A microsimulation approach for modelling the growth of small urban areas

MSc Dissertation

Master Programme in Planning and Project of the Urban Environment
School of Engineering
University of Porto
Portugal

Author
Nuno Eduardo Norte Pinto
Civil Engineer (University of Coimbra)
Lecturer at the Civil Engineering Department
School of Technology and Management
Polytechnic Institute of Leiria
Portugal

Director
António José Pais Antunes
Professor at the Civil Engineering Department
School of Sciences and Technology
University of Coimbra
Portugal

Porto, December 2006

Resumo

O estudo apresentado nesta dissertação centra-se no uso de um modelo baseado em autómatos celulares (AC) para a simulação de fenómenos de evolução urbana no contexto de pequenas áreas urbanas. Os modelos baseados em AC são objecto de intensa investigação teórica e prática, tendo daqui resultado uma série de modelos operacionais aplicados em diversas regiões e áreas urbanas em todo o mundo. Este estudo foca alguns aspectos que ainda não estão devidamente estudados, relacionados com a escala do problema e com características das células. É avaliada a aplicação de um modelo baseado em AC para a simulação de pequenas áreas urbanas. A utilização de células irregulares baseadas em unidades espaciais de informação estatística é considerada uma importante evolução para a aplicação de modelos baseados em AC, uma vez que a informação demográfica e socio-económica está disponível com este nível de desagregação. A calibração do modelo é realizada com recurso a um algoritmo de optimização denominado Particle Swarm. O modelo foi aplicado a uma série de problemas teste que foram desenvolvidos para avaliar o comportamento do modelo face aos diversos parâmetros de calibração. O modelo foi posteriormente aplicado a um estudo de caso centrado no município de Condeixa-a-Nova, seleccionado devido às elevadas taxas de crescimento demográfico e de área construída verificadas na última década censitária. O modelo produziu resultados promissores tanto para os problemas teste como para o estudo de caso, indicando a sua aplicabilidade para áreas urbanas de pequena escala. Os resultados mostram ainda a possibilidade do uso de células irregulares.

Abstract

The study presented in this dissertation is centred on the use of a cellular automata (CA) model for the simulation of urban change phenomena in small-size urban areas. CA-based models are currently under intensive theoretical and operational research, with a series of models being applied to several urban and metropolitan areas around the world. The study focus on important issues that were not fully considered so far regarding cell characteristics and problem size. The application of the model to small-size urban areas is evaluated. The feasibility of using irregular cells based on census tracts is also assessed. The use of irregular cells is believed to be an important development of CA models for their application, as demographic and socio-economic data is usually available for irregular census tracts. The calibration of the model is made through an optimization procedure based on a Particle Swarm algorithm. A series of theoretical test problems are used to evaluate the behaviour of the model and the method used to calibrate its parameters. The model is also applied to the case study of Condeixa-a-Nova, a small municipality in the Portuguese central region. The case study was selected because of the high growth rates of population and built area verified throughout the last decade. The model produced promising results not only for the test problems but also for the case study, indicating its ability for dealing with small-size urban areas. Also, it has shown that is feasible to use irregular cells.

Agradecimentos
Acknowledgments

Num trabalho de longos meses, como o que foi desenvolvido neste curso de mestrado que agora termina com a publicação desta dissertação, é sempre necessário contar com (e recorrer a) ajuda de quem tem algo de construtivo para o nosso trabalho.

Uma palavra sincera de agradecimento é devida ao Professor António Pais Antunes, orientador desta tese. A forte motivação que sempre me incutiu e a total disponibilidade com que sempre atendeu as minhas solicitações foram, e continuarão certamente a ser nos trabalhos que continuaremos a desenvolver, inexcedíveis.

Outros contribuíram com algo para este trabalho, entre dicas e ajudas, sem as quais seria mais difícil atingir o objectivo final. Agradeço ao Bruno Santos e ao Pedro Nuno pelas suas ajudas com papers e com a sua gestão, ao Eduardo Ribeiro pela sua infindável paciência para com as minhas dúvidas no processamento de texto e afins e ao Zé Carlos pelas importantes dicas sobre SIGs.

Outros ainda deram um apoio fundamental na logística da coisa, não só durante o período de elaboração da dissertação como também durante o ano escolar do mestrado. Agradeço à minha sogra, D. Helena, pelas incontáveis viagens para Coimbra, à Susete pela sua incansável disponibilidade para brincar e à Xana pequena e ao Sérgio por me acolherem no Porto.

A pedra de toque desta minha empresa foi a alegria da minha filha Diana, que sempre me acarinhou com um sorriso à prova de saudades e com uma compreensão invulgar para a sua tenra idade.

