# IMPROVING FORECASTS OF LAND USE WITH REGIONALIZED MAPS IN THE SLEUTH MODEL

ELLEN CRISTINA WOLF ROTH[1], ANDRÉ KOSCIANSKI [2]
Universidade Tecnológica Federal do Paraná (UTFPR).
Rua Dr Washington S Chueire, 330 - CEP 84017-220 - Ponta Grossa - PR - Brasil
[1] ellenwolf1@hotmail.com, [2] koscianski@utfpr.edu.br

ABSTRACT

The expansion of cities has significant impacts on economy, ecology, and quality of life, among other sectors. This process affects a large fraction of the world's population and draws attention of administrators, investors, and scientists. Simulation is an important tool to understand and administer the growth process; however, cities are complex systems, and computer models capture only a fraction of their dynamics. Precision is compromised by simplifications as data averages, and the difficulty to represent human aspects decisive in the evolution of cities. This study tries to mitigate these issues by integrating qualitative information in forecasts computed with the model SLEUTH. Simulations were regionalised using a socioeconomic and historical perspective, which can be explored with other tools. The method was compared to the traditional approach, and the results confirmed a better match between the simulation and the city characteristics.

Keywords: land use and cover change; computational simulation; regionalized simulation; historical and sociological characteristics; SLEUTH model.

MEJORA DE LA PREDICCIÓN DE LOS USOS DEL SUELO CON MAPAS REGIONALIZADOS EN EL MODELO SLEUTH

RESUMEN

El crecimiento de las ciudades tiene impactos en la economía, la ecología y la calidad de vida, entre otros aspectos, afectando a una gran fracción de la población mundial. La simulación es una herramienta importante para comprender y gestionar este proceso; sin embargo, las ciudades son sistemas complejos y los modelos computacionales capturan solo una fracción de su dinámica. La precisión se ve afectada por la simplificación de los datos, y la dificultad de representar aspectos humanos decisivos en la evolución de las ciudades. Este estudio intenta mitigar estos problemas integrando información cualitativa en la predicción calculada con el modelo SLEUTH.

Las simulaciones, que pueden ser exploradas con otras herramientas, se regionalizaron utilizando una perspectiva socioeconómica e histórica. El método se comparó con el enfoque tradicional, y los resultados confirmaron una mejor coincidencia entre la simulación y las características de la ciudad.

Palabras clave: cambio de usos y cubiertas del suelo; simulación por ordenador; simulación regionalizada; características históricas y sociológicas; modelo SLEUTH.

## 1. Introduction

Cities are the stage of the life of most of the world's population, reaching proportions greater than 80 % in some countries (Rodríguez and Meneses 2011, Ratcliffe *et al.* 2016, United Nations 2018). The interactions between a city and its surroundings include a myriad of processes, as flows of materials, gas emissions, and waste; production of noise and light pollution; modifications to the landscape exemplified by water streams buried under streets, creation of parks and artificial lakes. Cities attract the installation of industries and spawn new activities, leading to self-sustaining cycles and the transformations of increasingly large areas. This ensemble of processes affects the environment and people, and creates challenging problems for public administration and private initiatives. The planning of investments and actions in this context entails scenarios with conflicting objectives regarding factors as ecology, economy, and quality of life. Besides the issue of conflicting objectives, a major obstacle complicates the task of administering urban expansion: the difficulty of calculating forecasts.

The interactions between human activities and land engender a complex system, with dynamics that are difficult to characterize. This notion is known in all fields of science, with examples as the functioning of live organisms in Biology, the web of effects linking micro and macro economic systems, or coupling between thermic, hydraulic and mechanical phenomena in systems studied in Physics. Urban sprawl is influenced by very different factors, such as the distribution of roads, local topography, social history, and economic activities. If taken individually, those different elements encompass phenomena and causal relations that are reasonably understood (Santé *et al.* 2010, Triantakonstantis and Mountrakis 2012, Brown *et al.* 2012). However, the exact evolution of a city is a result of the composition of all those socioeconomic and environmental aspects and form systems that are unique (Rocha 2012). Because of its practical importance, the study of methods to compute forecasts of land use attracted increasing interest in recent years. This field of study makes use of computer tools to process large amounts of data.

Computer models entered practical use during the Second World War, with the development of the Monte-Carlo simulation technique as a major example. Initially explored by their capability to do repeated tasks and calculations, computers and programming languages rapidly evolved to go beyond pure analytical models as those expressed with differential equations, providing scientists with the possibility of describing natural processes in the form of algorithms. This ability is explored nowadays to represent the functioning of cities and their internal components. However, as the representations get more complex and capture more details, it becomes increasingly difficult to assess the quality of a given model. This constitutes a classic example of a wicked problem (Ritchey 2013). In other terms, the choice of the mechanisms to be

modeled, the level of precision and other requirements cannot be set in advance, as the consequences of these decisions on the results are hard to foresee (Rae and Wong 2012).

Models of land use must draw a compromise between the flexibility needed to represent a rich set of situations, and the accuracy expected to reproduce a given scenario, and which may involve specific requirements (Sohl and Claggett 2013, Wahyudi 2016). Choosing a minimal set of data and parameters that can reliably provide answers is a difficult task, which gives rise to different models and a history of research that strives to improve understanding of the problem (Torrens and O'Sullivan 2001, Rocha 2012). In the context of city dynamics, the design of a model is based on some general guidelines that provide a safe starting ground. For instance, spatial correlation is a ubiquitous factor to weight variables and processes; it can be exemplified by the gravity metaphor (Fotheringham 1985), or the definition of attraction and repulsion forces that drive the choice for future land use (Aguilera *et al.* 2011). Clusters of activity, especially reflecting the use of land for economic activities (industry and commerce) are influenced by transportation costs (Fujita and Thisse 1996); in fact, distance to roads and to the city center is a variable frequently found in urban models (Santé *et al.*. 2010). Information about topography is a basic element, with special attention to slope (Clarke and Gaydos 1998). Finally, some models try to represent patterns of human behavior that influence city dynamics; a technique suited to this end is computational agents (Parker *et al.* 2003). However, representing fine details as individual market operations or entrepreneurial initiatives poses serious difficulties concerning model construction and validation, because of human unpredictability and lack of data (Crooks *et al.* 2008, Triantakonstantis and Mountrakis 2012). Integrating qualitative information and rethinking the role of specialist knowledge in a simulation are some alternatives to escape the limitations of a purely mechanical approach (Houet *et al.* 2016, Kok and Verburg 2007).

The different techniques used to implement computer simulators of Land Use and Cover Change (LUCC) affect the translation from world views into code in several ways, but a common trait is a difficulty to relate processes that occur at different spatial and time scales (Ménard and Marceau 2005, Brown *et al.* 2012). When a model averages behaviors over large areas or long periods, it loses local details that may have special interest to understand the processes occurring in different parts of the same city (Jantz *et al.* 2010). This issue can be seen as a corollary of an observation published half a century ago, which became known as Tobler's first law of Geography: "everything is related to everything else, but near things are more related than distant things." (Tobler 1970). Another limitation present in analytic and algorithmic techniques is the difficulty to integrate abstract, qualitative information into a model; this problem has been called the difference between "narrative and number" or "story and simulation" (Kemp-Benedict 2004, Alcamo 2008). This suggests the integration of the two views: using the strength of computer models to explore the mechanical aspects of the system, and expert intuition to guide the preparation of data and filter and interpret the results.

