Developing A Spatio-Temporal Ambient Population Estimation Model using Epidemics-based Geosimulation

Mehdi Khakpour
Department of Geography, Norwegian University of Science and Technology
Dragvoll, 7491
Trondheim, Norway
Mehdik@svt.ntnu.no

Jan Ketil Rød
Department of Geography, Norwegian University of Science and Technology
Dragvoll, 7491
Trondheim, Norway
Jan.rod@svt.ntnu.no

Abstract

While estimating the population-at-risk is believed to be a crucial part of any disaster management planning process, there are very few modelling structures, by which we can generate controlled estimates of the number of people at risk in the wake or before a major hazard. In this research, a spatio-temporal ambient population model is developed on an epidemics-based geosimulation platform. The methodology used in this paper is based on building a geosimulation structure of urban movements that follows the analogous behaviours of the spread of invasive diseases in an assumedly homogeneous population. An attraction-based cellular automaton model is the driver of the geosimulation allowing different land use categories to obtain different estimations of losing or gaining the population in different times of the day.

It is inferred from this research approach that it is possible to develop epidemics-based geosimulations of people's movements through out a random working day in a very fine resolution. This model can be used to generate synthetic ambient population maps of urban areas, which form a type of demographic information that could be used in several urban planning as well as disaster management assessments by studying the urban movement patterns or making controlled estimates of the number of people at risk.

Keywords: Geosimulation, Epidemics, Cellular Automata, Population-at-risk, Synthetic Ambient Population Maps.

1 Introduction

Estimating the population-at-risk is believed to be a critical factor of any risk assessment process as part of the disaster preparedness planning as well as post-disaster response operations. While the highest priority in any disaster management structure is given to the human lives, the size of any formal evacuation planning, search and rescue operation, recovery and emergency housing, etc. must necessarily be incorporated to a fairly reliable estimate of the number of people being exposed to the considered hazard at the prospective location and time, often referred to as the ´Ambient Population´[3]. Obviously, the efficiency of such planning is, to a large extent, a matter of the accuracy as well as the spatio-temporal resolution of the available population data. For example, knowing how many people are at risk of a possible major hazard such as a hurricane can clearly determine what capacity of evacuation routes or how many logistical vehicles are needed or how much food or medicines must be stored when planning for an emergency response operation.

A fairly recent approach within urban planning has risen on the basis of looking at the cities and their services as complex organisms [1,7] . This analogy enables the urban planners to develop models in order to understand the dynamics of different functions in a city and to study the disruptions of systems and services and improve their resiliency [2].

The term ´Epidemics-based Geosimulation´ is introduced in this article in order to refer to a particular type of geosimulation, in which the spatially explicit elements of the simulated phenomenon are assumed to spread through time and space, following stochastic epidemic modelling approaches. In this particular sense, cellular automata and agent-based models turn into a discrete modelling environment of autonomous geographical elements (e.g. gridded land cells) or agents (e.g. people). The behaviour (or the state) of the system is explained using an analogy to the spread of invasive diseases in an assumedly homogeneous population. In this context, the cells or agents are still autonomously moving or changing status and their macroscopic behaviour could be emergent, but their decision making process is based on a set of epidemiologic deterministic laws that involve some level of stochasticity.

In this research, fine scale synthetic population maps are generated considering the city as a complex system using an attraction-based cellular automaton. The Ambient Population Estimation model is built on an analogy to formal SIS (Susceptible-Infected-Susceptible) epidemic models with the difference that we assume every grid-cell of the urban area changes its states in a Stable-Dynamic-Stable manner. This means that the population movements and transition between the cells are considered analogous to the invasive behaviour of some form of disease propagation that has neither any recovery nor any death (i.e. common flu).

2 Methodology

Fine scale synthetic population maps are generated based on a general assumption that the movements in an urban area can be modelled using a similar modelling structure to the classical SIS models. So if the urban area would assumedly be divided into unilateral grid-cells, each representing a population of individuals, every piece of land can represent either a susceptible or infectious individual that can change its status due to it's contact with the neighbouring cells. When people travel between the grid cells, every grid cell is assigned by an estimated number of people staying in that certain piece of land in a certain time of the day. Then in different times of the day, some areas in an urban area are more likely to attract or repel the population, while others are more likely to keep equilibrium between the number of people travelling to and from these places. The infectious state is assumed to be similar to a condition, under which a grid cell is more likely to lose or gain population. Similarly, the level of activity in a grid cell is assumed to be stable, when the number of people moving to and from that cell is almost the same.