Para a minha mulher Inês um beijo de obrigado do fundo do meu coração. Acreditou em mim e teve uma paciência de santa para aguentar estes longos meses de noitadas atrás de noitadas, de fins-de-semana arruinados por um trabalho por vezes difícil de vislumbrar.

Muito obrigado a todos.

Até à próxima, quando mais uma vez me sentirei à vontade para vos solicitar novas ajudas.

Contents

List of figures

List of tables

1 Introduction

In this chapter issues regarding problem contextualization, motivation and methodological approach will be addressed. It is important to clearly identify the problem at hand in order to correctly justify the need for the study and the application of a given approach. The main goals of the study will be stated and it will also be presented a short description of the methodological approach used. Finally, a brief overview of the thesis will be made.

1.1 Context

The rapid growth of urban areas is one of the most complex problems in urban sciences and raises important issues to the planning process. These issues regard not only the motives behind past evolution, but also the definition of new strategies and policies capable to respond to the needs of the present contexts and to predict and control the evolution towards a sustainable future.

The complexity of this type of problems is such that there are no simple ways to achieve a solution (the very concept of "a solution" is far from being close to planning), nor these solutions are based on a single approach. The comprehensive character of the planning process is simultaneously a strength, because it results from the consideration of all ingredients of the problem (physical, sociological, economic, historical, among several others), and a potential weakness, because the problem becomes more and more complex as the natural evolution of societies takes place, especially in the current globalized context, demanding from planners new levels of commitment and accuracy in their research and work.

A comprehensive approach to the planning process raises the necessity of integrating knowledge from different areas of science, including operational research and other mathematical techniques. This knowledge can be used to develop models that aim to explain urban phenomena, retain knowledge from urban systems, and forecast planning scenarios.

The Portuguese urban planning panorama is interesting at several levels, as a result of the context within which the country evolved, from an undeveloped situation in the 1970s (in terms of physical and social infrastructure, and, more generally, urban quality) to a period of growth triggered by the country's integration in the former European Economic Community (EEC) in the mid 1980s.

The urban population in Portugal was, in 1970, about 66 percent of the total population. By 2001, this figure has grown up to 75 percent (Marques, 2004), meaning a growth rate of 41 percent in this period (Table 1 and Table 2).

Table 1 Evolution of the Portuguese population since 1950 (inhabitants) (Marques, 2004)

	1950	1960	1970	1981	1991	2001
Mainland Portugal*	7895153	8251175	8008000	9337660	9375926	9869343
Urban Areas	4571460	4975394	5253784	6640217	6831176	7407061
Non-Urban Areas	3323663	3275781	2754216	2697443	2544750	2462282
Urban Areas/Mainland	58%	60%	66%	71%	73%	75%

* Does not include Azores and Madeira islands

Table 2 Evolution of the Portuguese population since 1950 (variation) (Marques, 2004)

	1950/1960	1960/1970	1970/1981	1981/1991	1991/2001
Mainland Portugal*	4.5	-2.9	16.6	0.4	5.3
Urban Areas	8.8	5.6	26.4	2.9	8.4
Non-Urban Areas	-1.4	-15.9	-2.1	-5.7	-3.2

* Does not include Azores and Madeira islands

In Figure 1 is represented the evolution of the Portuguese urban system for the period 1950 to 2001. The backbone of the Portuguese urban system is the axis formed by the two largest cities, Lisbon and Porto, each one heading metropolitan areas with more than one million inhabitants. Together, they sum up about 39 percent of the national population (Marques, 2004). The rest of the urban population, which amounts to another 36 percent of the Portuguese population, occupies a set of mid-size cities with no more than 100,000 inhabitants, located between or around those metropolitan areas. Each one of the mid-size cities is located at the functional centre of a group of smaller towns, with a population ranging from a few thousands to no more than 20,000. The concentration of these urban areas in the coastal zone generated, during the past decades[1], a structural unbalance with the off-coast regions of the country.

The asymmetry of the urban system is still being aggravated by a smaller but continuous flow of population and jobs from off-coast regions (with a strong rural character and a diffuse urban land use) to the littoral. The large coastal urban area, from Setúbal (in the southern Lisbon metropolitan area) to Braga (50 kilometres north of Porto), continues to attract population from other regions as a result of the concentration of employment and new urban functions.

[1] This trend has begun in the late 1940s, when large migration flows from rural areas of the country led to the concentration of population near the coast, especially in Lisbon.

Large sums of structural funds allocated to Portugal by the European Union were spent in the development of new infrastructures in urban areas, particularly in large cities located in the coast due to the concentration of population. Those funds were spent mostly on high density urban areas. However, smaller cities have also benefited from the application of some of these funds, particularly those funds oriented to the construction of water distribution networks, solid waste management systems, road networks, and other basic infrastructure.

