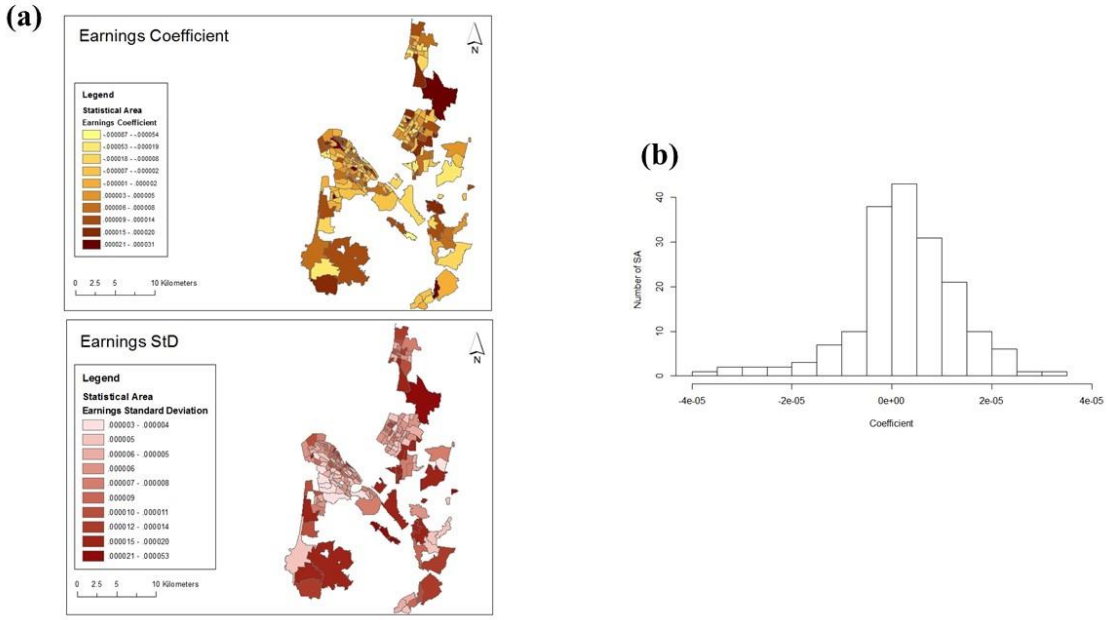

Figure 2. (a) resulting coefficients for the correlation between households' earnings and the number of cars by statistical area, for the project study area, (b) the overall distribution of the coefficients.

Consequently, a revision of the procedure is required in order to account for interrelationships. Accordingly, a more probabilistic approach has been defined. This approach requires formulating assumptions regarding how two (or more) variables are jointly distributed prior to the allocation. Realizing the number of assumptions required, we omit several the variables from the analysis, retaining only the most essential attributes. Relying on the functional forms derived from these assumptions, probability functions for individual *x* representing traits *y* are formulated. The values of coefficients of these relationships are derived via statistical estimation using census-tract level data. For example, the probability of a household owning a car is related to income and household size, and the coefficients for each of these variables are derived using elasticities computed from Berry's (1994) market share model. Each attribute is allocated in such a way to individuals until the sum-total derived from population counts is achieved. Figure 3 presents these characteristics and the relations between them.