Cellular automata (CA), that are known to be as simple model of a spatially extended decentralized system, are made up of a number of individual components, called cells, each of which communicate between themselves following a set of simple rules [5]. Each individual cell is in a specific state, which changes over time depending on the state of its local neighbours and various inputs from outside the automaton. Despite the simplicity, CA is considered as a flexible and efficient tool being used in modelling numerous physical phenomena, particularly because of their capability of working as an alternative to differential equations. They have also become very popular during the past decade for being well suited to handling geographic phenomena [6].

Hence, the attraction-based CA is formulated as:

$$ST^{t+1}\{i,j\} = f_{MN}(ST^{t}\{i,j\}, P_{ST}^{t}\{i,j\}) \quad (1)$$

Where $ST^{t+1}\{i,j\}$ and $ST^{t}\{i,j\}$ represent the status of cell $\{i,j\}$ at time t+1 and t respectively, $P_{ST}^{t}\{i,j\}$ is the probability of cell $\{i,j\}$ transition to the state ST at time t, and f_{MN} is a transition function, averaging the cell probabilities in a Moore neighbourhood of the cell $\{i,j\}$ including itself.

So,

$$f_{MN} \xRightarrow{Transition\ Rule} ST = \begin{cases} Dynamic & \sigma \geq \beta \\ Stable & \sigma < \beta \end{cases} \quad (2)$$

and,

$$\sigma = \left(\sum_{M_1}^{M_9} P_{ST}^{t}\{idx + M\}\right)/9 \quad (3)$$

Where β is a probability by which the Dynamic state (Infectious) spreads through the Stable state cells, σ is an average of the estimated probabilities assigned to the cell and all its eight Moore neighbours, idx is the Matrix index of cell $\{i,j\}$ in an $m \times n$ Matrix and M is a set of offset values that indicate the nearest neighbours of the designated cell as illustrated here:

$P_{ST}^{t}\{idx\}$ is the probability of cell $\{idx\}$ holding state ST, which is based on the $AtV\ \{idx\}$, the attraction value at cell $\{idx\}$ and $RAND^{t}\{idx\}$ is a random variable that is generated in every iteration without a memory as following:

$$P_{ST}^{t}\{idx\} = (AtV\ \{idx\} * RAND^{t}\{idx\}) \quad (4)$$

In which, $AtV\ \ \{idx\}$ is calculated in every grid cell by summing the Euclidean distance of the cell to the attraction points multiplied by their associated attraction weight.

The CA's development direction is guided towards the main population attraction points (known as the city centroids) that attract the majority of the population transition during different time intervals of the day [4]. In simple words, work places and universities attract population counts during the morning rush hours of a city, while residential lands play the same role in the afternoon when people are mostly heading back to their homes. This is analogous to the concepts of the conventional gravity-based transport modelling, where highly populated city-cores generate trips by attracting and repelling population during the 24 hours of a day. The algorithm chooses the largest population clusters as different attraction points in different time intervals of the day. This attraction rule is simply implemented into the model by estimating the cells' population probability according to their Euclidean distance to the nearest attraction point and their associated attraction weight.

In this framework, where the gridded maps resemble a large matrix of land cells, MATLAB that is mostly popular for its' significant efficiency in matrix manipulations, implementation of algorithms and interaction with other programing languages, is chosen to perform the numerical computations . The use of matrix cells representing the gridded land cells could be particularly helpful in a discrete simulation architecture, in which there are several data layers for the mapped area, namely the land use category, accessibility, attraction points, Day/Night population. These numerous data layers can easily be represented by several uniform matrices that interact with each other and get updated in every time step without losing their spatial precision.

The CA implemented here, is designed to drive the population counts on an another gridded map, in which every grid cell has a primary value for the population count and is assigned with a transitional value in every time step. If a cell holds the Stable state, the number of population count of the corresponding cell on the visualization map does not change. In turn, when a cell's state shifts to Dynamic, then the algorithm either adds or reduces the population count, depending on the time interval of the day, the cell's land use category and the amount of population on the move in each time step. Population-on-the-move is also calculated based on a random probability from 0 to 1 assigned to the number of cells that are losing population count. This value is then distributed uniformly to the cells that are subject to gaining population counts.