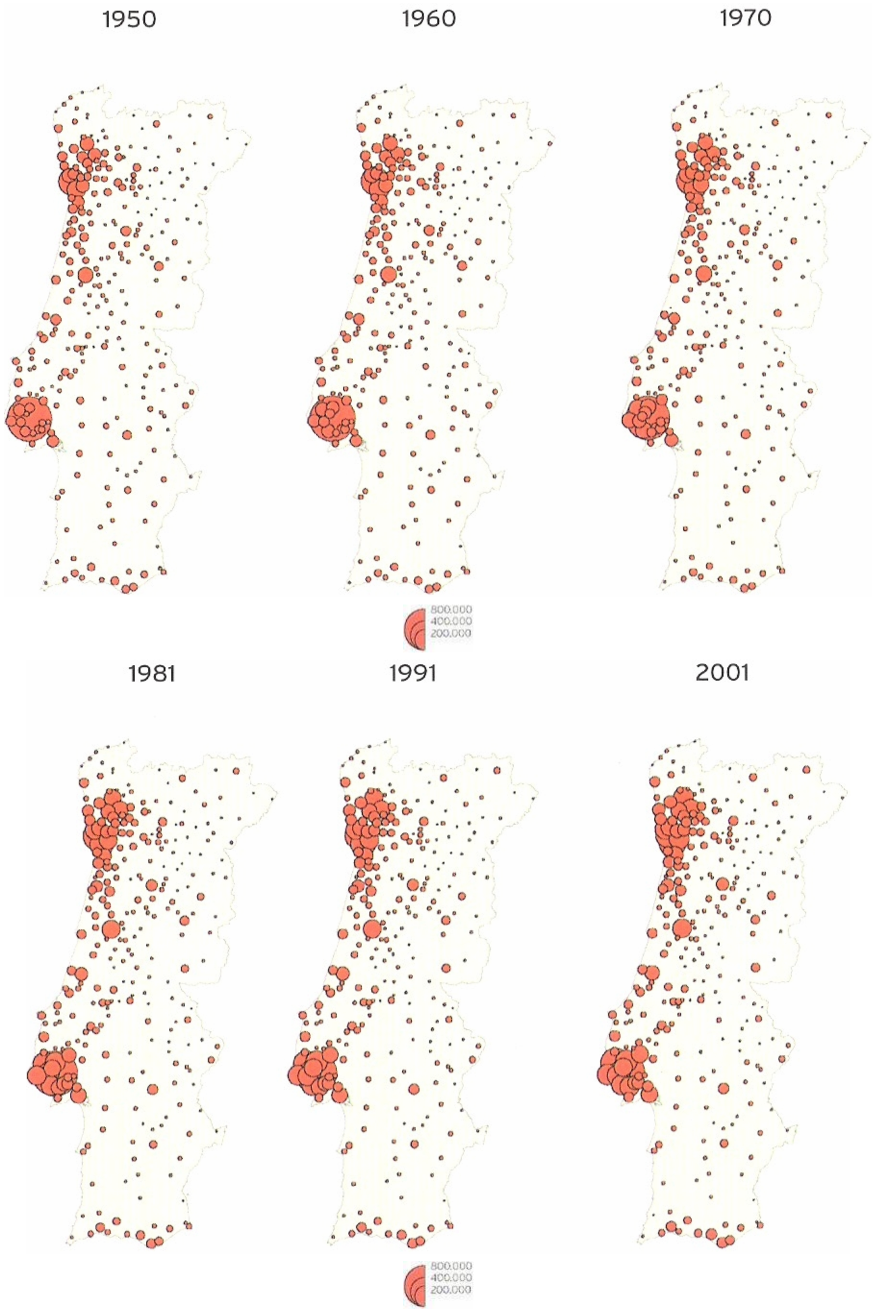


Figure 1 Evolution of the Portuguese urban system from 1950 to 2001 (Marques, 2004)

The smaller urban sub-systems, formed by groups of small towns located within a short distance from mid-size cities, have experienced in the last decades important developments due to public investments co-supported by the EU structural funds. The increase in quality of life experienced by small towns induced a strong growth dynamics, as a result of new or renewed urban services and infrastructures. This growth dynamics promoted the location of new urban functions transforming these small towns into larger towns with low land prices when compared with the central urban areas of those small conurbations. As a consequence, the growth rate of urban areas was very high for a large number of municipalities all over the country. In Figure 2 it is possible to verify the growth trend that all but small-size urban areas have registered for the last 50 years, with a strong concentration in Lisbon's metropolitan area until the beginning of the 1980s. However, it is interesting to note that small-size urban areas have inverted the trend and started to grow after 1991.