The objective of the present study was to assess modifications on computations of the evolution of land use, produced by combining the capability of the algorithmic model SLEUTH with qualitative information not represented in computer code nor captured by statistics and databases. Two methods to segment the area of study were compared with calculations for complete maps. The city chosen for the study has historical and physical aspects that arguably make simulation harder. The hypothesis was that the regional approach would have an impact on the forecasts, possibly revealing a more detailed picture.

## 2. Background

One of the first known studies about city structure was conceived in 1826 by the German economist Von Thünen, drawing relations between the price of goods, distance to market, and price of land (Roth 2019). It was the seed for several models describing city dynamics, which could incorporate higher levels of detail with the advent of computers, more than a century later. Simulations of traffic and other urban processes gained impulse in North America, following the rapid rise of the automobile industry around the 1970 decade (Batty 1976). Approximately, in the same period, Geographical Information Systems started to define an area of research on its own (Goodchild 2018). As new computational techniques evolved, they could be explored in combination with analytic methods. The following examples compare some of the most notable approaches and highlight differences.

Markov Chains assimilate patterns of transitions using probability matrices and have been implemented with Artificial Intelligence techniques as Neural Networks (Aaviksoo 1995, Wang and Li 2011). Simulators based on stochastic models may operate without inner knowledge about a system; in other words, causality chains are not implemented in the code, which reproduces phenomena following a black-box paradigm (Rivals and Personnaz 1996). By contrast, Econometric Models and System Dynamics are two examples of techniques that decompose the phenomenon of city growth in a set of processes, adjusting hypotheses and parameters in order to fit reality (Sanders and Sanders 2004, Haase and Schwarz 2009). Such analytic tools benefit from a strong mathematical background but generally do not incorporate information as local topography. As a result, they can compute total projections such as the number of new houses, but cannot represent finer details as the spatial patterns produced during growth. On the other extreme, there are Agent Based Models (ABM), which can represent entities and their interactions with a high level of detail using computer code (Macal and North 2010, Crooks and Heppenstall 2012), but it is difficult to assert that the right aspects of the system were modeled (Filatova and Verburg 2013).

Cellular Automata (CA) is one of the most used solutions in LUCC simulation (Triantakonstantis and Mountrakis 2012). The structure of CA is conceptually simple but can accommodate complex models, with characteristics as non-linearity and emergent processes and properties, or even chaotic behavior (Wolfram 1984, Rocha 2012). The next section makes an overview of Cellular Automata and the simulator SLEUTH used in the present study.

2.1. Cellular Automata and Urban Simulation

Models built with Cellular Automata are composed of a matrix of individual elements, which is also typical of Geographical Information Systems (Creighton *et al.* 1959). Each cell operates independently and in parallel, having its own attributes and internal state. Each cell evolves in discrete time steps according to rules that may depend on the state of neighbors, parameters and other input data (Bandini *et al.* 2001). This organization makes it equally possible to represent the local characteristics of a neighborhood, and the global effects and trends of an entire city. Some simulators combine CA with different techniques as Markov Chains (Arsanjani *et al.* 2013, Rimal *et al.* 2018), Neural Networks (Basse *et al.* 2014, Ozturk 2015) or Agent-Based Models (Loibl and Toetzer 2003, Mustafa *et al.* 2017, Clarke 2018).

geofocus

Revista Internacional de Ciencia y Tecnología de la Información Geográfica
International Review of Geographical Information Science and Technology

Grids of cellular automata are surprisingly powerful devices; a simple matrix of size $10 \times 10$ containing Boolean cells can represent $2^{100}$ states, an astronomical quantity. While flexibility is a desirable feature for any model, it also makes it harder to determine system trajectories. In the case of urban sprawl, constraining the search space can help to keep the system under control (Rocha 2012). Partitioning the map reduces the number of variables and increases the possibility of the model adapting to regional characteristics (Shen *et al.* 2009, Liu *et al.* 2012). A recent study by Dutta and Das (2019) exemplifies the fact that spatial metrics and growth patterns may present important differences across areas inside the same agglomeration.

2.2 The SLEUTH model

SLEUTH is one of the most well known Cellular Automata (CA) models in the field, with two decades of experiments documented in the literature (Clarke and Gaydos 1998, Clarke 2018). Its name is an acronym of the input layers used by the software: Slope, Land use, Excluded areas, Urbanization, Transportation and Hillshade. Data are represented as images, and the model requires four types of information:

- Slope, corresponding the topography of the region, derived from a Digital Elevation Model (DEM).

- Excluded areas, which indicate regions as lakes, swamps or parks, that will not be converted to urban usage.

- Road network, representing roads that are significant for the future process of city growth; and

- Urban footprint, with a map that distinguishes different land uses.

In its basic operation mode, the simulator classifies land into two states: cells are either urban or not. The software also allows users to define an arbitrary set of categories of land use, as 'industrial', 'commerce' and 'residential area'; if this feature is activated then a second mechanism is added to the execution, based on computational agents. In the present study, it was not enabled.

Instead of trying to represent internal processes and detailed decision mechanisms that occur in a city, as it happens in some urban sprawl models, SLEUTH took a more cautious approach and defined four types of growth that should capture land-use changes over time. They are combined in the CA grid by means of the sequential execution of rules, which define four growth types illustrated in Figure 1.
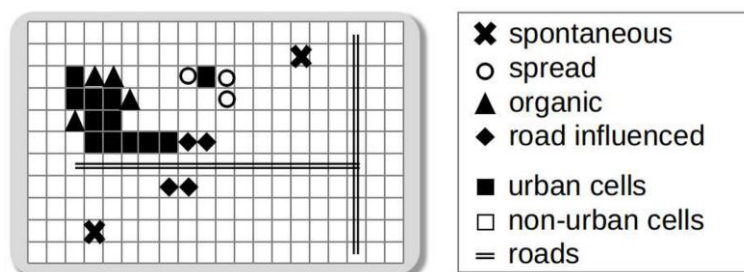


**Figure 1. Types of growth simulated by SLEUTH.**

Source: the authors.

Spontaneous growth corresponds to the random conversion of cells, from unused to urbanized. In the process of Spread, new occupied areas stem from isolated urban cells. Organic growth is a version of Spread where the seed is a group of at least three cells. Different from the previous processes that create isolated fragments, organic growth corresponds to the sprawl of existing urban areas. The last process is the Road Influenced growth, which spawns new urban cells located along roads. All these rules are activated in a non-deterministic fashion, meaning that there is a degree of randomness implemented in the algorithms. Simulation results are averaged over several runs using a Monte-Carlo approach.

Execution of the software is controlled by five global parameters, listed on Table 1.

**Table 1. SLEUTH parameters.**

| Parameter | Meaning |
|---|---|
| Diffusion | Controls generation of new cells scattered on the matrix and along roads |
| Bread | Adjusts the probability of a new urban fragment being expanded |
| Spread | Controls the organic process of expansion |
| Slope | Adjusts overall sensitivity of the model to steep terrains |
| Road Gravity | Affects how roads pull urbanization |

Source: the authors, from software documentation.

Calibration is the first execution mode of Sleuth. Adjusting parameters is a crucial task in the application of a model, being affected by the quality of information available and the size of the search space (Clarke 2004, Amujal 2015, Chen *et al.* 2017). This process can combine algorithmic procedures with expert knowledge, in order to tune configurations and steer the simulation trajectory or help select the most probable scenarios (Volkery *et al.* 2008, Houet *et al.* 2016, Gounaridis *et al.* 2019). In SLEUTH this task is performed iteratively by the user, based on logs of execution that collect several statistics. The metric chosen to guide the calibration, in most studies using SLEUTH, is Lee-Sallee (Lee and Sallee 1970). It compares two binary images: the output of a run of the simulator and an actual map. It is calculated as a ratio A/B, where A is the count of urban pixels present at the same coordinates (x,y) on both images (a Boolean intersection); and B is the count of urban pixels on both images, minus A (a Boolean union). The higher the value of the metric, the closer the agreement between the images.