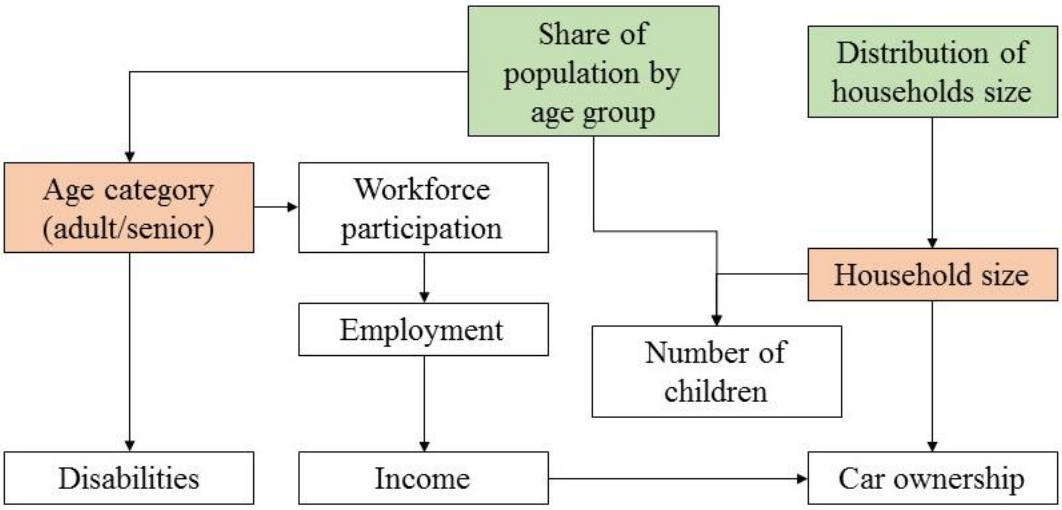

Figure 3. Revised conceptual framework for allocation algorithm. Green – initial data, orange – key variables.

Another issue that has been reconsidered relates to the spatial allocation of individuals and households to assets and buildings. In previous work, this was achieved by computing income-based and rent-based scores for each household and asset, in accordance, and matching households and assets based on these scores. This ad hoc approach has no real theoretical foundation. To rectify this situation a new allocation procedure has been developed grounded in the reverse of the Alonso (1960) bid-rent curve model. The original model derives a land use price structure based on the revealed preference structure of households as expressed in their propensity to trade-off space and accessibility. Our data includes information on price structure and node location, and thus allows us to derive estimates regarding local preferences, as described by a parameter capturing the substitution of location and area.. The substitution parameter replaces the scores computed in the previous version of the spatial matching algorithm. For more details, see Samuels & Felsenstein (2017).

4 Integrating the Labor Market into the AB-Model

Previous modeling efforts have considered the dynamics of the land-use system, human mobilities, and housing market following a disaster. Treatment of the labor market has been only partial. Its dynamics affect the entire system: by setting income levels and rents, as well as attracting or repelling populations and firms, it serves as a spatially- and functionally-organizing element of the system (Lowry, 1964; Scott, 1988).

Following the existing formulation of the modeling framework, the addition of the labor market component required the simulation of both stock-level (demand for labor) and individual-level (supply) dynamics. The algorithms for this are based on a theoretical model that relates to worker productivity, capital, and costs on the firm side, and utility on the worker side. This is developed into a set of conditions which determine the change in labor demand and wage levels in accordance with changes in capital stock and previous wage levels. These conditions allow

for the integration of the stock-level component of the labor market into the framework (see Fig. 1). The integration of individual-level behavior is based on the translation of utility functions into rules guiding agent activity. In the case of the labor market, an agent determines its workplace in accordance with wage levels (in relation to a minimum wage level) and distance from home (see Grinberger et al., 2017b). The addition of labor market dynamics into the modeling framework is expected to enrich the outputs by exposing effects related to wage levels, employment levels, recovery of different sectors, and welfare effects.

5 Matlab as an AB Modeling Environment

The final development refers to the computing environment. The previous iterations of the framework relied on the Repast Simphony 2.0 development environment. This is a popular open source platform for AB programming, which also offers a solid spatial element in the form of different spatial contexts, including a basic visualization interface. However, this environment presents several constraints for the current application. Previous models using Repast simulated the behavior of about 20,000 different agents acting within a relatively small inner-city area. The computing times required for such simulations were relatively long, lasting more than 6 hours for a 1000 iterations-long simulation (each iteration representing a day) on a relatively capable computer. The current study has greater demands on computing resources and graphical representation. The population of agents is likely to increase by more than an order of magnitude and behavior within multiple municipalities will need to be simulated. Repast does offer a solution to this in the form of the RepastHPC platform (a high-performance computing platform), but this does not incorporate a GIS component. The second constraint relates to visualization capabilities. The graphic interface included within the Repast suite offers basic functionalities such as symbology settings. Yet, more advanced functionalities, such as 3D visualization essential for disseminating the results, is not included and cannot be integrated with ease.