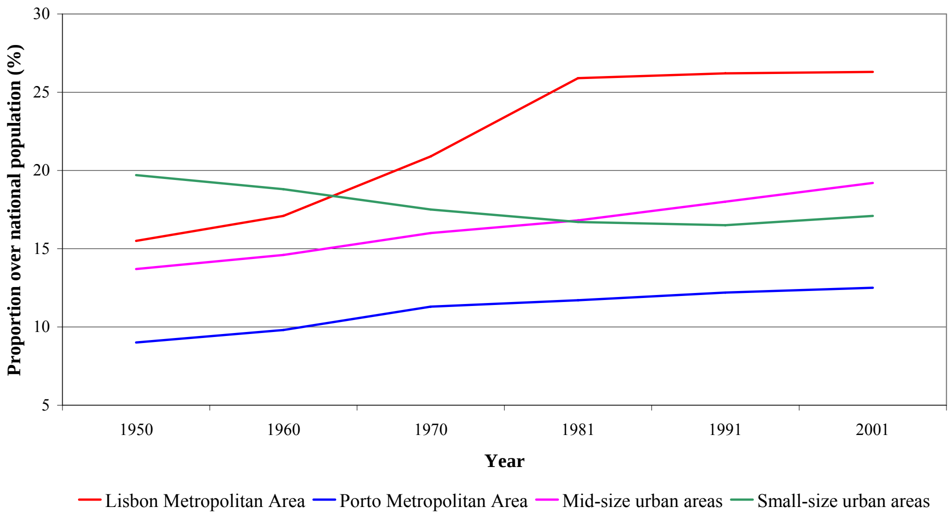


Figure 2 Proportion over national population

At the same time, new forms of municipal management based on direct negotiations with developers and land owners started to take place. This practice gave a new flexibility to planning processes, promoting the rapid increase of available land.

Finally, the economic impulse experienced by all sectors of the economy as a result of the integration in the former EEC brought lower interest rates, raising housing demand, with a strong preference for new construction rather than the regeneration of the existing housing stock.

While urban areas were expanding, the Portuguese planning system was entirely reformed. These two simultaneous processes led to a particularly sensitive context for planning practice in Portugal. The definition of new objectives and strategies in this context of legal change and rapid growth should have had the support of better tools, based on complex modelling techniques (such as Spatial Decision Support Systems, or SDSS). Their intrinsic ability to forecast and evaluate scenarios could be used to inform planners and the general public, and to assist elected local and national officials in decision making. However, in practice, they were rarely applied.

1.2 Motivation

The work presented in this dissertation was developed within a master course on planning and project of the urban environment, at the University of Porto, Portugal. This master course is strongly oriented towards the integration of different scientific areas, from regional and urban planning to environmental sciences and infrastructure engineering, and different types of practitioners, from urban planners and architects to civil engineers and geographers. It was interesting to notice during the teaching year of the course, particularly from the team work developed by multidisciplinary working groups, that there is a hard work that needs to be done, in Portugal, to achieve an effective integration of those scientific areas and practitioners.

The use of models in the planning practice in Portugal was very limited by the time this study was done. There are few reports on the application of models, all of them produced as a result of R&D projects developed at universities and research centres (Julião, 2001; Silva, 2002; Silva and Clarke, 2002).

Two main reasons are believed to be the cause for this lack of modelling in planning practice. Firstly, for all levels of the Portuguese administration – local, regional[2] and national – there was and still is a strong deficit of practitioners capable of ensuring a proficient planning practice. Particularly at the local level, these practitioners are usually engaged in municipal management activities or construction licensing rather than in planning activities. Secondly, the use of spatial models aimed to assist urban and regional planning has little tradition; there are some published studies in the fields of facility location (Antunes, 1999; Antunes and Peeters, 2000), solid waste collection management (Teixeira, Antunes and de Sousa, 2004), and transport network planning (Costa and Markellos, 1997; Antunes, Seco and Pinto, 2003).

1.3 Study goals

This study has two main goals: one regarding the assessment of the applicability of a microsimulation approach to the study of urban change in small urban areas; another regarding the exploration of some characteristics of the modelling technique used.

The model used in the study is based on a cellular automata (CA) discrete model. CA are a modelling tool with an intuitive spatial orientation and its application to urban studies is under intensive research for the past 15 years. In a simple definition “an

[2] Portugal does not have yet formal administrative regions, which are the object of a reform that is under public discussion for the last two decades. There is, however, an intermediate level of planning regions, the Regional Coordination and Developing Commissions, CCDRs, which can be considered as the seeds for the future regions.

automaton is a processing mechanism with characteristics that change over time based on its internal characteristics, rules and external input" (Benenson and Torrens, 2004) CA are based on a discrete set of spatial units called cells that together form a cell space. Each cell takes a given cell state from a finite set of cell states. Time is also considered in a discrete manner. The automaton then operates state changes over time according to a finite set of transition rules that can be of various types (deterministic, stochastic, unconstrained or constrained).