The SLEUTH model has a global character, where every cell follows the same algorithms that are configured by the set of five parameters. This strategy is not unusual and can be found in many simulators; however, it may limit precision since local characteristics tend to be averaged and diluted by statistics and algorithms dealing with large areas (Shen *et al.* 2009, Liu *et al.* 2012, Moghadam *et al.* 2018). Some general approaches to overcome this issue consists of using patches in place of cells (Li *et al.* 2017, Chen *et al.* 2017, Moghadam *et al.* 2018), incorporating qualitative input directly into the model (Houet *et al.* 2016), or using a territorial division guided by metrics (Kazemzadeh-Zow *et al.* 2017) or cluster analysis (Jantz *et al.* 2010). There are some studies for the specific case of SLEUTH. Houet *et al.* (2016) modified the software adding a layer of qualitative information. In the study of Jantz *et al.* (2010) maps were divided using an automatic procedure, based on k-means cluster analysis performed with eleven variables. Mahiny and Clark (2012) computed an urbanization suitability layer using a mathematical analysis of fifteen factors. The

present study adopts the view of combining context analysis and algorithmic model but doing this outside of the software (see for instance Kemp-Benedict (2004)).

## 3. Materials and Methods

3.1 Study area

Ponta Grossa is located in the south of Brazil, in the Paraná State, a federative unit of the country. It stands in a region of irregular topography, near hill ranges that separate the Atlantic coast from inland; its approximate location is shown in Figure 2. It occupies approximately 2054 km$^2$ and the last estimates for 2019 indicated a population of 351 736 inhabitants, a population density of 170 hab/km$^2$.
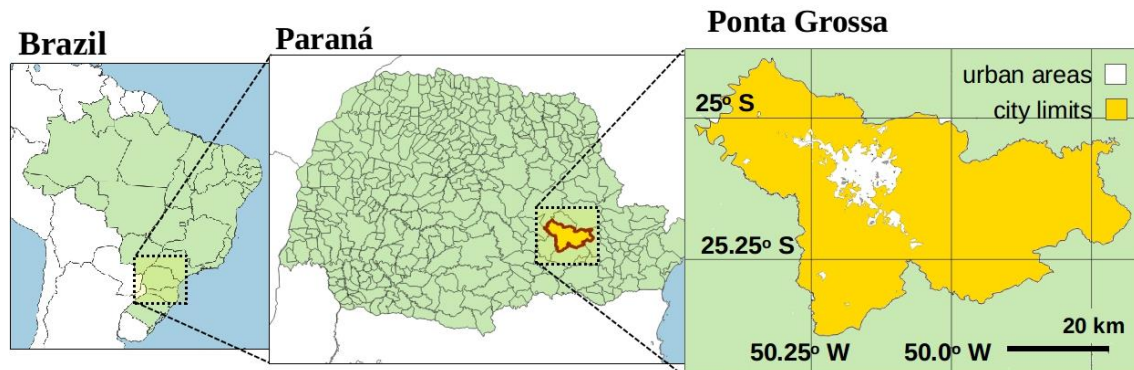


**Figure 2. Localization of Ponta Grossa.**
Source: the authors.

The initial settlement was established by herdsmen that crossed the region; records from the beginning of the XIX century estimate a population of 1330 citizens and 331 slaves in 1823 (Gonçalves and Pinto 1983). Incentives from the brazilian government attracted European settlers, and in 1887 more than 2000 immigrants were received in the region (Kubaski 2015). The arrival of a railroad in 1894 and the strategical position of Ponta Grossa at crossroads linking important cities were motors for its development along of the XIX century. Its commercial projection declined during the XX century with the advent of coffee exploitations in other cities, and subsequently the growth of Curitiba, the capital of Paraná State located between Ponta Grossa and the coast. The evolution of the population is shown in Figure 3, along with an approximation of the urban area taken from the images used in the study. A slight decrease in the speed of city expansion occurs around 2005-2008 and 2014-2017. Population growth exhibits a similar trend and overall slowed down in the last decade (IBGE 2019).
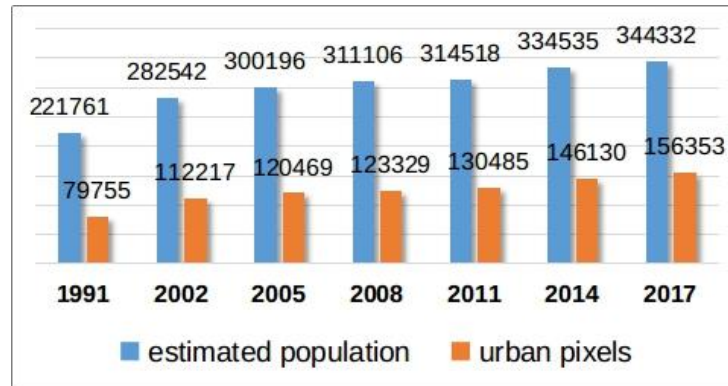
**Figure 3. Population statistics and count of urban pixels.**
Source: count of pixels extracted from input maps; population from IBGE (2019).

Ponta Grossa is the economic center of a region known as 'Campos Gerais' composed of 13 cities and approximately 760 000 inhabitants. The Gross Domestic Product (GDP) per capita of the city was approximately R$ 42 000.00 in 2017 (US$ 10 200.00, with values from January/2020), placing the city at position 530 out of 5570 localities in Brazil (IBGE 2019). The service sector and agroindustry are the two most important sources of revenue (Ponta Grossa 2018). The value of the GDP Human Development Index (HDI) assessed in 2010 was 0.763, corresponding to the position 320 in the country and qualitatively classified as high (IBGE 2019). Despite the favorable IDH indicator, important social disparities still remain. The value of the GINI index calculated for 2010 is 0.54, with 21 % of the population having a salary below the minimum reference value and half of the young adults not having completed secondary education. In 2018, more than 6000 irregular houses (non-authorized by the municipal government) were identified; the poorest ones were built on valleys bottoms (Ponta Grossa 2018).

Compared to other cities of Paraná, Ponta Grossa has a relatively complex topography and the territory was occupied over the years in a rather disorganized fashion, without effective urban guidelines (Silva 2013). The lack of infrastructures as viaducts frequently requires people to cover relatively long paths to join destinations that are not far at a straight distance. This also affects users of public transports, sometimes obliged to change buses up to three times (economic interests probably play also a role in this situation). These characteristics create a degree of isolation between regions, interfering with decisions regarding the choice of residence, school, work, and leisure activities.

Cities in Brazil are traditionally divided into neighborhoods known as '*bairros*'. These divisions are related to the Portuguese *freguesias* but do not have local administration. They are considered nonetheless in official statistics (health, security), are used in public government decisions (placement of schools, transportation, health centres), and are recognized by postal codes. Such neighborhoods emerge with the development of a city and are shaped by historical factors. For instance, in Ponta Grossa the neighbourhood known as 'Oficinas' (Portuguese for 'machine shop') concentrated services for the railroad that crosses the city, 'Olarias' had brick factories operating coal furnaces, while 'Nova Russia' was the home region for many immigrants from East Europe. Nowadays, Oficinas still concentrates commercial activities such as auto part shops, car dealerships and auto repair shops. Olarias, a neighborhood not especially sought for residences in the past,

nowadays represents a mixed market, with gated communities but also areas with very modest houses. Finally, 'Jardim América' is known for its concentration of high-valued properties.

