UCL CENTRE FOR ADVANCED SPATIAL ANALYSIS

UCL

WORKING PAPERS SERIES

Paper 176 - Dec 11

Calibrating Cellular Automata Models for Simulating Urban Growth: Comparative Analysis of SLEUTH and Metronamica

ISSN 1467-1298

Centre for Advanced Spatial Analysis University College London 1 - 19 Torrington Place Gower St London WC1E 7HB
Tel: +44 (0)20 7679 1782 casa@ucl.ac.uk www.casa.ucl.ac.uk

Calibrating Cellular Automata Models for Simulating Urban Growth: Comparative Analysis of **SLEUTH** and **Metronamica**

Donghan Kim & Michael Batty

d.h.kim@ucl.ac.uk & m.batty@ucl.ac.uk

Centre for Advanced Spatial Analysis
University College London
90 Tottenham Court Road
London, W1T 4TJ, UK

24 December 2011

Abstract

For almost two decades, cellular automata (CA) has proven to be a popular and sometimes effective modeling approach to the study of complex urban systems. Not only as a new methods for predictive simulation but also as a practical policy support tool, CA models have been applied to a large collection of diverse urban regions which now provide a good basis for comparative analysis. After sketching some basic ideas about how CA models can be applied to urban systems, we describe and then calibrate two well known and widely applied CA models, **SLEUTH** and **Metronamica**, to simulate the future urban growth of the Seoul Metropolitan Area, Korea. This is for the express purpose of generating the impacts of practical planning policies on the study area and of conducting comparative explorations of these CA models themselves. The results confirm the value of CA which provides a rich exploratory of knowledge for investigating dynamic urban growth systems and for evaluating the impacts of possible policy options. Moreover, the concurrent use of two generic CA models provides certain insights in the use and development of CA urban models in general and these two models in particular.

Keywords

Cellular Automata, Calibration, **SLEUTH**, **Metronamica**, Urban Growth, Comparative Analysis, Seoul Metropolitan Area

1 INTRODUCTION

Since the first generation of urban models was developed in the form of Land Use Transport Interaction (LUTI) models in the 1950's and 1960's, the dominant style of such model has evolved from an early focus on aggregate and comparative static, cross-sectional approaches to more detailed disaggregate and temporally dynamic procedures (Iacono et al., 2008, Batty, 2009). This is partly due to developments in gathering bigger, more individualized data sets, as well as through dramatic advances in computation but this is all predicated on the fact that there is now general agreement that cities need to be simulated from the bottom up rather than top down. Evolution and change is central to the way cities evolve and it is now widely regarded that such dynamics must be built into the structure of the most applicable models. While traditional urban models pay more attention to the static impact of transportation on land use change, recent urban models have thus focused on the dynamic transformation of urban morphology with much lesser emphasis on transportation *per se* and this has spurred the development of such models (Batty, 2004). Although these new styles of model have less explicit links to policy than their predecessors, such models still have an important role in informing policy-making and supporting the decision-making process.

Cellular automata models have been at the forefront of such disaggregate and dynamic approaches in that they provide an effective and somewhat neutral as well as generic representational framework for the study of land use change and urban growth. Since urban growth occurs in time and space, disaggregate and dynamic modelling approaches provide a useful knowledge base for understanding the urban growth process and the signature spatial patterns that represents their outcomes. In addition, while urban growth can be understood as a complex system characterised by processes and behaviours based on self-organisation giving rise to multiple non-linearities, CA modelling allows us to encapsulate these characteristics of urban systems.

Here we will demonstrate these ideas by applying **SLEUTH** and **Metronamica** to the Seoul Metropolitan Area (SMA), the capital region of Korea. The purpose of our presentation of urban growth simulation with these models is twofold: to calibrate the model for a large, typical urban case study area which is dominated by practical planning instruments such as greenbelts and new towns, and to explore the methodological implications of what these two models are able to predict for these regions through such empirical applications. Firstly, the research reported here seeks to explore future urban growth trends in the study area and the possible consequences of different planning options [1]. To this end, among multiple possibilities, two scenarios are presented, based on *business as usual* and *greenbelt deregulation*. These scenarios are used for each model and the results and implications are discussed at the end of each simulation. Secondly, since the use of two generic models provides enables us to develop a basic comparative framework, this research goes beyond this in seeking more general implications for urban CA models and their applications. However, the objective is not to evaluate the performance of each model *per se* but to sketch the dilemmas of model calibration and the difficulties of developing them for planning support. Although a close comparison of model outcomes is possible, this is quite limited since the two models have different processes of simulating model behaviours as well as different data requirements, despite being based on a generic CA structure. Rather, we try to draw broader implications for CA urban models and their uses in general. Discussion is provided in this regard in our concluding section.

2 CELLUAR AUTOMATA URBAN MODELS

CA systems were originally designed to study self-replication in the natural sciences, originally as computable systems in general and then in fields such as biology and physics. The approach first came to the attention of geographers, particularly Tobler (1979), in the early 1970s where he saw the correspondence between the development of CA by researchers such as Arthur Burke and John Holland at Michigan and his own work in cartographic representation. In this sense his paper on simulating Detroit (Tobler, 1970) launched the field but it was not until the late 1970s, that he first suggested that the geographic phenomenon could be translated into a cellular array and explored through CA mechanisms based on neighbourhood types and transition rules (Tobler, 1979). Tobler suggested five types of model that explain dynamic land use change in a cell space. Although some are not purely CA but cell-space systems and thus closer to the kind of raster operations that one sees in GIS, these models offer important insights such as the integration with GIS layers with principles of land

---

[1] The simulation work in this research was supported by a project from the **OECD** (Organisation for Economic Co-operation and Development). The project involved showing how new kinds of widely available land development models based on CA might inform planning policy in the Korean urban context.

use development based on the notion of CA neighbourhoods. The models proposed by Tobler (1979) can be described as follows:

$$c_{xy}^{t+\Delta t} \neq c_{xy}^{t} \tag{2.1}$$

$$c_{xy}^{t+\Delta t} = F(c_{xy}^{t}) \tag{2.2}$$

$$c_{xy}^{t+\Delta t} = F(c_{xy}^{t}, c_{xy}^{t-\Delta t}, c_{xy}^{t-2\Delta t}, \ldots, c_{xy}^{t-k\Delta t}) \tag{2.3}$$

$$c_{xy}^{t+\Delta t} = F(e_{ij}^{t}, f_{ij}^{t}, g_{ij}^{t}, \ldots, h_{ij}^{t}) \tag{2.4}$$

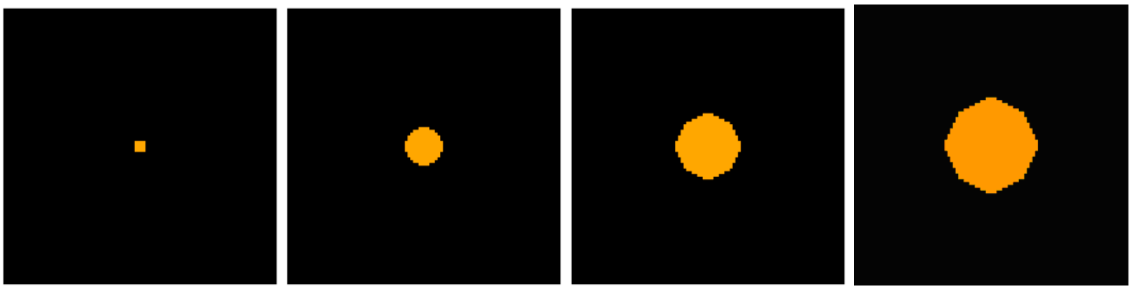
$$c_{xy}^{t+\Delta t} = F(c_{x\forall i, y\forall j}^{t}) \tag{2.5}$$

where c_{xy}^{t} is the land use category such as urban and rural at the cell location *x, y* at time *t*, and $c_{xy}^{t+\Delta t}$ is the land use category at the same location in the future. If Model 1 (equation 2.1) holds, this suggests an independent random land use change which has no relationship with previous land use at the spot. Model 2 (equation 2.2) simply notes that a land use change at location *x, y* at time $t+\Delta t$ functionally depends on the previous land use at that location. Model 3 (equation 2.3) defines historic land use change. Land use change in the future is a result of land use at that location in several previous time steps and this kind of model often appears in econometric formulations where variables at past time periods are lagged in time and influence. Model 4 (equation 2.4) proposes a multi variable (or layer) cellular operation. The land use depends on the several different additional factors at the same or different locations. Model 5 (equation 2.5) suggests an application of a typical CA system where land use depends on the land use of its neighbours in the previous time, that is where $x\forall i$ and $y\forall j$ represent the cell neighbours of x and y called i and j. Despite possible limitations such as complexity of the actual geography, Tobler (1979) concludes that these approaches make it possible to pursue the numerical study of non-numerical geographic systems.

This new approach to the study of geographical representation began to influence urban modelling in the 1980's and soon it became a dominant paradigm. Couclelis (1985) first presented an hypothetical CA model in an urban context, exploring how changes in individual cell states can represent large scale urban change. The development of new methods of urban morphology based on fractals provided a spur to the use of CA for the generation of many fractal shapes across different scales is essentially based on the CA algorithm (Batty and Longley, 1994). These developments were then followed by various proposals that CA might be used for actual urban systems growth which in turn is based on the notion that urban growth is fractal. Combined with GIS, packaged CA models focused on urban growth such as **SLEUTH** and **Metronamica** appeared in the mid to late 1990s with the CA urban model becoming one of the most popular approaches to the study of contemporary urban systems. However, the CA model has not simply remained as a new methodology. Linked to the complexity sciences, it has provided a much broader knowledge framework in which to understand urban systems in terms of interactions between their components, their spatial structure, and their temporal dynamics (Batty, 2005).

The very power of the CA model is the simplicity of its transition rules which gives rise to much richer resulting system behaviour than in other forms of model. Such complex system behaviour emerges from very simple local interactions between individual cells and this is the essence of emergence in terms of the way spatial patterns repeat themselves, in scaling self-similar fashion. As a proof of concept, before proceeding to the calibration of **SLEUTH** and **Metronamica**, we present a couple of abstract CA models to demonstrate how simple local transition rules can be used to emulate a complex urban growth. Furthermore, these examples also show how such basic patterns generated by simple local rules can be further augmented by global level rules which have an analogy with planning regulations and external investments.