One of the goals of this thesis is to assess the applicability of a modelling approach to the study of urban change in small urban areas. Considering, on the one hand, the necessity of improving the multidisciplinary work in planning and, on the other hand, the urgency of bringing a scientific approach to the analysis of urban phenomena, this study is believed to be strongly pertinent as it can introduce new tools to understand urban change phenomena in Portugal. Furthermore, the model developed in this study may also be the seed of a future planning tool aimed to real world application as a spatial decision support system. Another important issue regarding the applicability of a CA model (and of models of any type) is its calibration and validation for local conditions in opposition to the use of more generic models. This modelling approach is also relevant because this type of approach has rarely been used in the recent planning practice in Portugal.

Another goal of the study is the exploration of CA characteristics that were not fully addressed so far. The application of CA is widely published in several studies. Tobler (1979), Couclelis (1985) and White and Engelen (1993) worked on its theoretical issues regarding CA application to urban studies, Batty and Xie (1997) and Clarke, Hoppen and Gaydos (1997) worked on the application of important evolutions of CA to real world problems, Semboloni (2000) studies urban infrastructure development, O'Sullivan (2001b) uses an integrate approach based on CA and on graph theory to study gentrification, Semboloni (1997) and Ward, Murray and Phinn (2003) developed multi-scale urban models based on CA, and Silva and Clarke (2002) and Barredo, Kasanko, McCormick and Lavalle (2003) made straightforward applications of previously developed CA models to large metropolitan areas. However, all these studies are focused on large scale models, mainly regional and or metropolitan models, discarding the consideration of small, city problems. There is little work on the use of CA for modelling smaller areas. Therefore, an important area of interest in CA research is related with the scale of the problem, particularly with the feasibility of this type of models when the subject of study is a small urban area.

Finally, the last goal of this study is related with testing the use of irregular cells on a CA approach. There is only a small group of studies that developed CA models considering irregular cellular fabrics (Semboloni, 2000; Vandergue, Treuil and Drogoul, 2000; O'Sullivan, 2001a). This particular characteristic is of great importance for the simulation of local scale problems, where the traditional regular cells, obtained from satellite images, do not represent well the spatial structure of the territory. Irregular cells can be considered as “natural spatial cells”, that is, cells based on the use of the census tracts with a shape closer to urban form. This type of cells is thought to be far more representative of land use than regular cells, since census tracts are drawn taking the urban structure into account.

1.4 Outline of the dissertation

As mentioned above, a CA model is developed to assess its ability to model urban change in small-size urban areas. Model performance will be evaluated with a measure of coincidence between modelled and reference maps. Model calibration will use an optimization procedure based on the Particle Swarm algorithm aiming to maximize an objective function of the measure of coincidence. After the development of the model, a series of 20 test problems will be used to assess model performance. The model will also be tested with a real world case study selected among the Portuguese small municipalities.

This thesis is divided in two main parts: the first part is mainly dedicated to the formulation of the problem and to the development of a CA model and is addressed in Chapters 2 and 3; the second part is dedicated to the application of the CA model to a series of test problems and to real world case study as explained in Chapters 4 and 5.

Chapter 2 is a survey of the literature on urban modelling and its applications, where the use of mathematical models is discussed as a tool for the planning process. Particular attention is drawn to the use of CA as a modelling tool. The survey covers the mathematical concept of CA, its discovery and further development by geographers and urban planners. It also covers questions related to further relaxations, known model applications and CA models calibration.

Chapter 3 is dedicated to explain the methodological approach used in this study, regarding the use of a CA model, its characteristics, constraints and limitations. It will also be presented the optimization approach used to calibrate the simulation, based on the Particle Swarm algorithm. The integrated model that results from the joint application of CA and an optimization procedure will be assessed by its application to a series of theoretical problems generated specifically for this purpose.

Chapters 4 and 5 are devoted to the application of the CA model. Chapter 4 presents the model results for the application to a series of theoretical test problems. In Chapter 5 the results for the application of the model to a real world case study are presented.

Finally, in Chapter 6 an attempt is made to draw together all the issues raised by this study, concluding about the results of the model application and pointing out future lines of work.

2 Literature Survey

The main goal of this chapter is to present a review of the state-of-the-art practice in microsimulation applied to urban modelling. Sections 2.2.1 and 2.2.2 are focused on the use of models in urban sciences and on the relationship between theory and practical applicability. In Section 2.2.3 microsimulation and its main techniques will be presented.