3.2 Data gathering and pre-processing

Information had to be gathered from several sources to recover historical data. This includes IPLAN, a division of City Hall responsible for Urban Planning; IBGE, Instituto Brasileiro de Geografia e Estatística; satellite images from INPE, Instituto Nacional de Pesquisas Espaciais; the Aster Global DEM dataset, from NASA/METI; and Google Earth.

Classification of land use from remote sensing data is generally performed by applying automated procedures (Berberoglu and Akin 2009, Rozenstein and Karnielli 2011). These processes are harmed by obstacles as noise, insufficient or low-quality data, and introduction of errors with procedures as resampling and image correction (Manandhar et al 2009). In this study, city limits were visually identified in satellite photos and the corresponding urban footprints were manually painted. A sequence of black and white images was produced, spanning the period 1984 to 2017 in intervals of 3 years. Most studies presented in the literature employ larger intervals, but the use of a denser dataset was expected to leave fewer gaps to be guessed by the simulator, and SLEUTH documentation does not put a limit on the sampling rate.

The exclusion layer was obtained from IPLAN; it identifies preservation areas as a State Park (Vila Velha), which must be ignored by the growth processes implemented in the simulator. A roadmap for 2017 was extracted from the OpenStreetView database with the software Quantum GIS. Layers for precedent maps of 1984 and 1996 were created by hand, by inspecting satellite images and removing major roads that were not present at those years. Using a color-code defined by SLEUTH, roads connecting adjacent cities were marked with the value 100, main avenues of the city with value 50, and other important streets with the value 25. Slope and hillshade were obtained by processing the Data Elevation Model with Quantum GIS.

The final images used in the study had 1242 x 1339 pixels, corresponding to a scale of 30 meters and the same resolution was used in all the input layers. SLEUTH does not impose a minimal resolution, and the value chosen coincide with many studies reported in the literature (Triantakonstantis and Mountrakis 2012). A group of 3x3 cells in the present case is roughly equivalent to one city block. Examples of three city footprints and other images used as input to the software are shown in Figure 4.
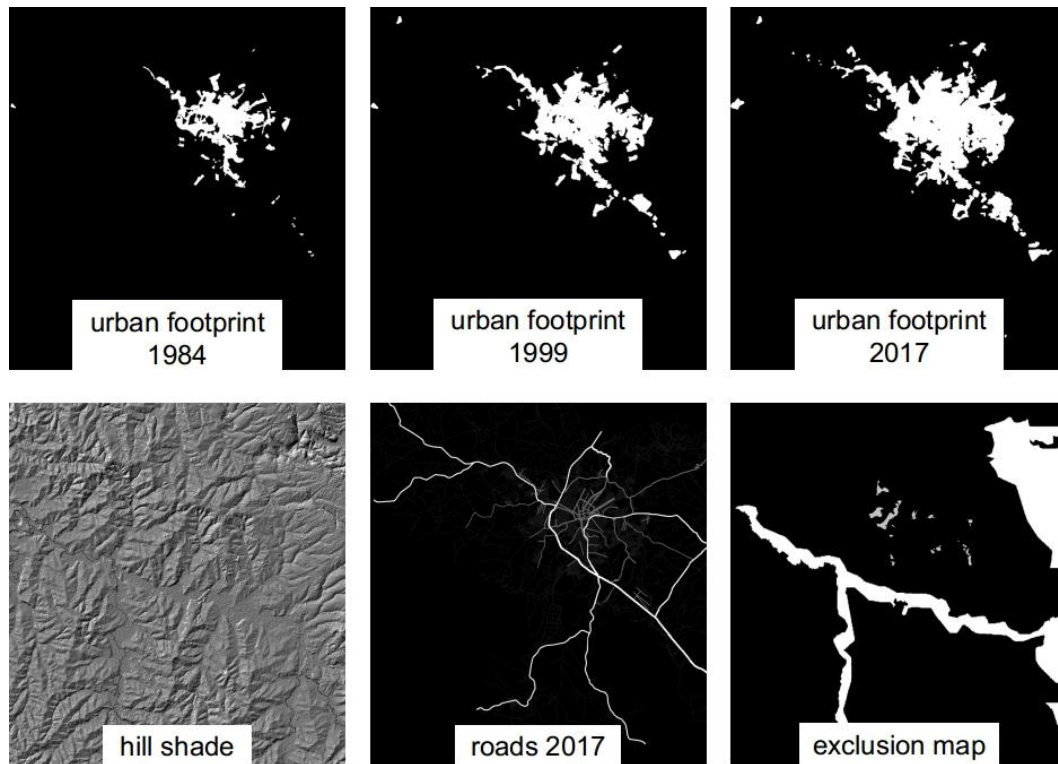
**Figure 4. Some Input Layers used in the study.**
Source: the authors.

3.3 Design of the experiments

Ponta Grossa has 14 neighborhoods (*bairros*), a number judged excessive to divide the map; the study sought to verify how local tendencies would be captured by the computational model, and not get to the extreme of simulating ensembles of city blocks. Thus, some criteria were initially defined in order to guide the partitioning. First, the division would have a low number of segments, of the order of four. A highway crossing the city and important avenues that function as axis that structure the urban morphology served as an initial reference. The combination of the transportation network and geographical features as valleys acts as barriers to locomotion, creating patterns that would not exist if the city had a flat topography. For instance, the tendency to avoid living far from work is stronger in Ponta Grossa than in Maringá, a city in the State of Paraná with a similar population and economy. Finally, banks, stores, pharmacies, and other amenities occur replicated in areas across the city of Ponta Grossa, allowing them to function more or less independently. These considerations guided the analysis and the city was divided into three regions of approximately equal occupation, shown in Figure 5.
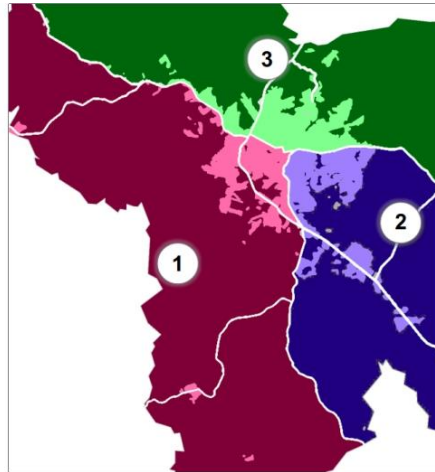
**Figure 5. City map subdivisions.**
Source: the authors.

Region 2 shown in Figure 5 is markedly industrial. It encompasses the neighborhoods of Oficinas and Olarias, and the largest industrial zone of the city. Regions 1 and 3 are more diversified, with commerce, residential areas, two public universities and one of the factories of BRF S.A., a Brazilian food producer among the greatest in the World. Sectors of service and commerce concentrate on the junction of the three regions. Maps prepared for the simulations included an overlap between regions, in order to permit interactions to occur from one partition to another. Three simulation scenarios were defined:

- C (complete): the maps encompass the whole city.
- S (sliced): layers were edited to remove parts of the city.
- E (excluded): the exclusion layer was used to mask removed regions.

The scenarios S and E, although in principle thought to be equivalent, were designed to confirm simulation results for each region using two experiments. The first phase of each experiment is the calibration to obtain the five model coefficients. These values have been dubbed as SLEUTH DNA (Gazulis and Clarke 2006) and they describe, under the model perspective, the major processes that drive the growth of a city. For instance, differences in concentration of commerce and residential settings might correspond to different 'DNA' types under the simulated perspective.