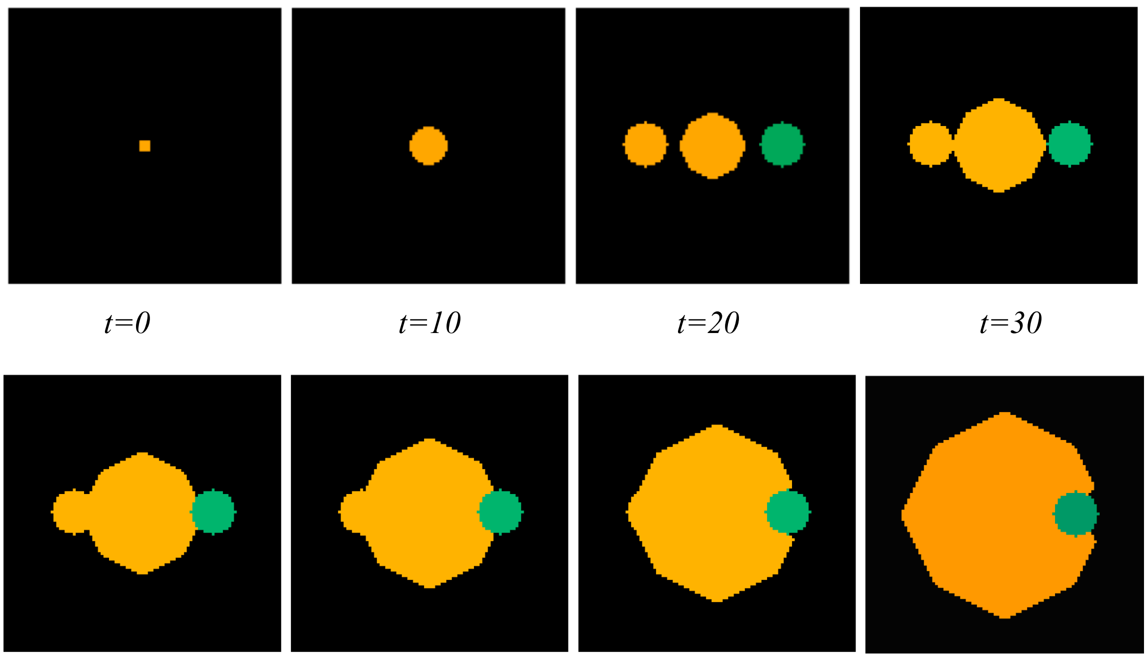
Imagine a two dimensional grid space which contains numbers of cells. Each cell has either of two possible states: urban or non-urban. Each cell checks the state of any cells which comprise its neighbours in its Moore (8 cell) neighbourhood in each time step and updates its own state, dependent upon the status of its neighbourhood in the next time step. Suppose that there is an initial urban centre composed of four urban cells. Let a vacant (non-urban) cell become an urban cell if three or more neighbouring cells are in the urban state. As shown in Figure 2.1, the resulting pattern is a mono-centric urban growth which is well explained in the domain of urban economics. However, note that what drives such growth here is not the decision-making of economic actors but simple interaction between cells. This clearly captures the strength of the neighbourhood effect in the CA model.



t=0 *t=10* *t=20* *t=30*

Figure 2.1. Simulation of Simple Concentric Urban Growth in a CA System

Dynamic and heterogeneous spatial conditions can add more reality to the above model. The following simulation is run with the same initial conditions and transition rules, but let us now imagine a new town development on the left side of the growing urban cluster and a new park on the right side. Such two different entities are introduced at *t=20* for this simulation. The new town itself is not growing although it is in an urban state since no cells in the boundary of the new town area have three or more urban cells in the Moore neighbourhood. The park is also static over time because the area does not contain live urban cells. However, the park is protected from urban development by global regulation while the new town is not protected since it is already urban.



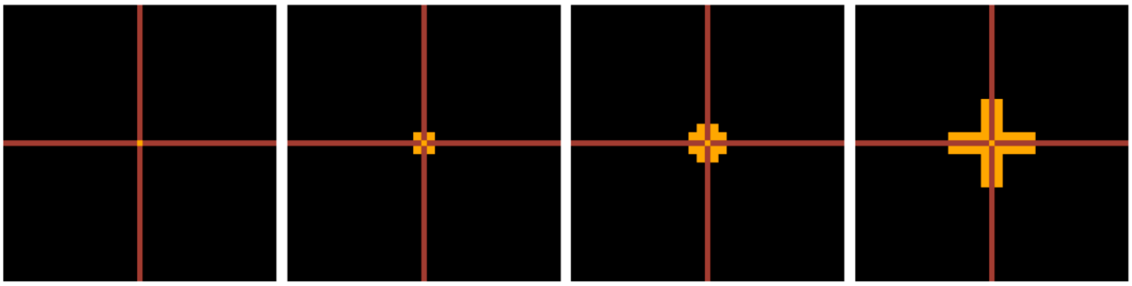
t=0 *t=10* *t=20* *t=30*

t=40 *t=50* *t=60* *t=70*

Figure 2.2. Simulation of Planned Development and Zoning Regulation

As the central urban cluster grows, these two urban clusters merge together forming a conurbation but the protected park area is not affected by ongoing urban growth. Figure 2.2 presents the sort of dynamic change that emerges in simulating such an urban growth system.

Now the following model assumes different initial conditions and transition rules. Let there be an urban core in the centre of space and a road network in its four perpendicular directions, i.e. north, south, east and west. It is an *a priori* condition here which mimics this possible real world geographic features but it might mirror the dynamic introduction of a new transportation network if necessary. Let a vacant cell become urban if there is more than one urban cell in the neighbourhood but only when there is a road cell in the Moore neighbourhood at the same time. The result is a linear urban growth along with the road network as shown in Figure 2.3.



t=0 *t=1* *t=2* *t=5*

Figure 2.3. Simulation of Road Dependent Urban Growth

It is clear that power of the CA model lies in its ability to represent complex system behaviour from such simple local interactions between cells. This of course is the basis of simple diffusion but the discreteness of the lattice on which development is played out leads always to some symmetry-breaking of rules when they are operated with some degree of random noise. The above mechanisms however can be the basic building blocks for developing a CA urban model. However, simulating actual urban systems requires much richer methods and transition rules which depend on many *ad hoc* constraints and invariably stochastic cell transitions. Various modifications in defining cells, cell states, neighbourhoods, and transition rules are also possible and indeed necessary for modelling actual urban systems. We now will demonstrate how CA urban models can effectively simulate urban growth systems by augmenting the standard and generic models with various mechanisms and constraints which characterize the specification and calibration of the **SLEUTH** and **Metronamica** models.

These comparisons will then give us the opportunity to develop further implications for the design and construction of CA urban models in general.

3 THE STUDY AREA

The Seoul Metropolitan Area (SMA) consists of Seoul city, the capital of Korea, and 32 surrounding municipalities. The area of the SMA in 2010 is approximately 11,801 km^2, 11.8% of the total area of Korea, but containing about half of the total population of Korea (some 49m in 2010). This is a large dispersed metropolitan area which is more or less comparable with the Greater South East in the UK. It is located in the north western part of the nation, and the area borders North Korea to the north. The Demilitarised Zone (DMZ) was installed as a buffer zone in the area between South and North Korea after the Korean War, and the area outside the DMZ, which is the northern edge of the SMA, is heavily militarised. To the east, the SMA borders the province Kangwon which is the most mountainous area of Korea. Thus the eastern part of the study area is dominated by a high-altitude area. On the west, it borders the West Sea which is an area containing flat plains and low rising hills. The southern part of the area also has relatively flat areas and it borders Chungcheong province. The Han River which is the main water source of the region flows from east to west in the middle of the region and through the city. Two upper rivers, the North and South Han River, merge outside of Seoul, and the river passes through the middle of Seoul city. The environs of Seoul city are protected by a greenbelt. The key features and overall characteristics of the study area are depicted in Figure 3.1.

The SMA has experienced diverse growth within relatively short time periods although the growth of the SMA is largely shaped by the influence of Seoul city. Urban growth in the SMA was centred on Seoul city until the 1970's. However, after the introduction of the greenbelt in the early 1970's, the physical expansion of Seoul city has been strictly regulated. While the greenbelt has successfully prevented further expansion of Seoul, it could not reduce the need for urban development itself, and as a result, new urban development has occurred in various locations outside the greenbelt in the SMA, thus leapfrogging the constrained area which is reminiscent of growth patterns in many other large cities such as London which have a long history of containment through green belt policies.

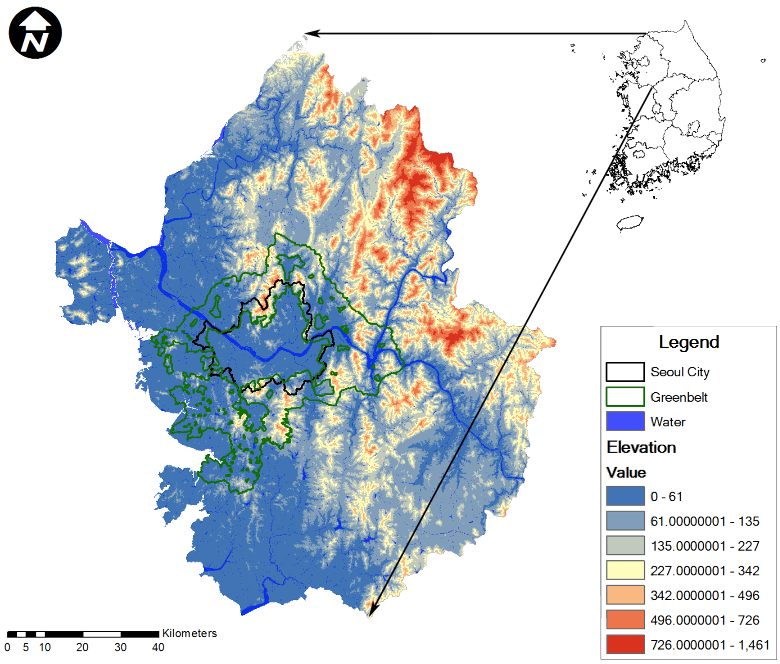


Figure 3.1.General Characteristics of the Study Area

As a result, the SMA as a whole has experienced dramatic population growth over the past decades. The region has been the centre of various high profile socio-economic and cultural activities in Korea – politics, finance, commerce, higher education, research and development, media, and entertainment. Such functional agglomeration once again has attracted population from elsewhere in Korea and from abroad. Although the population growth rate of the SMA has slowed since 2000, the region is still gaining population. The SMA's population increased from 22.0 million in 2000 to 24.3 million in 2009, which is 49% of the total population of Korea. If this trend continues, the population of the SMA will continue to grow, and according to the figures projected by the National Statistical Organisation of Korea, total population of the SMA will reach 25.7 million by 2020 and 26.3 million by 2030 when it is projected that more than 54% of Korea's population will live within the SMA. Table 3.1 summarises the past population growth trends of the SMA, and Table 3.2 shows the projected population growth.