Secondly, in Section 2.3, the use of cellular automata (CA) as a feasible approach to this type of problems is addressed. The mathematical concept of CA is presented in section 2.3.1. In Sections 2.3.2 and 2.3.3 issues regarding the use of CA in urban studies and recent evolutions are discussed. In Section 2.3.4 a series of important CA applications are presented. Section 2.3.5 is dedicated to the discussion over measurement techniques for CA models. In Section 2.3.6 a series of calibrations procedures for CA models are presented.

2.1 Overview

The use of mathematical tools to model a wide range of spatial problems has been classified for the last decades as an important approach to scientific planning (Batty, 1994). The growth of urban areas is one of the issues which have concentrated a large research effort (Tobler, 1970, 1979; Couclelis, 1985; White and Engelen, 1993; Batty, Couclelis and Eichen, 1997; Clarke *et al.*, 1997; Couclelis, 1997; Fragkias and Seto, 2005). Before the democratization of the use of computers, back in the 1970s, modelling approaches were usually applied on problems which demanded few information and, consequently, small computational effort. The information was taken in aggregate form, reducing the sensitiveness of those models, which were mainly static and deterministic. The models assumed an initial state of equilibrium, both in time and space, evolving to a new equilibrium after an exogenous stimulus (Waddell and Ulfarsson, 2004). In the mid 1970s, a period of reflection started after the publication of Douglass Lee article "*Requiem for large-scale models*" (Lee, 1973), where the role of models as tools for scientific planning was severely criticised.

After this first period of large scale modelling that somehow ended by the time Lee published his criticisms, a series of new attempts were made and a new era of modelling began with the new computational resources provided by micro computation.

Researchers seek new mathematical approaches founded in much more complex theories, such as discrete choice, agent-based simulation or CA, as the capacity of processing data at lower costs increased exponentially. At the same time, the development of new computers capable of producing better graphical representations of the problems brought new enthusiasm to the use of models (Klosterman, 1994). The new mathematical theories and modelling tools made feasible the use of disaggregate information and the application of stochastic approaches, reducing the scale of the problems down to agents themselves (individuals, households, trips), thus improving the feasibility of the models (Waddell and Ulfarsson, 2004). It was the beginning of the microsimulation era.

Several modelling tools used in microsimulation will be addressed, such as agent-based and multi-agent simulation, CA, discrete-choice, rule-based simulation, and geographical information system (GIS) based simulation. Their main characteristics, advantages/disadvantages, and application conditions will be discussed, as well as their applicability to real world problems.

2.2 Modelling and urban sciences

2.2.1 Modelling and the planning process

From the very early days of planning, like in several other areas of knowledge, a discussion over how it should be approached took place, creating the usual and most certainly necessary tension between theorists and practitioners.

In the specific case of planning, this issue becomes even more complex as it encompasses all sectors of society, from citizens to politicians, from bureaucrats to general stakeholders. Therefore, a necessity for creating a strong and organized planning system emerged from both planning theory and planning practice. As one of the most complex human activities, planning can use all the help it can get from a wide spread areas of knowledge (Couclelis, 2005).

After the blueprint planning era, where the final plan, the final image of the landscape or of the city represented the ultimate goal to achieve, planning become more comprehensive, incorporating social sciences, operations research, economic theories and regional science, trend introduced by the work of Rexford Tugwell and Harvey Perloff at the Chicago School (Klosterman, 1994).

The increasing interest in rational planning (a basic assumption of the Chicago School), was founded on the assumption that scientific approaches and systematic decision-making were the best way to deal with problems in such fields as management, politics and economics. Modelling was considered a major achievement of this scientific approach, therefore a new and important way to assist planning activities (Batty, 1994; Lee, 1994). These “new tools of planning” as Britton Harris called them in 1965, were thought to be a major technological breakthrough that would revolutionize the practice of urban policy making (Wegener, 1994).The idea of planning as a straight and rigid process that lead from a problem to a solution (materialized by the plan), without perturbations of any kind (judgement errors, public participation, monitorization, re-orientation of goals) changed to the cyclic approach proposed for example by McLoughlin (1969).

The introduction of computer sciences in the early 1950s brought new capabilities to mathematical calculus and data processing, limited only by the speed constraints of computer processors and available memory space. By that period the first models of cities were introduced. These models usually focused on transportation and land use allocation problems (Klosterman, 1994). During the 1960s, an important number of cities in the United States already had ambitious land use and transportation models running, and some of these models were already assisting planning activities (Lee, 1973; Batty, 1994; Klosterman, 1994).