Experiments were organized around two series, as shown in Figure 6. Using data up to 2014 a forecast was computed for 2017, allowing the validation of the experiments with available information. This was performed by comparing growth maps and actual images regarding two aspects: the distribution of new pixels across the output map should match areas that developed in Ponta Grossa; and the count of urban pixels in the projections should approximate the new area. The second test used the complete time series to obtain a forecast for 2020. The calculations were performed with the seven different setups (C, S1..3, E1..3), for a total of fourteen sets of results. Each of these sets involved four rounds of calibration, following the procedure recommended in SLEUTH documentation. In the first step, known as 'coarse', images had one-quarter of the full size. This was followed by the 'fine' calibration, with input files half the full size, using parameters

chosen from the previous step according to the Lee Sallee metric. In the third phase, images had full size (1242 × 1339 pixels). In a final process, named 'forecasting calibration', SLEUTH applies heuristics to modify the coefficients, responding to periods of sudden growth or stagnation. Each experiment is executed by the software several times, using an internal Monte Carlo algorithm to average values and compute, for each pixel, the probability of state change.
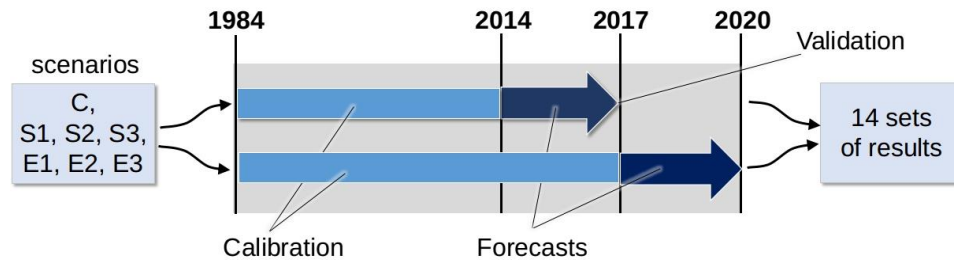


**Figure 6. Organization of the experiments.**
Source: the authors.

Source code was downloaded and compiled from the project site (National Center for Geographic Information and Analysis, USA) and, during the first runs, an internal error halted the experiments. The overflow of a variable was detected thanks to memory checks enforced by the compiler (gcc release 7) in approximately 20 000 lines of code. The error affects a memory area known as the top of stack and should cause no consequences for users of old compilers.

## 4. Results

4.1 The calibration phase

The coefficients obtained with the calibration processes provide a general picture of how the model tries to assimilate the city of Ponta Grossa. Values from the final phase of the calibration are shown in Figure 7. The letters indicate the simulation scenario (complete, sliced, excluded) and numbers identify each of the 3 regions.
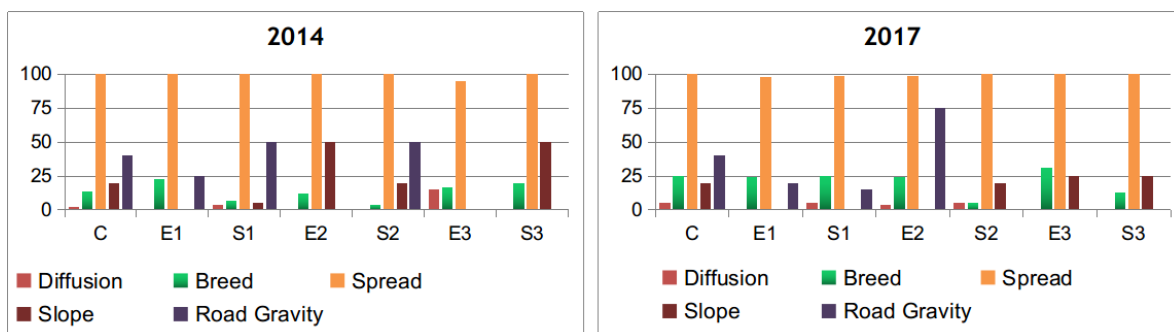


**Figure 7. Parameters obtained in Calibration.**
Source: the authors, from simulation results.

Values for the parameter Diffusion were very low and suggest that the emergence of new urban cells rarely occurs far from the city center. Breed, associated with the apparition of new urban 'islands' showed the same tendency. Spread, associated with the so called organic growth, had consistently high values; it was the most significant coefficient in the experiments. This trend represents the fact that the boundaries of the city advance over empty areas and represent most of the new urban blocks. The coefficient Slope presented mixed results; the values found during calibration do not follow any clear pattern; in scenario C, terrain topology showed little impact on results; in S1, this aspect was even less significant, with values near 0. By contrast, in S2 Slope presented high values, while in E2 - the same geographical region - results were incongruous: higher in 2014 and lower in 2017. S3 and E3 also showed erratic changes.

In principle, these results might be an indication of the model struggling with the erratic occupation of land. Compared to economic forces, topography played a minor role in the determination of city sprawl. Historically, patches of flat terrain were prioritized for agriculture, and, as a matter of fact, the city center lies in a region of irregular topography. Figure 8 illustrates part of the intricate boundaries between urban and rural areas in the city, still present today.



**Figure 8. Agricultural land around urban areas in Ponta Grossa.**
Source: Google Earth.

The last coefficient, Road Gravity, did not have a significant influence in most of the experiments, with the notable exception of scenarios S2 and E2, where it attained values above 50. This region developed along a boulevard that leaves the city towards Curitiba, the capital of the State of Paraná. The State Highway was a natural choice for the construction of factories and warehouses, and in the 1980 decade, the municipality defined an Industrial Sector in the region, with tax incentives. The values for all the coefficients are obtained and refined in three rounds of calibration, following the instructions in the documentation of the software. At this point, it is interesting to observe the LeeSalle metric, as it is the criteria guiding the selection of values and shows the overall precision of the simulation before the forecasts are calculated. Plots of the values of the metric along the process are shown in Figure 9.
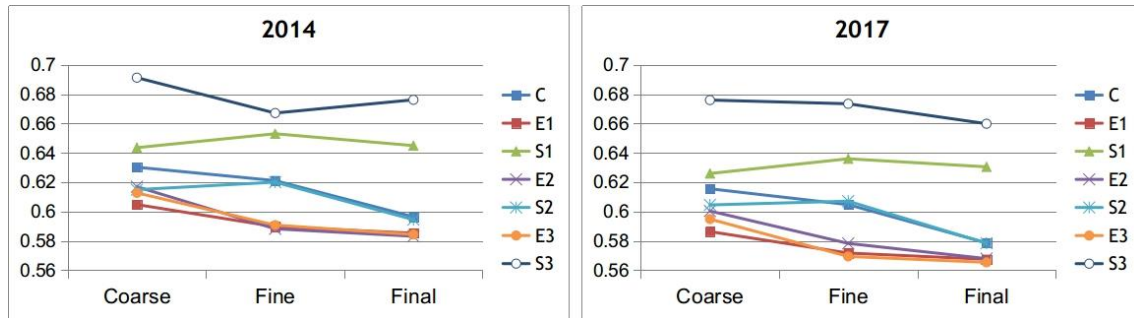
**Figure 9. Lee-Salee Metric.**
Source: the authors, from simulation results.

Most of the curves are decrescent and converge to similar results. The decreasing values of the metrics are somewhat expected since higher levels of detail make it harder for the simulator to match the city evolution. Scenario S3 consistently had the best output, followed by S1. The lower values were found in the three type-E setups, and the simulation of the complete city occupied an intermediate position among all tested scenarios.

4.2 Forecasts of land use

SLEUTH generates output in the form of data tables and images with color-coded pixels that represent the probability of new urban areas. These probabilities are obtained from the Monte Carlo approach. All simulations generated relatively modest growth, but followed the tendencies found in the calibration. The results are summarized in Table 2.