Table 3.1.Past Population Growth

	1985	1990	1995	2000	2005	2009
Whole country	40,419,652	43,390,374	44,553,710	45,985,289	47,041,434	49,773,145
SMA	15,803,288 *(39.1%)*	18,573,937 *(42.8%)*	20,159,295 *(45.2%)*	21,258,062 *(46.2%)*	22,621,232 *(48.1%)*	24,379,491 *(49.0%)*

Source: Statistics Korea, accessed 23/12/2011, http://www.kosis.kr/abroad/abroad_01List.jsp

Table 3.2. Projected Population Growth

	2010	2015	2020	2025	2030
Whole country	48,874,539	49,277,094	49,325,689	49,107,949	48,634,571
SMA	24,336,199 (49.8%)	25,191,245 (51.1%)	25,786,378 (52.3%)	26,161,866 (53.3%)	26,315,824 (54.1%)

Source: Statistics Korea, accessed 23/12/2011, http://www.kosis.kr/abroad/abroad_01List.jsp

Intensive population growth in the SMA has resulted in the dramatic conversion of open space into urban built up areas. The historic urban extent clipped from land cover data by the Korean Ministry of Environment catches rather well the past urban growth trend of the SMA. It is clearly observable that the total urban built-up area of the SMA has significantly increased over time. The data shows that the urban built-up area increased from approximately 5.7 percent of the total land area in 1985 to 15.2 percent in 2006. The increase has slowed during the 2000's, along with a slowing population growth rate of the SMA. However, the urban built-up area in the SMA is continuously increasing, consuming available open space and damaging the natural environment. In addition, scattered urban development is much more notable than 10 years ago. The overall urban growth pattern and the changing form of the study area is depicted in Figure 3.2.

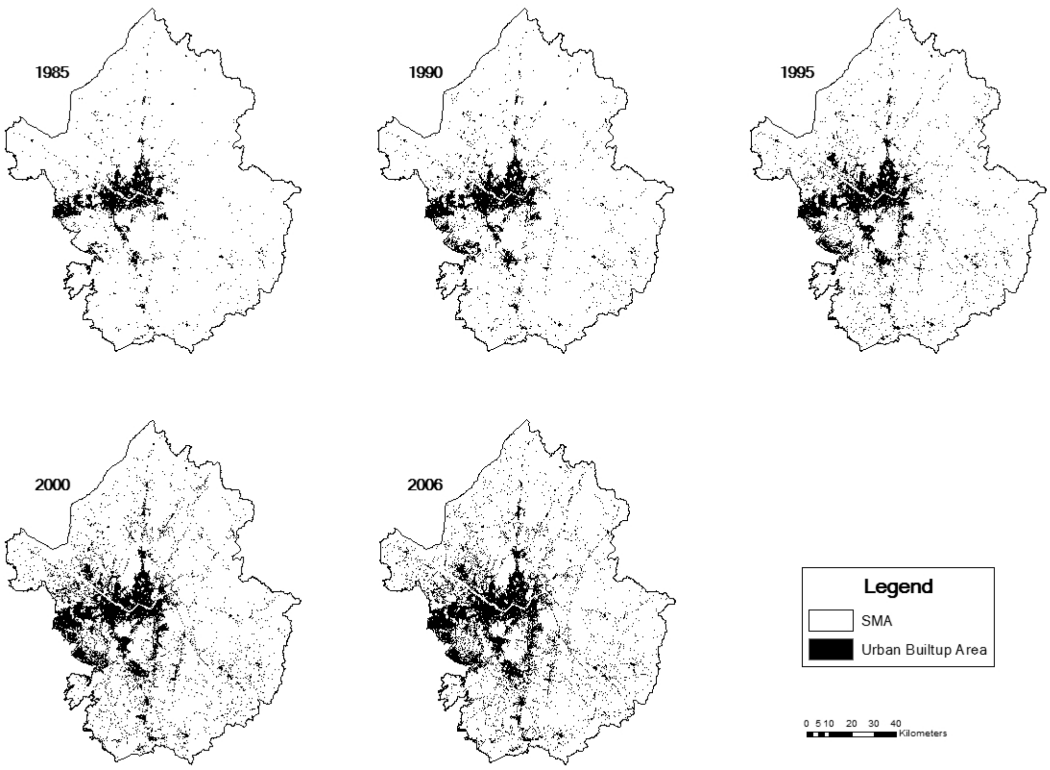


Figure 3.2. Changes in the Urban Built-up Area in the SMA, 1985-2006

.

Two types of urban development have shaped the overall urban growth of the SMA. Firstly, the public sector has led large scale development in the SMA. In an effort to resolve the housing shortage problem in the capital city Seoul, a series of major new town developments took place in the 1990's in areas close to Seoul such as Bundang, Ilsan, and Pyeongchon. More new town development but at a much the smaller scale has occurred more or less continuously at further distances due to the depletion of large scale vacant sites near Seoul. Secondly, small scale development by private developers has followed these larger developments, eventually resulting in a serious urban sprawl problem in the SMA. As a result, the SMA has suffered greatly from urban sprawl over the last decade. Necessary policy measures have been taken to stop undesired urban sprawl, but the small scale and dispersed development pattern still dominates current urban growth in the SMA[2].

---

[2] The sprawling urban growth pattern was identified by conducting a patch analysis of land cover data for 2001 and 2009. Whereas the number of urban patches increases from 10472 to 22900, average patch size decreases from 13.41 to 6.44. The result implies that the urban built-up area has increased but at a much smaller scale and higher urban density.

This rapid urban growth and sprawl in the SMA has resulted in diverse urban problems such as raised infrastructure costs and damaged natural environments. It is foreseeable that future growth would further consume vulnerable agricultural areas. Urban development in the SMA occurs wherever it is possible. However, in a democratic market regime, there are no absolute means to prohibit such spontaneous urban development. It is thus particularly important to understand how complex urban growth occurs and how certain policy actions can intervene to reduce the problems of growth and sprawl. We thus move onto examine what two urban growth models might be able to tell us about urban growth in Seoul, and we begin with the **SLEUTH** model.

4. THE SLEUTH MODEL

4.1 Model Overview

SLEUTH is an urban growth and land use change simulation model originally developed by Keith Clark at the University of California, Santa Barbara, in the early 1990s under the auspices of the US Geological Survey (USGS) and Environmental Protection Agency (EPA). The model was initially applied to the San Francisco Bay area from 1993-1997 (Clarke et al., 1997) and since then the model has been applied to over 100 cities and urban regions around world (Clarke, 2008). The model has provided useful understanding of urban growth and its implications for planning policies in diverse regions and it is clearly one the most widely used CA based urban growth simulation model which focus on urban growth and development (Jantz, et al., 2010). It is worth noting however that this class of model is based on a very different set of assumptions from the other main class of LUTI models that focus much more on activity location and spatial interaction than on actual physical development and land use that is the focus of these CA type models.

The model name is an acronym for six types of spatial data layer: **S**lope, **L**and use, **E**xclusion, **U**rban, **T**ransportation, and **H**illshade. Except for the hillshade layer, which is optionally used as a backdrop image for visualisation purposes, all other five layers are essential for model calibration and future simulation. The model requires greyscale 8 bit GIF images as an input data format which have a pixel (cell) value from 0 to 255. Relevant cell values for each layer are assigned in this range. All input images must be spatially consistent. They must have the same spatial resolution (size of individual cell) and spatial extent (size of entire cell space) so that the cells in all layers can be properly aligned. Since **SLEUTH** is a pure model without data processing capability as part of its core software, such input data need to be pre-prepared with external GIS and image processing software.

A standard CA system framework first of all forms the backbone of **SLEUTH**. The model adopts the core elements of CA systems to simulate urban growth: 1) the cell: the basic computational unit in a CA system. A cell size is defined as an input data resolution in **SLEUTH**; 2) cell space: a two dimensional array of cells. It is defined by the dimension of the input data; 3) cell state: An attribute value assigned to the cell. Each input layer holds relevant cell values between 0 and 255; 3) neighbourhood: the spatial relationship of one cell to another. **SLEUTH** uses a classic Moore neighbourhood, 8 cells based on a 3x3 grid of which the central cell is the focus of the neighbourhood, and 4) transition rules: conditions governing the change of a cell state from one to another. It is typically defined by the states of neighbouring cells in the case of simple CA systems. In the case of **SLEUTH**, the cell transition occurs in the urban layer, but the model incorporates additional information from reference layers such as slope and transportation as well as information from model parameters.

Based on such CA system fundamentals, the urban growth dynamics is jointly determined by a range of additional functions and methods in order to capture realistic urban system behaviour. Basic building blocks are 1) suitability conditions, 2) growth rules, 3) growth coefficients, and 4) self-modification rules. The suitability condition globally filters out those cells that are not subject to future growth and also defines basic potentials for urban growth. This condition is defined by two input layers: the exclusion and the slope layer. The area in the exclusion layer is literally excluded from future growth. In addition the areas with slopes greater than 21 percent are also excluded by default (note that this threshold can be modifiable). All other areas are relevant to future urban growth, but the potential for urbanisation is calculated by the slope value at each cell and the globally defined slope coefficient.