The solution to these constraints lies in transitioning to another development platform. While this requires much work in translating the model, the advantages make the effort worthwhile. Also, since the allocation procedure will also undergo major revision, the transition creates an opportunity to import it into the new environment thus directly integrating it into the model building stage. Previously the allocation procedure was coded separately from the model in SQL Server. Matlab is a general modeling tool that has been used in for the purpose of building agent-based models (Nikolai & Madey, 2009). It is particularly popular in the field of ecology (where ABMs are often called individual-based models; Lorek & Sonnenschein, 1999; Rashleigh & Grossman, 2005; Beekman & Lew, 2008; Kool, 2009). The advantages of Matlab lie in its high-performance computing abilities, its inclusion of spatial tools and models and its incorporation of models for volume visualization. As such, it provides for the needs of the current implementation. One drawback is its proprietary nature which is contrary to the open source environment of the Repast suite.

6 Conclusions

This paper has reviewed the latest developments in the DIM2SEA simulation framework. Three main developments are noted: the revision of the allocation algorithm to control for joint distributions of socio-economic variables and for the spatial patterns of residential preferences; the addition of two labor-market-related sub-models to the ABM and the transition of both the allocation algorithm and the ABM to the Matlab platform. These three developments present the latest stage in the evolution of the simulation model. We feel that this furthers the DIM2SEA agenda of developing a large-scale simulation framework for the long-term effects of a disaster on a metropolitan area. With a more accurate initial representation of the agents and environment within the model and the addition of labor market representation, we expect to produce better and richer simulations. The transition to a new development environment lays the technical and technological base for simulating at a large scale with relatively short computing times.

7 References

Alonso, W. (1960). A theory of urban land market. *Papers and Proceedings of the Regional Science Association, 6*(1), 149-157.

Beekman, M., & Lew, J.B. (2008). Foraging in honeybees – when does it pay to dance? *Behavioral Ecology, 19*(2), 255-261.

Berry, S.T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics, 25*(2), 242-262.

Chen, X., & Zhan, F.B. (2008). Agent-based modeling and simulation of urban evacuation: Relative effectiveness of simultaneous and staged evacuation strategies. *Journal of the Operational Research Society, 59*(1), 25-33.

Chen, X., Kwan, M.-P., Qiang, L., & Chen, J. (2012). A model for evacuation risk assessment with consideration of pre- and post-disaster factors. *Computers, Environment and Urban Systems, 36*(3), 207-217.

Crooks, A.T., & Wise, S. (2013). GIS and agent based models for humanitarian assistance. *Computers, Environment and Urban Systems, 41*, 100-111.

Dawson, R.J., Peppe, R., & Wang, M. (2011). An agent-based model for risk-based flood incident management. *Natural Hazards, 59*(1), 167-189.

Eicher, C.L., & Brewer, C.A. (2001). Dasymetric mapping and areal interpolation: Implementation and evaluation. *Cartography and Geographic Information Science, 28*(2), 125-138.

Ettema, D. (2011). A multi-agent model of urban processes: Modeling relocation processes and price setting in housing markets. *Computers, Environment and Urban Systems, 35*(1), 1-11.

Filatove, T. (2015). Empirical agent-based land market: Integrating adaptive economic behavior in urban land-use models. *Computers, Environment and Urban Systems, 54*, 397-413.

Grinberger, A.Y., & Felsenstein, D. (2014). Bouncing-back or bouncing forward? Simulating urban resilience. *Proceedings of the Institution of Civil Engineers: Urban Design and Planning, 167*(3), 115-124.

Grinberger, A.Y., & Felsenstein, D. (2016). Dynamic agent based simulation of welfare effects of urban disasters. *Computers, Environment and Urban Systems, 59*, 129-141.

Grinberger, A.Y., Lichter, M., & Felsenstein, D. (2015). Simulating urban resilience: Disasters, dynamics and (synthetic) data. In S. Geertamn, J. Ferreira Jr., R. Goodspeed & J. Stillwell (Eds.), *Planning Support Systems and Smart Cities* (pp. 99-119). Springer.

Grinberger, A.Y., Lichter, M., & Felsenstein, D. (2017a). Dynamic agent based simulation of an urban disaster using synthetic big data. In: P. Thakuriah, N. Tilahun & M. Zellner (Eds.), *Seeing Cities Through Big Data* (pp. 349-382). Springer: Dordrecht.