The 24-hour simulation time is divided into 290 time-steps, each of which corresponds to 5 minutes in the real world. A tolerance of 1 to 10 minutes was first obtained according to the standard travel-time commonly used in urban transport models and then calibrated based on the simulation's smooth functioning.

3 Simulation Application: Estimating Ambient population for Trondheim Urban Area

After replicating the simulation for 10^6 times, the most fitted datasets were then used to represent the final output of the model forming the synthetic ambient population maps for the 24 hours of a random working day in Trondheim.

Figure 1 shows the dynamics of the ambient population within the 24-hours of a random working day in Trondheim, comparing population distribution in an urban area in different times of the day. It reveals the very different patterns of population transition trends within the city's urban areas in day versus the night-time.

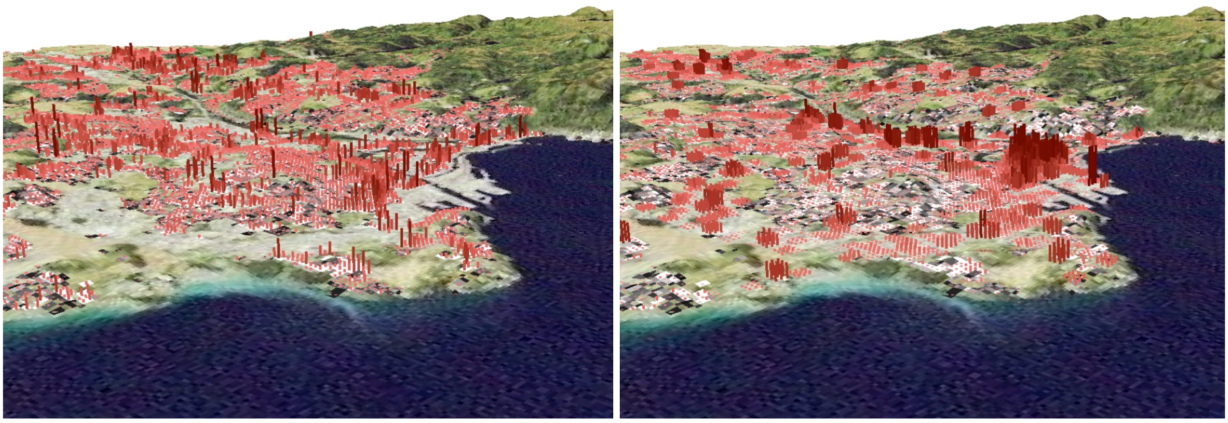

Figure 1: A Synthetic Ambient (Day) Population Map of Trondheim (Norway) based on an epidemics-based geosimulation (on the right), compared to its so-called Night Population Map based on the national census counts (On the Left).
Source: Maps generated by the second author in ESRI ArcGIS 10.1. Algorithms developed by the first author in MATLAB.

There are several places on the day population map, clearly indicated by the dark red colour, that show high-density population areas, which could cause a significant under-estimation in the process of calculating the number of people-at-risk.

4 Model Evaluation

Adequate calibration and validation, which is commonly referred to as model evaluation, is critical for ensuring the credibility and the usefulness of any geosimulation structure.

In this research, a two dimensional evaluation process consisting of operational and numerical validation is believed to be comprehensive enough to reveal the model's deviation from the real world. The model's randomly selected parameters were calibrated first by human inference and replicative mapping techniques. Then the operational validations, which basically evaluate the quality of the model and the way it operates, were implemented by running a series of comparative tests between the geosimulation results and the available real-world datasets. The numerical validation is complementarily implied in form of quantitative comparison between the synthetic maps and the available census datasets, generated by statistical surveys. As there are obviously no actual real-world ambient population datasets in a fine scale spatio-temporal resolution, the decline in the amount of standard errors (RMSE and MAE) is considered as a state of goodness-of-fit in the model. The simulated datasets are therefore compared to the real world population census data for two different times of the day: 1-Night-rest (00:01Am), 2-Day-rest (12:01Pm). The first data set is simply the residential population data set and the second one is a population count based on the number of registered employees in every business address point.