The growth rules form the core of urban growth dynamics in **SLEUTH**. Under the Moore neighbourhood configuration, this defines how individual cells become 'urban' or remain 'non-urban' when they meet certain conditions. **SLEUTH** defines four types of growth rule which occur sequentially and iteratively: spontaneous growth, new spreading centres, edge growth, and road-influenced growth. A set of four growth types forms one growth cycle which represents one year in the simulation environment. Spontaneous growth represents the random urbanisation of land. It simulates the small scale low density urban development which occurs independently from existing factors such as urban clusters and transportation networks. A new spreading centre determines whether isolated single urban cells generated in the previous step will become new urban centres which have a capacity for further urban expansion.

Once the cell is selected as a new spreading centre, two neighbouring cells are additionally converted into urban cells forming an urban block which has three of more urban cells. Edge growth further defines urbanisation from the established spreading centres. This type of growth simulates the expansion of existing urban clusters into their surroundings. If a non-urban cell has at least three urbanised cells in its neighbourhood, then the non-urban cell has a certain probability of becoming an urban cell. Road-influenced growth, as the name suggests, represents urbanisation largely directed by transportation networks and hence by accessibility. In this growth step, growth is jointly determined by the existing transportation network and the most recent urban development generated in the previous three steps. This consists of a range of steps affected by different coefficients, but in a nut shell, it ultimately generates spreading centres adjacent to the road networks, allowing urbanisation of up to two cells along the road. The above four growth rules are controlled by five growth coefficients: namely dispersion, breed, spread, slope, and road gravity. Each parameter has a value from 0 to 100 and guides single or multiple growth rules.

In addition to the growth rules and coefficients that invoke and control urban growth, another rule set kicks in to complete the urban growth dynamics of **SLEUTH**. While the five coefficients are defined as model parameters at the beginning of the simulation, the self modification rules at the global scale dynamically alter certain coefficients during the simulation runs. What this does is speed up or slow down overall urban growth. This self modification feature aims to add a degree of non-linearity to the overall urban growth system and to enable simulation of more realistic urban systems.

4.2 Input Data

SLEUTH runs over a grid space and derives model parameters from statistical analysis of raster based spatial data. Thus having good quality data is a first step in ensuring a successful implementation of **SLEUTH**. However, this project application to the SMA floundered on not being able to find data good enough for this simulation. After checking available digital data, we realized that custom data building would be the more desirable option to better calibrate the model. However, in order to minimise data preparation efforts, we decided to rely on the best available data even though we know that certain layers are incomplete and inaccurate. Inevitably this restricts the validity and usability of the simulation results. These issues with this input data will be further discussed later, but first descriptions about each input data and acquirement detail will be presented in the order in which we build and consider the layers in the acronym **SLEUTH**.

- The **S**lope layer can be typically processed from DEM (Digital Elevation Model) data. This study used the DEM data built and maintained by the National Geographic Information Institute in Korea and we utilised this to create the slope layer. The spatial resolution of original DEM data is 5m, and the base year is 2005.

- **L**and use data is not a requirement for the urban growth simulation. However, this research used land cover data to extract other required input layers such as urban extent and excluded area. The Ministry of Environment in Korea produces different types of land cover data for the nation. What is called the 'low resolution version' has a resolution of 30m and has 7 land categories. It is processed from Landsat TM (Thematic Mapper) imagery. The 'mid resolution version' is processed from 2.5m resolution SPOT 5 (Système Probatoire d'Observation de la Terre 5) and KOMPSAT-2 (Korea Multi-Purpose Satellite-2) imagery. This version of land cover data is however further refined by actual field survey and published in a vector format. This means it can be converted into any raster resolution. It has 22 categories of land cover which break down the former 7 categories of the low resolution version. The high resolution version uses 1m spatial resolution imagery of KOMPSAT-2 (Korea Multi-Purpose Satellite-2) as source data. It has 41 land classifications which are further subdivided from the categories of the mid resolution version. But this high resolution only covers a part of the nation at the moment and in this project, we acquired the low resolution version for 1985, 1990, 1995, 2000, and 2006, and the mid resolution version for 2001 and 2009.

- The **E**xclusion layer was created from a combination of natural barrier and institutional regulations. The natural barrier simply included water bodies which are extracted from the low resolution version land cover data for 2006. Then this was combined with the greenbelt area which was obtained in vector format. A partial exclusion is not considered. The urban extent layer is extracted from the low resolution land cover data while the years chosen for the calibration are based on data at 1990, 1995, 2000, and 2006.

- The **T**ransportation layer was the most difficult input data to prepare for this simulation. Although various GIS data coverages covering a wide range of land use and transport data are commonly available nowadays, at least for the study area of the SMA, time series geographic data are extremely rare except for that processed from satellite imagery. In this project although we could obtain the whole road network data for the study area, (which is in vector format holding all information about road classes and types), we could not obtain dedicated historic transportation network data. Only a single time was available for the road data which is dated to the year 2005. At least two historic time points are necessary for model calibration. We consequently decided to use incomplete transportation data extracted from a series of land cover data. The alternative option was to extract the "transportation" category from the mid resolution land cover data. Two transportation layers, 2001 and 2009, were available as a result. However, this transportation data has quality issues, for it is not strictly speaking considered as route data but as land use area data and thus difficult to use as a proxy for transport networks. As non-dedicated road data, this not only includes road networks but also auxiliary transportation facilities such as car parks and even airport runways. It is clear that this needs to be much refined if it is to be seriously used for urban development simulations. Besides, this extracted data does not have attribute information about road hierarchies. In this case, major motorways and local roads will have same attractiveness level which is a somewhat unrealistic assumption. Despite these problems, we decided to use the data without custom manipulation since such corrections would have required significant time and cost with the quality of improvement still remaining in doubt.

- **H**illshade is a grey scale image that facilitates the interpretation of the terrain surface. If overlaid as a background, it greatly enhances visual readability of the base map. It is typically produced from DEM data using simple automatic functions in many GIS applications. The hillshade layer for this study was also created from the DEM data described above using a standard GIS package.

All source data were collected under such conditions, but deciding a suitable resolution for calibration and simulation was a difficult decision. Technically the finest resolution possible for this case is 30m, which is the low resolution version of the land cover data. Although other sources are available at a finer resolution, resampling the 30m resolution data to a finer scale is pointless. Thus, in estimating computing power required and considering data quality, we initially adopted 50m resolution and prepared the input data accordingly. The grid dimension was 2650×3078 for the whole study area including the 'no data value' areas. However, the computer used for this simulation could not initiate the calibration, returning a memory error[3]. We then tried different levels of resolution and finally decided that 100m was an appropriate input data resolution. The grid size thus becomes 1325×1539 for the study area. Then the data were further re-sampled at 200m and 400m resolution for the calibration process, which is a requirement of the model. Details of input data layers are described in Table 4.1

Table 4.1. Source Data and Descriptions

Layer	Source	Raw Data Provider	Original Resolution	Base Year
Slope	Processed from DEM	National Geographic Information Institute	5m	2005
Land Use	Extracted from Low Resolution Land Cover	Ministry of Environment	Vector	1990, 1995, 2000, 2006,
Excluded	Extracted from Low Resolution Land Cover	Ministry of Environment	30m	2006
Urban	Extracted from Low Resolution Land Cover	Ministry of Environment	30m	1990, 1995, 2000, 2006
Transportation	Extracted from Mid Resolution Land Cover	Ministry of Land, Transport, and Maritime Affairs	vector	2001, 2009
Hillshade	Processed from DEM	National Geographic Information Institute	30m	2005

3 **SLEUTH** requires significant computing power. It is often necessary to use parallel computing or rewrite the source code to run the model for large areas at fine resolution.

4.3 Model Calibration

Running **SLEUTH** for predicting future urban growth requires the model to be calibrated beforehand. Generally speaking, the calibration of **SLEUTH** involves adapting the **SLEUTH** generic model to a particular study area by applying a parameter set unique to that area. More specifically, the main purpose of the calibration of **SLEUTH** in this case is to determine the best fit value for the five growth coefficients (dispersion, breed, spread, slope, and road gravity).

The calibration of **SLEUTH** is automatic and achieved by using a so called "brute force" algorithm and supported by related statistical methods. Examining all possible cases until a solution is found is a useful problem-solving strategy, but it is only practically possible with the use of large scale computation requiring significant run time. During the calibration, all possible combinations of parameter values are applied to the past urban seed data and then the simulated results are checked against the historic urban data to see if the model reproduces known observed growth patterns. However, **SLEUTH** does not automatically pick a single best fit parameter set as a result of calibration. It creates 13 metrics which can be used to evaluate the goodness of fit between the simulated and observed. In more detail, **SLEUTH** produces statistical correlation scores for 13 predefined measurements along with each combination of five parameters. The measurement metrics include the total number of urban pixels, urban clusters, urban edges as well as other features.

The calculation of statistical correlations for 13 metrics for every combination of parameters in each phase is automated, but the selection of a best range for next step is a role for the user. The difficulty is that each of 13 metrics compares different aspects of the spatial patterns. Thus there is no one right answer to evaluate the goodness of fit between the modelled and observed outcomes. Different researchers choose different measurements, but the LeeSallee metric has been among the most popular choices. However, recently Dietzel & Clarke (2007) have developed a new measurement, OSM (**O**ptimum **S**LEUTH **M**etric), and it has been claimed by the authors that the OSM is a better measure than other 13.

The whole calibration process is broken down into three consecutive steps which gradually narrows the search range for optimal coefficient values and increases the resolution of input data. The ranges of each coefficient derived from the first step are entered into the second step and the same goes for the rest of the steps. The first calibration phase, termed coarse calibration, explores the entire range of coefficient values with large increments in parameter values. A quarter of the resolution images from the original full input resolution are used for this initial step. The second step, fine calibration, explores the narrowed coefficient values using a smaller increment. This step uses half resolution images and produces further narrowed coefficient values. The third step, final calibration, uses full resolution images and examines further narrowed ranges with a much smaller increment. Then, a single best fit parameter set is determined here.