**Table 2. Number of generated pixels in forecast results.**

| Setup | C | E1 | E2 | E3 | S1 | S2 | S3 |
|---|---|---|---|---|---|---|---|
| **New pixels** | 578 | 206 | 266 | 180 | 285 | 470 | 184 |
| **Proportion** | 100 % | 31 % | 40 % | 27 % | 30 % | 50 % | 19 % |

Source: the authors, from simulation results.

Column C of Table 2 corresponds to the complete approach, therefore the proportion of new pixels is equal to 100 %. The remaining columns show the distribution of pixels and the relative proportion found in each of the three regions of the map. The sum of pixels in scenarios S and E is slightly greater than in C, because of the small overlap between regions.

By evaluating the accuracy of projections with respect to the Lee-Salle metric, the best results were those produced with the S approach. The region 2 of the map dominates the expansion, and this is in good agreement with the dynamics observed in the city. New gated communities and residential areas have been projected along the axis that cuts this region, which is impelled by the concentration of industrial activity.

During simulation SLEUTH adjusts coefficients, seeking to match periods where city expansion accelerates or slows down (Clarke and Gaydos 1998). The modifications are applied depending on the rate of simulated growth and parameters from a configuration file; this can be found in the code, in the subroutine coeff_SelfModication [sic]. Log files from the experiments

registered that all coefficients except Slope were reduced, and this happened along the whole simulation period beginning in 1984. This behavior of the software is in accordance with the tendency followed by the city, as shown in Figure 10.
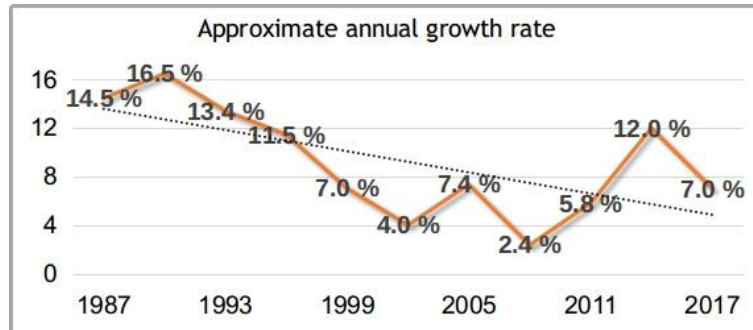


**Figure 10. Annual percentage of city growth.**
Source: IBGE (2019).

The plot in Figure 10 represents the percentage of growth according to the series of city images; the dotted line is a linear regression of the data. The growth deceleration that happens in the city was somewhat over-emphasized by the simulations.

The SLEUTH DNA of the seven scenarios was compared with the results obtained in the simulation of other 21 cities (Gazulis and Clarke 2006, Pramanik and Stathakis 2016). The Euclidean distance between 5-dimensional points was computed for all cities in the comparison and the seven simulation approaches. The best matches found are shown in Table 3.

Distinctive characteristics in the evolution of land use were captured by the model and reflected in the set of coefficients of the corresponding input maps. Two regions exhibited noticeable patterns; they were region 1, which shows similarity with the SLEUTH DNA of Chian Mai; and region 3, with coefficients close to those found for Houston. These regions had the best values of the Lee-Sallee metric in the S approach. Region 2, which presented the most pronounced variations in growth, also showed mixed results with regard to the coefficients. Despite fluctuations of the coefficient Road Gravity, the set of values found in the projections of 2014 and 2017 confirms the importance of the transportation network in region 2.

**Table 3. SLEUTH DNA of each simulated region.**

| Simulation scenario | Corresponding City | Euclidean Distance |
|---------------------|--------------------|--------------------|
| C | Dhaka City (Bangladesh) | 27.5 |
| E1 | Chian Mai (China) | 22.9 |
| S1 | Chian Mai (China) | 26 |
| E2 | Dhaka City (Bangladesh) | 22.7 |
| S2 | Houston (USA) | 16.7 |
| E3 | Houston (USA) | 32.4 |
| S3 | Houston (USA) | 19.1 |

Source: the authors, from simulation results.

The simulation using the segmented approach was able to distinguish the dynamics of regions and adapt to different patterns of land use. Regions 1 and 3, in particular, consistently presented coefficients similar to other cities previously analyzed with the SLEUTH model. The results also validate the procedure that guided the partitioning, involving the analysis of the city under historical and social perspectives to define how the map was divided; from that point, the simulation is executed in a rather mechanical manner, found in many studies of city sprawl. The results can be contrasted with similar research using the SLEUTH model. In the study of Houet *et al.* (2016), the global character of the model was kept intact; this means that coefficients are still averaged over the entire regions being simulated, as in scenario C of the present work. In the study of Jantz *et al.* (2010) maps were divided using an automatic procedure; however, this process requires additional tools and relies on numerical data that in principle does not reflect historic and social elements as those considered in the present paper. In Mahiny and Clark (2012), a suitability layer reflects characteristics such as proximity to rivers, forest density, and climate classification; their main interest was on sustainable land management and the method requires additional analytic data.

## 5. Conclusions

The results of the study confirmed the initial hypothesis, and regionalized simulations showed a better calibration of the model. This was indeed observed in the values of the Lee-Sallee metric, especially for the approach using segmented maps. The use of both a segmented map and the exclusion layer (respectively scenarios S and E) was initially thought as a simple validation test and was expected to give similar results; as the experiments unfolded, it became clear that this assumption did not hold and the reasons had to be examined.

Some fluctuation of coefficients can be regarded as normal in view of the nondeterministic character of the simulation. However, coefficients as Slope and Road Gravity presented significant differences between scenarios S and E. Inspecting the code of the software it was ascertained that the exclusion layer does not affect the calculation of Lee-Salee metric, although it effectively disables the urbanization process. The scenarios E1, E2, and E3 produced the poorest results. According to the logic implemented in the software, during execution with these scenarios a single region Ek was modified but the calibration metric was computed for the entire set $E1 \cup E2 \cup E3$. The extent to which results can be affected by this logic of the algorithms is difficult to assess. For instance, while the cellular automata model was able to prioritize organic growth, mixed values were obtained for coefficients in region 2. This point might deserve specific tests.

The time interval between images, 3 years, was shorter than other experiments in the literature; an additional test of the simulator was conducted using a larger interval of 6 years and running only the initial phases of calibration with complete maps. The Lee-Salee metric attained 0.7, but once again the reasons for this behavior of the simulator are hard to determine with certitude; a possible explanation is the fact that removing 2 out of 3 images of the series, erratic growth patterns became less pronounced in the input and this had a positive impact on the calculations. One possible study would be to check the simulator sensitivity to random input patterns. A few different metrics of urban sprawl have been evaluated in the context of SLEUTH studies. Interestingly, metrics as Lee-Salee, recommended in the documentation of the software, are

only a by-product of model execution, not used as a criterion to guide the simulated trajectory. There is instead a set of heuristics in the code which depend on statistics as the rate of growth.

The city of Ponta Grossa, chosen for this study, presents some distinctive characteristics. Historical changes of land use occur in disorganized patterns, as a consequence of municipal legislation being weakly enforced and social and economical forces that exerted a considerable influence on the territory. The concentration of certain activities in specific zones, during the XIX century, can still be perceived in the current distribution of residential and commercial areas. The urbanization of the city also challenges the logic associated with topography, especially as slums and irregular residences advance over areas that should resist this kind of land use. This phenomenon can be observed in other Brazilian cities and in the present case, it contributed to the rise of irregular morphology and the evolution of the transportation network with irregular layouts.