In 1973 Douglass Lee published in the Journal of American Planners one of the most important papers ever published in the field, the famous "*Requiem for large-scale models*" (Lee, 1973). He identifies in this paper seven major "sins" of large-scale models: (1) *hypercomprehensiveness*: the early models tried to replicate too large and too complex a system in only one model, in a time when urban knowledge still was taking its first steps; (2) *grossness*: aiming for a large number of results obtained from the models, the information outputted by them was too rough to be use in practice; (3) *hungriness*: these models demand an enormous amount of data; (4) *wrongheadedness*: the models often deduced behaviours for some relationships that could not bet generalized for a different subset of data obtained from the same problem; (5) *complicatedness*: the results of those models were somewhat so complex that they usually need some kind of exogenous intervention that would rebalance the output, with the consequent lost of scientific validity; (6) *mechanicalness*: the systematic errors due to mathematical processes usually produced large amounts of untraceable errors; and finally 7) *expensiveness*: the first theoretical and operational models from the 1960s were to much expensive, in the order of a few millions of dollars. Lee also emphasised the fact that, at the time, no model had produced any kind of relevant theory, as well as no model was founded on strong theoretical grounds.

To make some progress in this field at this critical turning point, Lee draws four major conclusions: (1) models should be more intuitive for potential users; (2) models should combine strong theoretical foundations, objective information and judgement, in order to eliminate the empiricism and the abstract, mainly futile, theorizing; (3) planners should start from simple, well defined problems, towards methods aimed to well identified purposes; and finally (4) models should be simple by nature, since complex models had failed to simulate real life.

Some of the criticism made by Lee where limited by its own boundaries: computational technology was still in a very initial phase, which was, by itself, a strong limitation to the development of models (no matter what the chosen scale was), both in available mathematical tools that were able to be implemented and in data processing capability.

But another author, Gary Brewer, published at the same year of "*Requiem…*", results from his work on the organizational limits to the development of large-scale models. He argued that, rather than theory, technology, data availability, or technical expertise, the inherent difficulties in adapting organizations to technology was the cause for the misuse of models (Batty, 1994).

Batty presents three major achievements that took place in the 1970s, despite this disturbance (Batty, 1994): first, a series of modest but steady refinements on the side of practical applications of land use and transportation models; second, and perhaps the most significant one as Batty argues, the introduction to the urban systems theory of the general concept of optimization, linked to several studies on econometrics and market behaviours; and third, the most challenging one, the incorporation of time – the dynamic behaviour – that could only take place after the development of several new mathematical concepts during the 1960s.

The year of 1973 was the first moment of retrospective for the urban modelling science. Many authors argued opposite opinions on these subjects, but the sense that large scale urban modelling had had its days was evident. Lee's criticisms made a huge impact on the planning community, by the time when the modelling practice had acquired a rudimentary organization, and some tradition and the scientific approach applied to planning was making its first "incursions" in Europe (Batty, 1994). It is symptomatic that the number of papers published on large-scale modelling decreased dramatically, and for a long period of time the subject was, despite a hand full of works, practically put aside (Klosterman, 1994). And Lee's "*Requiem…*" still is, in our days, one of the most cited papers in the field of urban modelling.

One of the consequences of the social change of paradigms that occurred in the 1970s with the end of the post-World War II economical boom and with the early signs of weakness shown by the welfare state, was a shift on the planning horizon adopted by the majority of practitioners. The object of planning was re-centred on short term goals, on immediate solutions to problems, rather than to accomplish ambitious long-term strategic objectives, a characteristic of large-scale planning as it was proposed by the Chicago School, for example. It was the shift from planning to management (Batty, 1994).

Another crucial concept of planning had passed virtually untouchable during these turbulent times for the science of modelling: the vital need for large amounts of information. In the 1960s, alongside with modelling techniques, the management of large information systems were the cutting edge of scientific planning (Batty, 1994). The advent of sophisticated data base management systems provided powerful tools for planners as they could now process a larger amount of more disaggregate data in increasingly shorter amounts of time.

The next step that took place in the 1980s was the diffusion of the concept of geographical information systems (GIS). Although, for many years until the mid 1990s, these software tools were mainly used for cartography processing and mapping (Batty, 1994), the integration of new, sophisticated, built in modelling with GIS has provided new grounds for the planning activity (Takeyama and Couclelis, 1997; Wagner, 1997; Batty, Xie and Sun, 1999).

At present, new paths are being explored. Computer capabilities are now at a level which planners and modellers in the 1960s would consider almost science fiction. Whereas location and transportation were the key to the first generation of urban models, the dynamics of growth and diffusion phenomena at a fine scale are the subjects of the new generation of micro-scale modelling (Batty, 2004). Object-oriented programming brought new and powerful tools for modelling at a disaggregate micro-scale (Benenson, Omer and Hatna, 2002; Barros, 2005; Benenson, Aronovich and Noam, 2005; Semboloni, 2005). Micro-simulation is now a reality, supported by a series of techniques such as CA and agent-based simulation. GIS and data base management are two of the most developed areas in software, with a series of commercial products that easily and by lower costs provide the necessary data processing capacity. The developments in the last two decades both in data availability and computational capacity have created a big impulse on the use of models, clarifying their role as a scientific approach to the highly comprehensive planning process (Couclelis, 2005).