However, the set determined in the final phase is not yet complete. It is not ready to be used for the prediction mode. Although the whole calibration consists of the above three steps, one more additional treatment is necessary to get the best fit parameter values for prediction runs. Due to the self-modification function of **SLEUTH**, these starting values of coefficients will be altered at the end of simulation year. The self modification increases or decreases the growth coefficient values as the simulation continues. To initialise the future simulation, it is desirable to use the values at the end year of the calibration than those at the beginning. A solution is thus obtained by using the best fit coefficients derived in the final phase and running the model again over the calibration period. Then the model will produce 'self-modified' coefficients.

Our research adopted the standard three step calibration process described above and used the OSM to evaluate a goodness of fit. The calibration was conducted over the data between 1990 and 2006. The initial phase was the coarse calibration. Re-sampled images with a resolution of 400m were used. The entire range from 0 to 100 of coefficient values was assigned with an increment step of 25. A low number, 4, of Monte Carlo iterations was assigned. The result of the coarse calibration phase was evaluated using the OSM, and then the ranges were selected from the top 5 scores. The result obtained in the coarse phase was then entered for the initial coefficient ranges of the second phase involving the fine calibration. The resolution of input images was reduced to a half from full resolution for this step, and the number of Monte Carlo iterations was increased to 7. The result was also analysed using the OSM, and then the ranges for the next phase were selected from the top 5 OSM scores. In the last phase, the final calibration, the ranges obtained from the fine calibration are applied to full resolution images. Now the aim is to determine a single best set rather than a range. The number of Monte Carlo iterations was increased to 10 for this step. As a result of this step, the best coefficients were selected from the top OSM score: 100 for the dispersion coefficient, 91 for the breed, 1 for the spread, 63 for the slope resistance, and 61 for the road gravity. However, since these are the best set for the beginning calibration year, the ones at the end of calibration after the self modification step is made, are necessary for future simulation. A higher number of 100 Monte Carlo iterations were conducted to find the final coefficients for the prediction, which means 100 simulations were run with the best parameter set produced in the final stage and then the coefficient values presented at the ends of calibration year are averaged over 100. The chosen ranges for each coefficient in each step as well as the final values after the application of self modification rule are described in Table 4.2.

The derived parameters through such a calibration process characterise past urban growth patterns of the study area although these values are bounded by the quality of the input data. If this issue is not considered, some local characteristics can be inferred. A low value of the dispersion value implies that small scale urban sprawl is less dominant in the area. Low scores of the breed and spread coefficients show that such isolated urban developments are not likely to become spreading centres thus attracting new urban developments in their surroundings. The high value of slope resistance tells us that the urban growth of the study area is greatly limited by topography. Finally, the low value of the road gravity parameter implies that the urban growth in the area is less affected by transportation networks.

Table 4.2. Calibration Results

	Selected Values in Each Step			
	Coarse	**Fine**	**Final**	**Self Modification**
Dispersion	100-100	100-100	100	21
Breed	75-100	90-100	91	1
Spread	1-1	1-1	1	19
Slope	50-75	55-65	63	100
Road	50-100	60-70	61	1

4.4 Simulation Results

We present two scenarios to simulate different growth dynamics generated by distinctive policy options: business as usual and deregulation of the greenbelt. There could be many different scenarios depending on the policy interests at hand, but these two scenarios are chosen to illustrate important and expected urban growth momentum for the study area. While the first scenario focuses on the extension of the *status quo*, the second scenario emphasises the effect of the new policy action. Followed by the calibration on data between 1990 and 2006, the prediction was run from 2006 to 2030. Although the data used, especially the transportation layer, are incomplete and inaccurate, the model produced convincing results in comparison with the general characteristics of the study area.

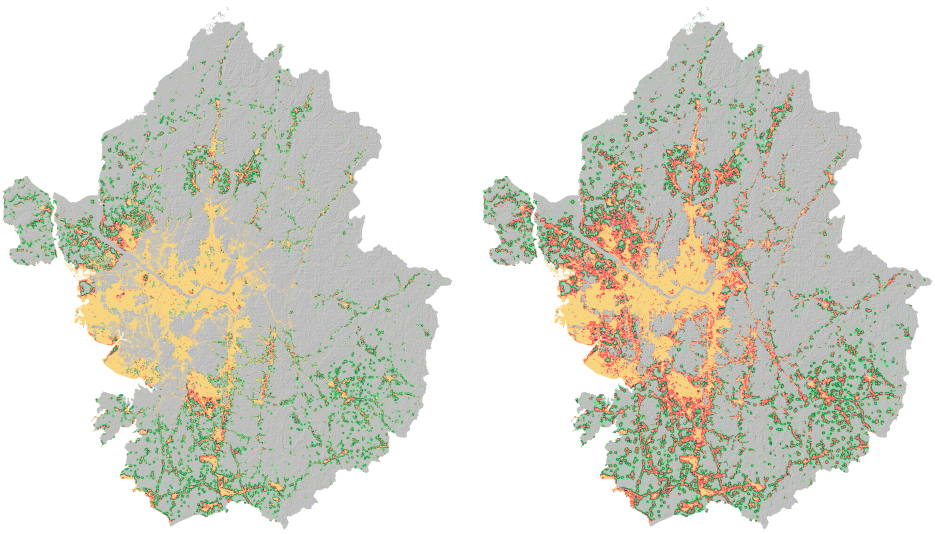
The 'business as usual' scenario assumes the future as an extension of past trends. The term 'business as usual' is often used to describe no particular additional intervention in the future, and this is often compared with other scenarios with intended policy actions. Typically no significant regional constraints are considered in the case of the 'business as usual' scenario. However, since a greenbelt has protected expansion of Seoul city over the past decades, such a constraint is included in the exclusion layer for this scenario. This scenario allows urban development to continue but restrains the growth in the designated greenbelt area. The total simulated urban area by 2030 in this scenario is approximately 2264 km^2, and about 19.9% is urban in whole study area. The net increase of urban land is 4.8%, compared to the urban land in 2006. In terms of spatial allocation, this scenario generates less urban development around Seoul city but creates small urban clusters up and down the study area where slope values are relatively low. Such urban forms are more dominant in the areas outside the greenbelt implying leapfrogging sprawl.

The greenbelt deregulation scenario removes the greenbelt restriction while maintaining all other conditions used for the first scenario. This can allow maximum development for the region. In this scenario, urban land increased from 15.1% in 2006 to 27.5% in 2030. The total simulated urban area by 2030 is about 3126 km^2. In terms of urban form, this scenario showed more clustered development around Seoul city. However, considering the growth rates and patterns in the region during the past decades, this is too radical a pattern of urban growth and looks implausible. New growth not only occurred in the removed green belt area but also in all other areas of the SMA. The reason seems to be due to the values of the model parameters. The best parameter set which was derived when the greenbelt existed, worked for the first scenario but not for the second one. Thus a new parameter set is desirable for this scenario, but this possibility not yet been attempted.

Table 4.3. Simulation Results by 2030

	Total Urban Cells 2006	Total Urban Cells 2030	Percent Urban2006	Percent Urban 2030
Business as Usual	171774	226406	15.1	19.9
Greenbelt Deregulation		312555		27.5

A computer simulation model can act like a virtual laboratory which enables the exploration of various 'what-if' scenarios. However, consideration of a certain policy intervention which can abruptly alter future growth trends over the best parameter set derived from the past patterns can return unexpected results. As shown in the simulations in this study, the parameter set derived from analyzing past patterns worked with the scenario of no change but produced too much growth in the outcomes for the scenario based on greenbelt removal. A new calibration with an improved data set and close examination of why this happened could be a possible future extension of this simulation. Overall growth patterns and characteristics are given in Table 4.3 above and Figure 4.1.



(a) Business as Usual (b) Greenbelt Removal

*Figure 4.1. Results of Urban Growth Simulation with **SLEUTH** at 2030*

5. THE METRONAMICA MODEL

5.1 Model Overview

Metronamica is a CA based land use change model, developed and managed by the Research Institute for Knowledge Systems (RIKS). The model is built upon the pioneering work of White and Engelen (1993) and White et al (1997) who introduced a constrained and integrated CA urban model. The model was firstly applied to the city of Cincinnati, USA. Up until now, it has been applied to a large number of cities and regions around world, including Dublin (Ireland), Milan (Italy), Wuhan (China), Vitoria-Gasteiz (Spain) as we as many other places where land use change dynamics and possible consequences of alternative policy options have been simulated (a complete list of applications as well as the software specification is given in RIKS, 2011). The model was designed to study changes among multiple land use classes, but it is also possible to focus on the dynamics of urban and non-urban land conversion.

Metronamica employs the basic principles of CA modelling but more greatly relies on a series of innovative methods whereas **SLEUTH** is simpler and more traditional Three key characteristics distinguish the **Metronamica** model from conventional CA models: distance decay functions, integration with GIS, and constrained cell transition. Firstly, the model uses a larger concentric neighbourhood configuration and incorporates the notion of distance decay into its modelling framework to define the relationship between a cell and its neighbours. Conventional CA models usually use either 4-cell von Neumann or 8-cell Moore neighbourhood configurations. Here only immediately adjacent cells - the Moore neighbourhood - is the default neighbourhood and affect the centre cell's transition. However, it is more realistic to assume a larger neighbourhood interaction in the case of urban models because the land use state is not only affected by its immediate surroundings but also by features in more remote locations. To this end, **Metronamica** defines the size of a 196-cell concentric neighbourhood, a radius of 8 cells from the centre cell, as a default neighbourhood although its size can be adjusted as a model parameter. The centre cell has a one-to-many relationship in the neighbourhood, and the strength of this relationship generally diminishes as the distance increases thus implying distance decay. The collective influence from all cells to the centre cell in a given neighbourhood is defined as a neighbourhood effect in the model. Defining the degree and magnitude of such neighbourhood effect is a matter of model calibration, and this will be discussed later. It is worth noting that it is this property that destroys the concept of strict emergence in the model and forces comparisons to LUTI models in which interaction fields based on distance decay are central to the notion of the way cities and regions are organized. The argument is often that if CA models are relaxed in this way, they lose their pedagogic and informative value in terms of simulating emergence.