The increasing amount of available data, represented by official municipal databases, collaborative projects as WikiMapia and OpenStreetMaps, and even private services as iFood or Uber, suggests new avenues of research. One example would be to partition maps according to displacement of people and goods, inferred from sources as Waze, Uber and iFood. Combining other data sources, as financial, and the possibility of full automation of the task, are potential research questions. The concept of DNA might be explored further. For instance, it might be derived from the analysis of current data, instead of being the result of the calibration of the simulation; this way it could guide map segmentation. All those situations might benefit from techniques from fields as big data and data mining.

## Acknowledgements

## References

Aguilera, F., Valenzuela, L. M. & Botequilha-Leitão, A. (2011). "Landscape metrics in the analysis of urban land use patterns: A case study in a Spanish metropolitan area", *Landscape and Urban Planning*, 99, 3-4, 226-238.

Amujal, C. (2015). "Using cellular automata to analyze and model urban growth: a case study of the greater Copenhagen area from 1990 to 2010", Master's thesis, Aalborg University, Denmark.

Aaviksoo, K. (1995). "Simulating vegetation dynamics and land use in a mire landscape using a Markov model", *Landscape and urban planning*, 31, 129–142, 1995.

Alcamo, J. (2008). "Chapter Six The SAS approach: combining qualitative and quantitative knowledge in environmental scenarios", *Developments in integrated environmental assessment*, 2, 123-150.

Arsanjani J. J., Helbich M., Kainz W. & Boloorani A. D. (2013). "Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion", *International Journal of Applied Earth Observation and Geoinformation,* 21: 265-275.

Bandini S., Mauri G. & Serra R. (2001). "Cellular automata: From a theoretical parallel computational model to its application to complex systems", *Parallel Computing*, 25, 5, 163–168.

Basse, R. M., Omrani, H., Charif, O., Gerber, P. & Bódis, K. (2014). "Land use changes modelling using advanced methods: Cellular automata and artificial neural networks. The spatial and explicit representation of land cover dynamics at the cross-border region scale", *Applied Geography*, 53, 160-171.

Batty, M. (1976). "Urban modelling". Cambridge: Cambridge University Press.

Berberoglu, S. & Akin, A. (2009). "Assessing different remote sensing techniques to detect land use/cover changes in the eastern Mediterranean", *International Journal of Applied Earth Observation and Geoinformation*, 11, 1, 46-53.

Bhatta B., Saraswati S. & Bandyopadhyay D. (2010). "Urban sprawl measurement from remote sensing data", *Applied Geography,* 30, 4, 731–740.

Brown, D. G., Walker, R., Manson, S. & Seto, K. (2012). "Modeling land use and land cover change". In. Gutman, G., Janetos, A. C., Justice, C. O., Moran, E. F., Mustard, J. F., Rindfuss, Cochrane, M. A. Land Change Science, 395–409. Dordrecht: Springer.

Chen, Y., Liu, X. & Li, X. (2017). "Calibrating a Land Parcel Cellular Automaton (LP-CA) for urban growth simulation based on ensemble learning", *International Journal of Geographical Information Science*, 31, 12, 2480-2504.

Clarke K. C. (2004). "Toward geocomputational honesty in urban modeling". Atkinson P., Foody G. M. & Darby F. W. S. E. (Eds). Geodynamics, Boca Raton: CRC Press, 215–232.

Clarke K. C. & Gaydos L. J. (1998). "Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore", *International Journal of Geographical Information Science,* 12, 7, 699–714.

Clarke, K. C. (2018). "Cellular automata and agent-based models", Handbook of regional science, Berlin Heidelberg: Springer, 1-16.

Creighton, R, L., Junior, J. D. C. and Finney, G. S. (1959). "Data processing for city planning", *Journal of the American Institute of Planners*, 25, 2, 96-103.

Crooks A.T., Heppenstall A.J. (2012). "Introduction to Agent-Based Modelling", In: Heppenstall A., Crooks A., See L., Batty M. (eds) Agent-Based Models of Geographical Systems. Springer, Dordrecht.

Dietzel C. & Clarke K. C. (2016). "The effect of disaggregating land use categories in cellular automata during model calibration and forecasting", *Computers, Environment and Urban Systems,* 30, 1, 78–101.

Dutta, I. & Das, A. (2019). "Exploring the dynamics of urban sprawl using geo-spatial indices: a study of English Bazar Urban Agglomeration, West Bengal", *Applied Geomatics*, 11, 3, 1-18.

Filatova, T., Verburg, P. H., Parker, D. C. & Stannard, C. A. (2013). "Spatial agent-based models for socio-ecological systems: challenges and prospects", *Environmental Modelling & Software*, 45, 1-7.

Fotheringham, A. S. (1985). "Spatial competition and agglomeration in urban modelling", *Environment and Planning A*, 17, 2, 213-230.

Fujita, M. & Thisse, J. F. (1996). "Economics of agglomeration*", Journal of the Japanese and International Economies*, 10, 4, 339-378.

Gazulis N. & Clarke K. C. (2006). "Exploring the DNA of our regions: Classification of outputs from the SLEUTH model", In: Yacoubi S. E., Chopard B. & Bandini S. (Eds), Cellular Automata, ACRI 2006. Lecture Notes in Computer Science, 4173, Berlin: Springer, 462–471.

Goodchild, M. F. (2018). "Reimagining the history of GIS", *Annals of GIS*, 24, 1, 1-8.

Gonçalves, M. A. C. & Pinto, E. A. (1983). "Ponta Grossa-um século de vida (1823-1923)" [Ponta Grossa - one century of life (1823-1923)], Ponta Grossa: Ed. Kugler.

Gounaridis, D., Chorianopoulos, I., Symeonakis, E. & Koukoulas, S. (2019). "A Random Forest-Cellular Automata modelling approach to explore future land use/cover change in Attica (Greece), under different socio-economic realities and scales", *Science of the Total Environment*, 646, 1, 320-335.

Haase, D. and Schwarz, N. (2009). "Simulation models on human-nature interactions in urban landscapes: a review including spatial economics, system dynamics, cellular automata and agent-based approaches", *Living Reviews in Landscape Research*, 3, 1–45.

Houet T., Aguejdad R., Doukari O., Battaia G. & Clarke K. C. (2016). "Description and validation of a non path-dependent model for projecting contrasting urban growth futures", *Cybergeo: European Journal of Geography*, DOI: https://doi.org/10.4000/cybergeo.27397.

IBGE Instituto Brasileiro de Geografia e Estatística (2019). "SIDRA, Sistema IBGE de Recuperação Automática" [SIDRA, IBGE System for Automatic Recovering], https://sidra.ibge.gov.br, Retrieved 01-04-2020.

Jantz, C. A., Goetz, S. J., Donato, D. & Claggett, P. (2010). "Designing and implementing a regional urban modeling system using the SLEUTH cellular urban model", *Computers, Environment and Urban Systems*, 34, 1, 1-16.

Kazemzadeh-Zow A., Shahraki, S. Z., Salvati L. & Samani N. N. (2017). "A spatial zoning approach to calibrate and validate urban growth models", *International Journal of Geographical Information Science,* 31, 4, 763-782.

Kemp-Benedict, E. (2004). "From Narrative to Number: A Role for Quantitative Models in Scenario Analysis", In: 2nd International Congress on Environmental Modelling and Software - Osnabrück, Germany.

Kok, K. & Verburg, P. H. (2007). "Integrated Assessment of the land system: The future of land use", *Land use policy*, 3, 24, 517-520.

Kubaski, L. (2015). "Imigração e educação dos poloneses em Ponta Grossa, PR" [Polish imigration and education, in Ponta Grossa, PR], M.Sc. Dissertation, Universidade Estadual de Ponta Grossa.