Grinberger, A.Y., Samuels, P., & Felsenstein, D. (2017b). Integrating labor market dynamics into the DIM2SEA agent-based model: A framework for analysis. *DIM2SEA Project, Working Paper No. 9.*

Hepenstall, A.J., Evans, A.J., & Birkin, M.H. (2005). A hybrid multi-agent/spatial interaction model system for petrol price setting. *Transactions in GIS, 9*(1), 25-51.

Kool, J.T. (2009). An object-oriented, individual-based approach for simulating the dynamics of genes in subdivided populations. *Ecological Informatics, 4*(3), 136-146.

Linard, C., Gilbert, M., Tatem, A.J. (2011). Assessing the use of global land cover data for guiding large area population distribution modeling. *GeoJournal, 76*(5), 525-538.

Lorek, H., & Sonnenschein, M. (1999). Modelling and simulation software to support individual-based ecological modelling. *Ecological Modelling, 115*(2), 199-216.

Lowry, I.S. (1964). *A Model of Metropolis*. Santa Monica, CA: Rand Corporation.

Magliocca, N., McConnell, V., & Walls, M. (2015). Exploring sprawl: Results from an economic agent-based model of land and housing markets. *Ecological Economics, 113*, 114-125.

Magliocca, N., Safirova, E., McConnell, V., & Walls, M. (2011). An economic agent-based model of coupled housing and land markets (CHALMS). *Computers, Environment and Urban Systems, 35*(3), 183-191.

Mennis, J. (2003). Generating surface model of population using dasymetric mapping. *Professional Geographer, 55*(1), 31-42.

Nikolai, C., & Madey, G. (2009). Tools of the trade: A survey of various agent based modeling platforms. *Journal of Artificial Societies and Social Simulation, 12*(2), 2.

Olner, D., Evans, A., & Heppenstall, A. (2015). An agent based model of urban economics: Digging into emergence. *Computers, Environment and Urban Systems, 54*, 414-427.

Park, A.J., Tsang, H.H., Sun, M., Glasser, U. (2012). An agent-based model and computational framework for conuter-terrorism and public safety based on swarm intelligence. *Security Informatics, 1*(23). http://dx.doi.org/10.1186/2190-8532-1-23.

Rashleigh, B., & Grossman, G.D. (2005). An individual-based simulation model for mottled sculpin (*Cottus bairdi*) in a southern Appalachian Stream. *Ecological modelling, 187*(2), 247-258.

Reibel, M., & Bufalino, M.E. (2005). Street-weighted interpolation techniques for demographic count estimation in incompatible zone systems. *Environment and Planning A, 37*, 127-139.

Saltze, P., Beck, E., Douvinet, J., Amalric, M., Bonnet, E., Daudé, E., … Sheeren, S. (2014). OXI-CITY: an agent-based model for exploring the effects of risk awareness and spatial configuration on the survival rate in the case of industrial accidents. *Cybergeo-European Journal of Geography, 692*. https://cybeergeo.revues.org/26522

Scott, A.J. (1988). *Metropolis: from the division of labor to urban form*. Berkley, CA: University of California Press.

Samuels, P. & Felsenstein, D. (2017). *Deriving Spatial Preferences from Rents: The Alonso Model in Reverse*. DIM2SEA Project, Working Paper No.XX .

Zhu, Y., Ferreira J. Jr. (2015). Data integration to create large-scale spatially detailed synthetic populations. In: S. Geertman, J. Ferreira Jr., R. Goodspeed & J. Stillwell (Eds.) *Planning Support Systems and Smart Cities* (pp. 121-14). Springer: Dordrecht.

Zimmerman, B., Nawn, D., Wang, Y., Kuhlman, B., Sochats, K., Luangesorn, L., … Bidanda, B. (2010). *Dynamic Model Generation for Agent-Based Emergency Response Simulation. Paper presented at ESRI International User Conference.* California: San Diego (Retrieved from) http://proceedings.esri.com/library/userconf/proc10/uc/papers/pap_1078.pdf.