Secondly, the model integrates the CA modelling framework with GIS technology. The use of GIS data not only makes it possible to initiate the model from within the actual geography but also suggests a way of taking into account the effect of various driving forces contributing to land use change dynamics. In addition to the interaction within the neighbourhood, **Metronamica** further integrates GIS data in order to introduce the influence of additional key factors: zoning, suitability, and accessibility. Consequently the model assumes that the land use change is jointly brought about by an interaction between four major factors: spatial interaction with surrounding land uses, zoning, suitability, and accessibility. On the other hand the integration with GIS does not necessarily only mean the use of GIS data. The model also incorporates GIS technologies to analyse and visualise input data as well as model outcomes. Thirdly, the model constrains the total amount of cell transition through the use of exogenous variables. In a general CA system, cell transition is only governed by local interactions, not by other mechanisms. This then leads to an unpredictable global level outcomes. However, the constrained CA model **Metronamica** globally regulates the occurrence of local patterns. In other words, the model does not sum up all possible changes at the local level. The model calculates a ranked score for each cell and then makes an allocation considering the total amount defined. The rank score, termed transition potential, is calculated for each cell in each time step by using the above four major factors: spatial relationships with surrounding land uses, and overlaid zoning, suitability, and accessibility information. No matter how high the transition potential, the cell's future transition can be limited depending on its exogenous parameters. In this way only limited numbers of cells are allowed for state change in each time step. The merit of this approach is that the model can incorporate meaningful indicative values from various socio-economic macro models and data. For instance, the exogenous parameter can have meaning for macro level land use demand.

Built on the above conceptual framework, the following equations best describe key determinants of the transition potential in detail as well as the elements of model calibration.

$$\hat{N}_{i,j} = \begin{cases} N_{i,j}(1+e), & if\ \alpha \geq 0 \\ N_{i,j}, & else \end{cases} \tag{5.1}$$

$$T_{i,j} = \begin{cases} \hat{N}_{i,j} S_{i,j} Z_{i,j} A_{i,j}, & if\ \hat{N}_{i,j} \geq 0 \\ \hat{N}_{i,j} \left(2 - S_{i,j} Z_{i,j} A_{i,j}\right), & else \end{cases} \tag{5.2}$$

where $N_{i,j}$ represents the neighbourhood potential in a cell i for an actively changeable land use class j before the consideration of random perturbation effect, α is a parameter which decides the existence and extent of the stochastic perturbation [0, 1], and e is a random value taken from a Weibull distribution (1/ α, 1). $\hat{N}_{i,j}$ is the neighbourhood potential after taking into account the random effect. Respectively in the cell i for the land use class j, $S_{i,j}$ is the suitability, $Z_{i,j}$ denotes the zoning, and $A_{i,j}$ stands for the accessibility. $T_{i,j}$ is the resulting transition potential score which varies with four main factors as well as consideration of the random disturbance.

The neighbourhood potential forms the core of the transition potential, while three other factors augment the CA dynamics by bringing essential factors relevant to land use change. It is worth investigating the neighbourhood effect in more detail. The neighbourhood effect is defined by:

$$\hat{N}_{i,j} = \sum_{b \in S(a)} I(a,b,d) D(a,b) \tag{5.3}$$

where *S(a)* represents the neighbourhood of a cell *a*, *b* is a member of *S(a)*, and *D(a, b)* is the Euclidian distance between the cell *a* and *b*. *I (a, b, d)* is the influence function describing the style and strength of relationship between the cell *a* and *b,* which is also affected by the distance *d* between cell *a* and *b*.

Thus the neighbourhood potential is the sum of the distance and influence function for each cell in the neighbourhood. The use of a fixed neighbourhood size for all cells in the study space and an Euclidian distance between one cell and another implies that the influence function is a key parameter for determining the neighbourhood potential. One land use type may attract or repulse another type by varying degrees. All cells in the neighbourhood are related to the centre cell in one way or another. In general cells in nearer locations in the neighbourhood will have a larger influence. However, an opposite case also exists, and such an effect is not likely always to be linear. Many different types of neighbourhood effect can exist in the real world. Thus, identifying/defining such relationships is one important element of the calibration of **Metronamica** model. The model provides some predefined rule sets which describe the influence relationship between land uses in order to facilitate the calibration process but users need to be immersed in their own application so that unique rule sets can be identified.

5.2 Input Data

Metronamica requires five GIS based input layers: land use, suitability, zoning, accessibility, and the boundary. Layers such as suitability and zoning are actually value added information that hold composite scores. In that case, the layer requires additional input factor data. This study relied on the data available from the public sector rather than custom built data. Although there was an accuracy problem such as inconsistencies between land use maps at different years, generally fine scale spatial data were available for the given study area. The study area for this simulation is the same as for **SLEUTH** above, the Seoul Metropolitan Area (SMA). The following section describes the data set used for model calibration and simulation run for the study area, some of which was used above in the **SLEUTH** application.

For the land use map, what is called the 'mid resolution land cover data' produced by Ministry of Environment in Korea was used. We will repeat the data specification to remind the reader of the nature of this data. Such land cover data is fundamentally based on the 2.5m resolution SPOT 5 (Système Probatoire d'Observation de la Terre 5) and KOMPSAT-2 (Korea Multi-Purpose Satellite-2) imagery but published in vector format after refining the data by back-up field survey. The mid resolution land cover map originally had 22 land classes, but it was reclassified into 9 categories for this simulation: agriculture, forest, grass, barren, urban, wetland, water, recreation[4], and transportation. Then each of these was assigned to three land categories which is a requirement of the **Metronamica** model. Since the urban growth simulation is targeted, the function category includes only one land use class, urban. The vacant category consists of agriculture, forest, grass, and barren. This means that land uses in this category are available for future urban growth. The feature category is composed of wetland, water, recreation, and transportation. The land uses in this group will

[4] It includes open space not subject to urban growth such as golf courses and theme parks.

remain static during the simulation. The land use map for 2001 was used as a seed layer for the calibration, and the results were compared to 2009 data. Then the 2009 map was used as a seed for the future simulation.

The suitability layer mainly takes account of the terrain condition of the study area. It was created by jointly considering the height and slope condition. After firstly excluding the area over 200m, it classified the area into four categories with percent slope values. The area with slopes over 20 percent was set to have 0 value which excluded them from urban growth. The slope values from 20 to 11 and 10 to 5 were assigned to values of 2 and 1 respectively. The values from 5 to 0 were classified as 3, which mean the lowest topographical resistance. The zoning layer included the greenbelt information which is the most important spatial regulation in the study area. Except for the area protected by the greenbelt, future urban growth is permitted without further restrictions. Two different versions of greenbelt data were prepared in order to assume different planning scenarios. One represented the currently active greenbelt. The other reflected possible adjustments which are part of an ongoing planning policy agenda for the study area. Detailed descriptions will be dealt in the relevant scenario section.

The accessibility layer used a comprehensive road network data for the study area. The level of accessibility was defined into 4 levels depending on the type of roads: highway, major road, minor road, and local road. In addition, the newly proposed high speed railway routes and stations were prepared for an alternative scenario to simulate the impact of such a new railway system. The details of this will be also described along with the relevant scenario. A spatial resolution of 50m was decided for the simulation. Generally finer scale data better describes geographic details but there is a trade-off between data resolution and computing resources. Technically the finest resolution possible for this simulation is 5m. However such fine scale was never likely to be practical for this study. The research initially tried to run the calibration with 25m resolution data which gives a grid size of 5292×6168 for whole study area. However as with **SLEUTH** the system could not be run at this level of resolution. After exploring alternatives, we finally decided to use 50m resolution, which gives a grid size is 2649×3084 in this case. Details of input data are presented in Table 5.1.

5.3 Model Calibration

The calibration process of **Metronamica** can be generally broke down into four phases although they are not exactly sequential: 1) specification of an exogenous parameter that controls the total quantity of land use change, 2) definition of the neighbourhood effect which details the relationship between land uses and governs the resulting local level land use patterns, 3) determination of the random perturbation parameter that adds a degree of stochasticity in land use distributions, and 4) calibration of suitability, accessibility, and zoning that reflects the CA based dynamics in geographic heterogeneities. By adjusting 2), 3), and 4), the modeller creates the transition potential score (See Equations (5.1) and (5.2) and then sets a cutline by specifying 1).

Table 5.1. Input Data and Descriptions

	Layer	Source	Raw Data Provider	Original Resolution	Base Year
Land Use		Reclassified from Mid Resolution Land Cover	Ministry of Environment	Vector	2001, 2009
Suitability	DEM	Original	National Geographic Information Institute	5m	2005
Suitability	Slope	Converted from DEM	National Geographic Information Institute	5m	2005
Suitability	Water Body	Extracted from Mid Resolution Land Cover	Ministry of Environment	Vector	2009
Zoning	Greenbelt	Original	Ministry of Land, Transport, and Maritime Affairs	Vector	2008
Zoning	Greenbelt Adjustment	Original	Ministry of Land, Transport, and Maritime Affairs	Vector	2010
Accessibility	Road Networks	Original	Ministry of Land, Transport, and Maritime Affairs	Vector	2005
Accessibility	GTX[a] Routes and Stations	Original	Gyeonggi-Do[b]	Vector	2013
Area Boundary		Processed from Administrative Boundary	National Statistical Office	Vector	2005

[a]stands for Great Train eXpress, which is a new metropolitan high speed rail system currently under planning for the study area.
[b]Local government that covers most of study area except for the Seoul and Incheon city.

Firstly, the calibration should start with a definition of exogenous constraints. To initiate the calibration, it is necessary to identify the total number of cells for each actively modelled land use type for the beginning and end calibration years. As a constrained CA model, the

Metronamica model allocates this quantity in the study space based on the transition potential scores. A misplacement of the global constraint would generate unrealistic results by allocating excessive or insufficient land use change. A simple way is to count the number of cells from the land use maps. However, this is a physical observation which bounds the calibration results to the observed quantity of land use cells; it is not a socio-economic prediction or projection which can be used to generate future growth. The constraints for the future may be simply extended from the cell counts of past land use classes. If a socio-economic link is desired for the model, a new exogenous assumption, analysis, or projection of these land requirements is necessary.