Lee D. R. & Sallee G. T. (1970). "A method of measuring shape", *Geographical Review,* 60, 4, 555–563.

Li X., Gong P., Yu L. & Hu T. (2017). "A segment derived patch-based logistic cellular automata for urban growth modeling with heuristic rules", *Computers, Environment and Urban Systems,* 65, 140–149.

Liu X., Sun R., Yang Q., Su G. & Qi W. (2012). "Simulating urban expansion using an improved SLEUTH model", *Journal of Applied Remote Sensing*, 6, 1, 1-20.

Loibl W. & Toetzer T. (2003). "Modeling growth and densification processes in suburban regions-simulation of landscape transition with spatial agents", *Environmental Modelling & Software*, 18, 6, 553-563.

Macal, C. M. & North, M. J. (2010). "Tutorial on agent-based modelling and simulation", *Journal of Simulation*, 4, 151–162.

Mahiny A. S. & Clarke K. C. (2012). "Guiding sleuth land-use/land-cover change modeling using multicriteria evaluation: towards dynamic sustainable land use planning", *Environment and Planning B: Planning and Design*, 39, 5, 925–945.

Manandhar, R., Odeh, I. & Ancev, T. (2009). "Improving the accuracy of land use and land cover classification of Landsat data using post-classification enhancement", *Remote Sensing*, 1, 3, 330-344.

Ménard, A. & Marceau, D. J. (2005). "Exploration of Spatial Scale Sensitivity in Geographic Cellular Automata", *Environment and Planning B: Urban Analytics and City Science*, 32, 5, 2005.

Moghadam, S. A., Karimi, M. & Habibi, K. (2018). "Simulating urban growth in a megalopolitan area using a patch-based cellular automata", *Transactions in GIS*, 22, 1, 249-268.

Mustafa, A., Cools, M., Saadi, I. & Teller, J. (2017). "Coupling agent-based, cellular automata and logistic regression into a hybrid urban expansion model (HUEM)", *Land Use Policy*, 69, 529-540.

Ozturk, D. (2015). "Urban growth simulation of Atakum (Samsun, Turkey) using cellular automata-Markov chain and multi-layer perceptron-Markov chain models", *Remote Sensing*, 7, 5, 5918-5950.

Parker, D. C., Manson, S. M., Janssen, M. A., Hoffmann, M. J. & Deadman, P. (2003). "Multi-agent systems for the simulation of land-use and land-cover change: a review", *Annals of the association of American Geographers*, 93, 2, 314-337.

Ponta Grossa (2018). "Revisão do Plano Diretor Municipal de Ponta Grossa, Parte 3" [Revision of City Development Plan of Ponta Grossa, Part 3]. City Hall.

Pramanik, M. M. A. & Stathakis, D. (2016). "Forecasting urban sprawl in Dhaka city of Bangladesh", *Environment and Planning B: Planning and Design*, 43, 4, 756-771.

Rae, A. & Wong, C. (2012). "Monitoring spatial planning policies: towards an analytical, adaptive, and spatial approach to a 'wicked problem' ", *Environment and Planning B: Planning and Design*, 39, 5, 880-896.

Ratcliffe, M., Burd, C., Holder, K. & Fields, A. (2016). "Defining rural at the US Census Bureau", *American community survey and geography brief*, 1-8.

Rimal, B., Zhang, L., Keshtkar, H., Haack, B., Rijal, S. & Zhang, P. (2018). "Land use/land cover dynamics and modeling of urban land expansion by the integration of cellular automata and Markov chain", *ISPRS International Journal of Geo-Information*, 7, 4, 154.

Ritchey, T. (2013), "Wicked problems", *Acta morphologica generalis*, 2, 1.

Rivals, I. & Personnaz, L. (1996). "Black-box modeling with state-space neural networks", *Neural Adaptive Control Technology*, 237-264, DOI 10.1142/9789812830388_0008.

Rocha, F. J. P. S. (2012). "Sistemas complexos modelação e geosimulação da evolução de padrões de uso e ocupação do solo" [Complex systems, modelling and geosimulation of the evolution of patterns of land use and occupation], Doctoral Dissertation, Instituto de Geografia e Ordenamento do Território, Universidade de Lisboa, Portugal.

Rodríguez, A. & Meneses, J. (2011). "Transformaciones rurales en América Latina y sus relaciones con la población rural" [Rural transformations in Latin America and its relations with rural population], Reunión de expertos sobre población, territorio y desarrollo sostenible, Comisión Económica para América Latina y el Caribe (CEPAL), Santiago de Chile, 16-17 August 2011.

Roth, E. C. W. (2019). "Urban Growth Forecast Using Segmented and Complete Maps with the Sleuth Simulator", Master's thesis, Federal Technological University of Paraná.

Rozenstein, O. & Karnieli, A. (2011). "Comparison of methods for land-use classification incorporating remote sensing and GIS inputs", *Applied Geography*, 31, 2, 533-544.

Sanders, P. & Sanders, F. (2004). "Spatial urban dynamics. A vision on the future of urban dynamics: Forrester revisited", In: Kennedy, M., Winch, G. W., Langer, R. S., Rowe, J. I. & Yanni, J. M. (Eds) International System Dynamics Conference at Oxford, System Dynamics Society.

Santé I., Garcia A. M., Miranda D. & Crecente R. (2010). "Cellular automata models for the simulation of real-world urban processes: A review and analysis", *Landscape and Urban Planning,* 96, 2, 108–122.

Sarkar, P. (2000). "A brief history of cellular automata", *ACM Computing Surveys*, 32, 1, 80-107.

Shen Z., Kawakami M. & Kawamura I. (2009). "Geosimulation model using geographic automata for simulating land-use patterns in urban partitions", *Environment and planning B: Planning and Design,* 36, 5, 802–823.

Silva A. C. (2013). "Identificação das Áreas de Vulnerabilidade Socioambiental Mediante Lógica Fuzzy – Estudo no Município de Ponta Grossa" [Identification of Social and Environmental Vulnerable Areas using Fuzzy Logic - A study in Ponta Grossa], M.Sc. Dissertation, Universidade Estadual de Ponta Grossa.

Sohl T. L. & Claggett P. R. (2013). "Clarity versus complexity: Land-use modeling as a practical tool for decision-makers", *Journal of Environmental Management*, 129, 235–243.

Tobler, W. R. (1970). "A computer movie simulating urban growth in the Detroit region", *Economic Geography*, 46, sup 1, 234-240.

Torrens P. M. & O'Sullivan D. (2001). "Cellular automata and urban simulation: where do we go from here?", *Environment and Planning B: Planning and Design*, 28, 2, 163–168.

Triantakonstantis D. & Mountrakis G. (2012). "Urban growth prediction: a review of computational models and human perceptions", Journal of Geographic Information System, 4, 6, 555–587.

United Nations. (2018). "2018 revision of world urbanization prospects".

Volkery A., Ribeiro T., Henrichs T. & Hoogeveen Y. (2008). "Your vision or my model? Lessons from participatory land use scenario development on a European Scale", Systemic Practice and Action Research, 21, 6, 459–477.

Wahyudi, A. (2016). "Modelling urban growth: Integrating both physical and human dimensions in developing world context", PhD Thesis, University of Queensland, Australia.

Wolfram, S. (1984). "Cellular automata as models of complexity", Nature, 311, 419–424.

Wang Y. & Li, S. (2011). "Simulating Multiple Class Urban Land-Use/Cover Changes by RBFN-Based CA Model", Computers and Geosciences, 37, 2, 2011, 111-121.