As shown in Figure 2, as the simulation approaches the time-steps, for which the real-world datasets are available, the error value declines too. This means that the approximate allocation of population counts is going to the right direction.

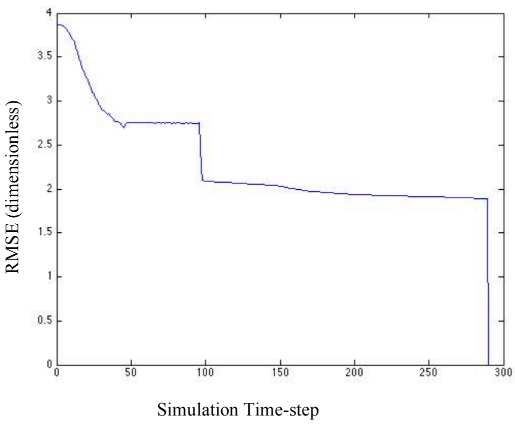

Figure 2: Plotting the RMSE of the model (vertical) in every time step of the simulation (Horizontal).

5 Conclusions

In this research, a methodological approach was developed and tested in form of developing an epidemics-based geosimulation of people's movements through out a random working day. This high-resolution tempo-spatial model, which should be considered more of a proof of concept rather than a formal predictive structure, can be used to generate synthetic ambient population maps of urban areas. This type of demographic information could be used in several urban planning as well as disaster management assessments by studying the movement patterns or making controlled estimates of the number of people at risk.

It is inferred in this article that, the movement patterns, assumed to be the emergent macroscopic behaviour of the cities, considered as complex systems, could be generated based on our basic understandings of cities' and people's microscopic behaviours, such as their travel habits. The decision making process of the moving crowd in a city, if assumed to account for this microscopic behaviour, can then be rationalized based on their obvious needs and the city's spatial characteristics. In other words, people can first be categorized based on their daily activities such as students, workers, etc., then assumed to move to their desired destinations based on their related land use category in different times of the day on a spatially explicit modelling platform of an urban area. For example, students go to institutional land use categories in the morning and commute back to the residential lands in the afternoon. Another approach in setting destination for individuals, employed in this research, was to set highly populated commercial and institutional lands as attraction points in the morning and highly populated residential areas as attraction points in the afternoon in a way that the autonomous land cells change their states based on the land use categories and day time intervals.
There are broader potential applications to the presented work. Generating fine-scale spatio-temporal synthetic population maps, which are critically beneficial to the accuracy as well as the efficiency of disaster management or any similar contingency plans, where initially targeted in this research, while several other applications of this model could still be experimented. In general, the presented model is believed to be able to be used as a powerful decision-support tool if integrated to any other spatial modeling infrastructure. It could either act as an exposure assessment tool or a model for pattern detection and trend prediction applications with a wide variety of implications, ranging from transport modeling and urban growth analyses to business development geo-analysis and climate-change adaptation strategy testing.

References

[1] Batty M, Marshall S. The Origins of Complexity Theory in Cities and Planning. In: Portugali J, Meyer H, Stolk E, Tan E, editors. Complexity Theories of Cities Have Come of Age [Internet]. Springer Berlin Heidelberg; page 21–45, 2012.

[2] Batty M, Xie Y, Sun Z. Modeling urban dynamics through GIS-based cellular automata. Computers, Environment and Urban Systems.1;23(3):205–33, 1999.

[3] Dobson JE, Bright EA, Coleman PR, Durfee RC, Worley BA. LandScan: A global population database for estimating populations at risk. Photogramm. Eng. Remote Sens.;66(7):849–57, Jul 2000.

[4] Roth C, Kang SM, Batty M, Barthélemy M. Structure of Urban Movements: Polycentric Activity and Entangled Hierarchical Flows. PLoS ONE. 7;6(1):e15923, Jan 2011.

[5] Torrens PM, Benenson I. Geographic Automata Systems. International Journal of Geographical Information Science. 19(4):385–412, 2005.

[6] Torrens PM, O'Sullivan D. Cellular automata and urban simulation: where do we go from here? Environment and Planning B: Planning and Design. 28(2):163–8, 2001.

[7] Vliet J van, White R, Dragicevic S. Modeling urban growth using a variable grid cellular automaton. Computers, Environment and Urban Systems. 33(1):35–43, 2009.