Secondly, the calibration of the neighbourhood effect is necessary. As described in Equation 5.3, the neighbourhood effect is a function of the distance and the strength of influence between land uses. The function can be a simple linear, quadratic, cubic, or more complex function depending on the characteristic of the system under study. A modeller must decide the type of relationship between land uses as well as the distant dependent magnitude of the relationship. To simplify the calibration process, the model introduced a spline interpolation method. Then the specification of four control points which have fixed and parameterised X and Y values define the neighbourhood influence function. The first point should be on the (X=0, Y=neighbourhood parameter 1). The zero value of X means that it is the centre cell itself in the neighbourhood, and the parameterised Y value represents an inertial force to remain as a current cell state, i.e. a given land use type. The second point should be on the (X=1, Y= neighbourhood parameter 2). The distance 1 is fixed by the model, but the Y value depends on user definition. The third point can be at any distance between the second and fourth with any strength value (X= neighbourhood parameter 3, Y= neighbourhood parameter 4). The last, fourth, point should be located in the (X=max distance on the neighbourhood, Y= 0). This limits the spatial boundary of neighbourhood influence. From here and beyond, the neighbourhood effect becomes zero. By specifying the values for the above four points, a modeller actually defines the influence function and its curve. Though simplified by an interpolation method, finding a relevant neighbourhood influence function from a vacuum for a study area is like finding a needle in a haystack. **Metronamica** assumes certain common basic patterns can exist between land uses in a general sense. Thus the calibration of the neighbourhood effect in **Metronamica** often starts with ones used for previous studies, especially the one originally applied to Cincinnati, USA (White et al., 1997). Then these functions are finely adjusted for the given study area.

Thirdly, it is also necessary to determine the random disturbance parameter. The random factor controls three aspects of emerging land use patterns: the density gradient of land uses, the seeding of new clusters, and the degree of irregularity of cluster boundaries (White and Engelen, 2003). In sum this determines the scatteredness of land use patterns as well as the geometry of individual land use clusters. A sound value of this parameter helps to preserve the stochastic nature of the urban system. Too low or high values result in unrealistic symmetry or disorder.

The final part is adjusting suitability, zoning, and accessibility factors. Physical and institutional suitability can have an effect on calibration since together they form the function of transition potential. But these are more close to description of initial (or interim) conditions, and thus they are less relevant to the model calibration. Accessibility is also a kind of condition, i.e. infrastructure. However, the influence weight is clearly matter of calibration. It determines the degree of land use change influenced by varying the type of road network.

The most effective means of calibrating **Metronamica** is by visual map comparison, followed by iterative changes of parameter values and investigations on the goodness of fit. Globally aggregated statistical metrics are less relevant to determine the goodness of fit. The total amount of growth generated by the model will be always the same since it is globally constrained by an exogenous parameter; thus what is important here in the calibration process is comparing locally distributed patterns. Unfortunately an effective method to make a local level comparison is not yet available. Although the Map Comparison Kit can create cell comparison statistics such as the Kappa[5], Kappa Location[6], and Kappa Histogram[7], these are not good enough to compare locally distributed spatial patterns, especially the patterns generated by the varying neighbourhood effect. Consequently the best resemblance is judged by human intuition. Then the question is when to stop the calibration and by what criteria? Unfortunately there is no right answer for this yet. It is the modeller's decision.

The **Metronamica** model was calibrated with reference to the above elements. The model was calibrated using land use data from 2001 to 2009, but as an urban growth simulation only the urban land use category was considered as the actively changing land use type. The total number of urban cells was counted for each year and used as the global constraint. The

---

[5] The product of Kappa Location and Kappa Histogram
[6] Counts the locations simultaneously taken by each land use category
[7] Compares the total number of cells in each land use each category

neighbourhood influence function for urban land use was initially defined using the default function in **Metronamica**, and then it was gradually adapted to the study area. Urban land use is generally irreversible, and the study area also presents such a nature. Hence the influence function was set to have a high inertia value with a positive agglomeration effect. A random coefficient of 0.6 was used in this regard. The suitability and zoning layer were not adjusted for the calibration. An importance weight and distance decay parameter for the accessibility layer was decided in a way that the model reproduces a similar road influenced growth. As explained before, the model calibration had to rely on the visual map comparison with reference to the Kappa statistics. The map comparisons were repeated until a suitable parameter set was found. After repeated trial and error, the final parameter set was determined. Detailed values are presented in Table 5.2.

5.4 Simulation Results

The two same scenarios, the 'business as usual' and the 'greenbelt deregulation', policy instruments are applied using the simulations with **Metronamica**. The model is run from 2001 to 2030 for each scenario. The simulation from 2001 to 2009 is used as the calibration period. Then the model is run up until 2030 with different policy scenarios after the calibration. However, the calibration of local level dynamics does not yield a value for the global constraint. For future simulation, a new exogenous assumption, analysis, or projection of the land demand is necessary. In order to make this logical, the macro level urban land demand for this simulation was derived from projected population growth for the study area. From the past trends of population and urban growth and the projected population growth published by the National Statistical Agency of Korea, the total demand for urban land in 2030 has been extrapolated accordingly. It is assumed that approximately 17.5% of the SMA would be the urban built-up area by 2030 (See Table 5.3).

Thus with regard to the total amount urban growth, there is no difference between the scenarios because this growth is exogenously defined with reference to projected population in this simulation. As a result, the total amount of urban land conversion at the end of simulation year is same for the two scenarios. Although the number of urban patches and average patch size show a subtle difference between the scenarios, the urban area as a whole is almost identical across the scenarios. The SMA continues to grow and consume agricultural land as expected. Urban growth would occur in a dispersed way and continue to cause a loss of open space, and the number of urban patches and their degree of dispersion would also continue to increase.

Table 5.2. Calibration Results

			Neighbourhood Effect[a]		Random Coefficient	Accessibility Weight	
	Global Constraint						
Value	Year 2009	590275	Point 1	0, 10000	0.6	Highway	10, 0.25
			Point 2	1, 40		Main	10, 1
	Year 2030	670309	Point 3	2, 12		Minor	10, 1
			Point 4	8, 0		Local	10, 0.5

[a] *urban to urban interaction*

Table 5.3. Configuration of the Total Urban Built-up Area by Scenario, 2009-2030

LandscapeMetrics Scenario	Total Urban Cell Count	UrbanBuilt-upArea (km^2)	Percent	Number of Urban Patch	Mean Urban Patch Size (hectare)
Business as Usual	796002	1990.0	17.5	28970	6.87
Greenbelt Deregulation				28976	6.87

On the other hand, Figure 5.2 and Table 5.4 highlight varied spatial distributions of different urban development scenarios and compare varying degrees of sprawl. A comparison of total new urban growth is made at a distance between 0-50km from the centre of Seoul and this more clearly exposes the differences[8]. Having the same amount of total urban growth, the scenario 1 (Business as Usual) has lower amount of new growth within the 50km circle. Scenario 2 (Greenbelt Deregulation) shows more new development than the scenario 1 in the same range.

[8] The distance was measured from the location of Seoul City Hall, which is generally considered as the centre of Central Business District (CBD) in Seoul.

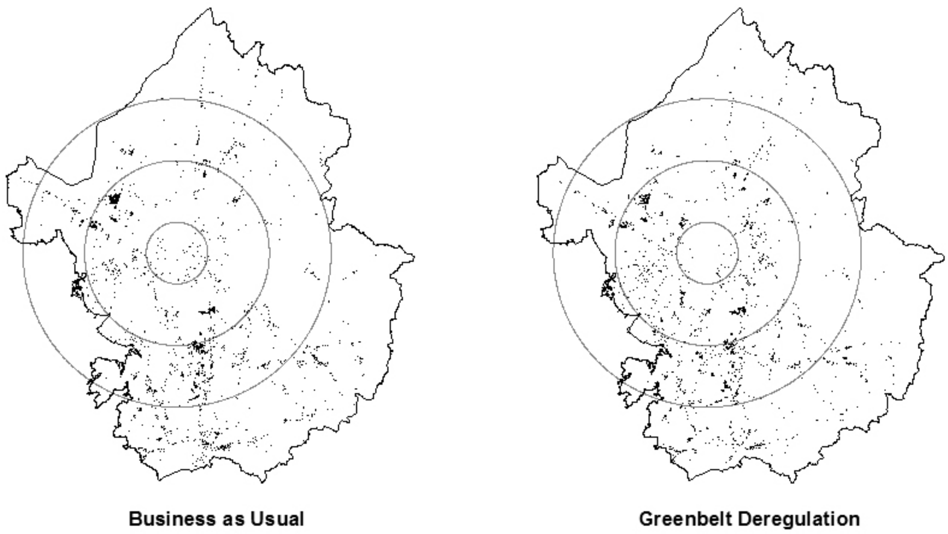


Figure 5.2.Comparison of New Urban Growth at 2030

Note: Above buffer rings are created at measured distances from Seoul City Hall. Each ring has a radius of 10km, 30km, and 50km respectively.

Table 5.4.Comparison of the New Urban Growth by Scenario & Distance, 2009-2030

Landscape Metrics / **Scenario**	**Within 10Km**		**Between 10-30Km**		**Between 30-50Km**		**Total(0-50Km)**	
	Cell Count	Area (km^2)	Cell Count	Area (km^2)	Cell Count	Area (km^2)	Cell Count	Area (km^2)
Business as Usual	1947	4.9	28900	72.3	36322	90.8	67169	167.9
Greenbelt Deregulation	1565	3.9	40057	100.1	32046	80.1	73668	184.2

Note: Areas are calculated from the cell count (Cell size = 50M x 50M). The figures present the different outcome states between scenarios but should not be regarded as an accurate prediction of future growth amount.

The result of these simulations reveals a paradox of greenbelt deregulation. As assumed and simulated, urban growth tends not to stop as long as the population and economy grow. Spontaneous growth without any further investment or regulation is likely to result in continuing leapfrog development affecting agricultural cities at further, more distant locations in the SMA. Deregulation of the greenbelt could prevent spontaneous growth in further parts of the SMA but it would harm previously protected areas near Seoul city.

6 DISCUSSION

6.1 Use of the Generic Model: Fulfilling Data Requirements

Generic urban models are pre-packed and ready to use for a wide variety of study areas without further development or modification. Fulfilling specified data requirements is a first step for successful implementation of generic models. In the use of dynamic CA models, such data requirements are imposed not only on the spatial dimension but also on the temporal one. CA urban models tend not to demand comprehensive spatial and/or aspatial data compared to the different styles of LUTI urban model. As seen in the **SLEUTH** and **Metronamica** models, such models can be run with less than several spatial input layers but generate future urban patterns even without the use of complex socio-economic data. Both **SLEUTH** and **Metronamica** do not require historic data for future simulations can be run from a single time point.

However, as a dynamic model, these models require historic spatial data to derive the best fit model parameters for future simulation and to ultimately bind the future simulation to the empirical ground. **SLEUTH** requires more intensive historic data than **Metronamica** does. However, this does not necessarily simply give a comparative advantage to a certain model. This is inherited from their different approaches to model calibration which will be discussed in the next section. Nonetheless, this study faced a major challenge with attaining historic spatial data, especially for the transportation network. Although dedicated custom data building was a possible option, this study relied on the data available from public sector which is the usual situation in reasonably well developed counties and those like Korea that have rapidly developed in recent years. This is necessary so that we can conduct urban simulation with the best available data.

Every model has a unique structure, hence it will have different data requirements. Whatever the requirements are, the model presents its own behaviour and outcomes based on such structure and requirements. Thus it is hard to evaluate a model simply with the data requirement. However, it is one thing that the urban modelling community should collectively think about. As Klosterman (2008) has pointed out, data available to planning practice tends to be inadequate but at the same time it is likely to be the best available data. Urban models should accommodate themselves to such conditions. In this way, urban models can be used not only by well-funded organisations but also by data-poor agencies and communities.

6.2 Calibration of the CA Model: Data Centred vs. Knowledge Oriented Approaches

Another key to the use of generic model is model calibration. Although model calibration is a necessity for any model if a practical application is aimed for, generic models usually have pre-defined calibration methods. This research has witnessed two types of model calibration: using the systematic quantitative method of the **SLEUTH** model and the qualitative approach of the **Metronamica** model. The former uses empirical data to derive the model parameters while the latter more relies on the area specific knowledge for model calibration.

The data oriented calibration method of **SLEUTH** enables a semi automatic calibration process. Although the determination of the best fit parameter set is ultimately made by the modeller, the model performs all the necessary computations and sums up the statistical results to compare simulation outcomes and actual data. As a result, **SLEUTH** requires multiple years of historic data for urban and transportation layers. At the same time, while a quantitative calibration method provides an objective measure to evaluate the goodness of fit of simulation results, the method is still limited in measuring the simulation outcomes at aggregate and global level. This means that the best parameter set determined by considering such measurements has a firm statistical representativeness. However, the simulated future from those parameters is an extension from the aggregated and averaged model outcomes, not from local peculiarities. Unfortunately, quantitative individual cell level comparison is not well developed in this field.

One the other hand, the calibration of **Metronamica** relies more on the study area specific knowledge than on the data itself. Although repetitive visual comparisons are necessary, this enables the modeller to conduct in-depth investigation of local patterns. Such characteristics of model calibration are basically due to a complex nature of the model structure and the difficulty of estimating different strata of parameter values from the single observed data of the land use map. An automatic extraction has been attempted (Straatman et al., 2004), but so far no single method has replaced a knowledge oriented calibration process specific to the **Metronamica** model. The pitfall of subjectivity of course does exist in this process. However, this does not necessarily mean such calibrations and hence the simulation results are unreliable. Compared to an automatic calibration based on statistical techniques, qualitative calibration has clear merit in bringing knowledge on spatial forms and pattern specific to the modelled area. Indeed this is the *de facto* method that enables the close examination of the local level patterns since no quantitative metrics can yet fully replace such a method.

The quantitative estimation of model parameters from data is more general and common to scientific models. However, it should be noted that such effort is bounded by the availability of data as well as the quality of data. On the other hand, knowledge oriented methods are less dependent on the data at least in the temporal dimension as seen in these simulations. While this can be a weakness, this would give a model a comparative advantage in supporting planning policies in data-poor conditions.

6.3 Beyond Behavioural Realism

One of the main strengths of CA is simplicity in model development. As demonstrated in Figures 2.1 to 2.3, a number of simple rules can generate certain urban growth patterns at a global scale. Since such rules are typically constructed on an *ad hoc* basis, model building is possible without the use of tested theory. Established generic models such as **SLEUTH** and **Metronamica** add more diverse elements to reproduce real urban systems, but they also greatly rely on such *ad hoc* model development strategy.

Not confined to the established available theories, CA urban models have answered what theory based models could not answer – realistic reproduction of urban systems without reliance on unrealistic assumptions. Thus the main trends of CA urban models have been centred on the pursuit of behavioural realism. The transition rules which form the core of such models mimic the behaviour of real urban systems based on an intuitive understanding of such systems. Then the models are calibrated over the observed land use data and used for practical applications, but the transition rules which generated the simulation results tend not to be examined further Moreover, the use of the random algorithm is almost essential to maximise such realism as we have seen the simulations in this research. As a result, although CA urban models have been successfully applied to the study of complex urban systems, they lack explanatory power and have not yet effectively yielded theories about how urban systems evolve.

Recently emerging research efforts to infuse a more rigid explanation of urban systems into CA or ABM (agent based model) is slowly pointing to a new synthesis in urban modelling (Brown and Robinson, 2006, Caruso et al., 2007, Filatova et al., 2009). Such approaches usually introduce micro economic theories to define cell transition rules and/or agent behaviour. This type of approach is not yet fully developed, but this new trend implies a need for theory-oriented, disaggregate, and dynamic urban models. In this way, CA urban models could provide much more realistic behavioural simulations of how urban structures emerge, evolve and regenerate themselves in such a way that they are useful and informative for enhancing policy support.

REFERENCES

BATTY, M. (2004) Dissecting the streams of planning history: Technology versus policy through models. *Environment and Planning B: Planning and Design,* 31**,** 326-330.

BATTY, M. (2005) *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals,* Cambridge, Massachusetts, The MIT Press.

BATTY, M. (2009) Urban modelling. In KITCHIN, R. & THRIFT, N. (Eds.) *International Encyclopaedia of Human Geography.* Amsterdam, Elsevier Science.

BATTY, M., and LONGLEY, P. A. (1994) *Fractal Cities: A Geometry of Form and Function*, San Diego and London, Academic Press, available online at www.fractalcities.org

BROWN, D. G. & ROBINSON, D. T. (2006) Effects of heterogeneity in residential preferences on an agent-based model of urban sprawl. *Ecology and Society,* 11.

CARUSO, G., PEETERS, D., CAVAILH, J. & ROUNSEVELL, M. (2007) Spatial configurations in a periurban city. A cellular automata-based microeconomic model. *Regional Science and Urban Economics,* 37**,** 542-567.

CLARKE, K. (2008) A Decade of cellular urban modelling with **SLEUTH**: Unresolved issues and problems. In BRAIL, R. K. (Ed.) *Planning Support Systems for Cities and Regions.* Cambridge, Massachusetts, Lincoln Institute of Land Policy.

CLARKE, K., HOPPEN, S. & GAYDOS, L. (1997) A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B: Planning and Design,* 24**,** 247-261.

COUCLELIS, H. (1985) Cellular worlds: a framework for modelling micro-macro dynamics. *Environment and Planning A,* 17**,** 585-596.

DIETZEL, C. & CLARKE, K. (2007) Toward optimal calibration of the **SLEUTH** land use change model. *Transactions in GIS,* 11**,** 29-45.

FILATOVA, T., PARKER, D. & VEEN, A. V. D. (2009) Agent-based urban land markets: Agent's pricing behaviour, land prices and urban land use change. *Journal of Artificial Societies and Social Simulation,* 12.

IACONO, M., LEVINSON, D. & EL-GENEIDY, A. (2008) Models of transportation and land use change: A guide to the territory. *Journal of Planning Literature,* 22**,** 323-340.

JANTZ, C. A., GOETZ, S. J., DONATO, D. & CLAGGETT, P. (2010) Designing and implementing a regional urban modelling system using the **SLEUTH** cellular urban model. *Computers, Environment and Urban Systems,* 34**,** 1-16.

KLOSTERMAN, R. (2008) A new tool for a new planning: The What if?™ planning support system. In BRAIL, R. K. (Ed.) *Planning Support Systems for Cities and Regions.* Cambridge, Massachusetts, Lincoln Institute of Land Policy.

RIKS (2011) **Metronamica**, http://www.metronamica.nl/ Research Institute for Knowledge Systems, P.O. Box 463, 6200 AL Maastricht, The Netherlands

STRAATMAN, B., WHITE, R. & ENGELEN, G. (2004) Towards an automatic calibration procedure for constrained cellular automata. *Computers, Environment and Urban Systems,* 28**,** 149-170.

TOBLER, W. R. (1970) A computer movie simulating population growth in the Detroit region. *Economic Geography*, 42, 234-240.

TOBLER, W. (1979) Cellular Geography. In GALE, S. & OLSSEN, G. (Eds.) *Philosophy in Geography.* Dordrecht, Reidel.

WHITE, R. & ENGELEN, G. (1993) Cellular automata and fractal urban form. *Environment and Planning A,* 25**,** 1175-1199.

WHITE, R. & ENGELEN, G. (1997) Cellular automata as the basis of integrated dynamic regional modelling. *Environment and Planning B: Planning and Design,* 24**,** 235-246.

WHITE, R. & ENGELEN, G. (2003) Kalibratieprocdure: A calibration procedure for constrained large neighbourhood cellular automata based land use models. *13th European Colloquium on Theoretical and Quantitative Geography.* Lucca, Italy,

WHITE, R., ENGELEN, G. & ULJEE, I. (1997) The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. *Environment and Planning B: Planning and Design,* 24**,** 323-343.