Models are shifting from a comprehensive perspective, the main assumption of the first generation of models, to a sketch-planning-type modelling, oriented for solving closely adapted local situations, standing out policy-oriented, practical goals rather than broad strategic goals, a characteristic of the former ones (Batty, 2004; Couclelis, 2005).

But some of the critics formulated in the 1970s remain present. There are some concerns that modellers still focus their main attention on model development rather than on the planning problem underlying the model. Urban simulation is as much a planning exercise in simulation as it is in planning sciences (Torrens and O'Sullivan, 2001). Couclelis stated that in all but trivial cases, the hope on good predictive models in the field of social phenomena is lost (Couclelis, 1997). There still is an apparent paradox on the fact that, as well intentioned modellers were creating land use models aimed to forecast future states of complex systems, these models helped little (or not helped at all) planners in their tasks (Couclelis, 2005).

2.2.2 The theory/practice dichotomy

The important notion of the conflict that exists in urban modelling between theoretical and practical perspectives, between the use of models and the practice of planning, is very well described by Couclelis (2005) (page 1359):

"Models are based on science; planning is about policy. Models are much better (…) at dealing with natural science problems; planning is mired in difficulties most often due to issues in the purview of social sciences. Models are usually developed from within particular disciplinary perspectives; planning must integrate across all domains. Models are about information and facts; planning is about interpretation and values. (…) Models codify uncertain knowledge; planning must lead to certain action. (…)"

This natural tension between theorists and practitioners, between modellers and planners was already mentioned in the previous section. This continuous tension between modelling and planning results from those general dichotomies mentioned before between science and policy, natural and social sciences, between analysis and synthesis, studying the past and preparing the future (Couclelis, 2005). Modelling was firstly considered the new grounds for a new scientific approach to planning (Batty, 1994; Wegener, 1994). Despite of the disturbance experienced by urban modelling throughout the past decades (or, in a more realistic perspective, since almost the beginning of its practice), the practice of planning was always considered intimately dependent of its theory, only varying the scale and the extent of dependence the later imposed to the former.

There still is a structural gap between planning theory and practice that only emphasizes the mistrust in the scientific approach made by the use of models. Modellers and planning support systems (PSS) developers must try hard on developing solutions that meet the planners needs (Couclelis, 2005). Therefore, applicability must be one of the major aims (if not the goal to achieve) of current and future studies in urban modelling, as it constitutes the means that will provide the needed validation to this scientific area. Modellers must accept that the role of scientific planning goes far beyond the implementation of forecasting models; at the same time, planners must find a balance between participation and systematic expertise (Couclelis, 2005).

The shift of paradigm from the pioneers of the 1960s to the current practice is mainly associated with the general perception that rational planning, aiming to understand and control the entire system, failed to accomplish the needs of more modest, incrementalist interpretation of planning (Wegener, 1994). In this context of integration of stakeholders, a major aim of current planning practice, the recent development of new methodologies for the integration of stakeholders' values, such as multi-criteria evaluation and the new Value Sensitive Design methodology (Friedman, Kahn and Borning, 2002) can improve the use of simulation in the planning process (Waddell and Ulfarsson, 2004). It is also very important to capacitate models to evaluate objectives that are stated in planning policies, even when they are somehow undefined (Waddell and Ulfarsson, 2004).

The integration of stakeholders and their value systems, along side with the fact that planning generates a variable set of goals and actions, imposes the consideration of uncertainty as a key factor for the success of the relationship modelling/planning (Couclelis, 2005). Three major roles for land use models are proposed by Couclelis in order to increase their mission to support planning: scenario writing (what may be), visioning (what should be) and storytelling (what could be) (Couclelis, 2005). Scenario writing is a notion that has its roots on modelling science, being one of its historical goals. Visioning is useful to integrate community interests and values in order to try to reach broad consensus on strategic matters. Storytelling can help to compare future desired and feared evolutions, in realistic terms that could effectively assist the planning process.

Since those uncertain times, a small but steady increase has been registered in the field of urban modelling. Time had tempered the experience and what had been accomplished between 1973 and the early 1990s could not be considered as failure (Batty, 2004). This increase promoted by the spectacular development of computational capacities, was studied by Wegener for the special issue of the Journal of American Planners published in 1994 (Wegener, 1994). Wegener assembled a list of groups and locals that were developing and implementing urban models of any kind with an operational perspective (see Figure 3).

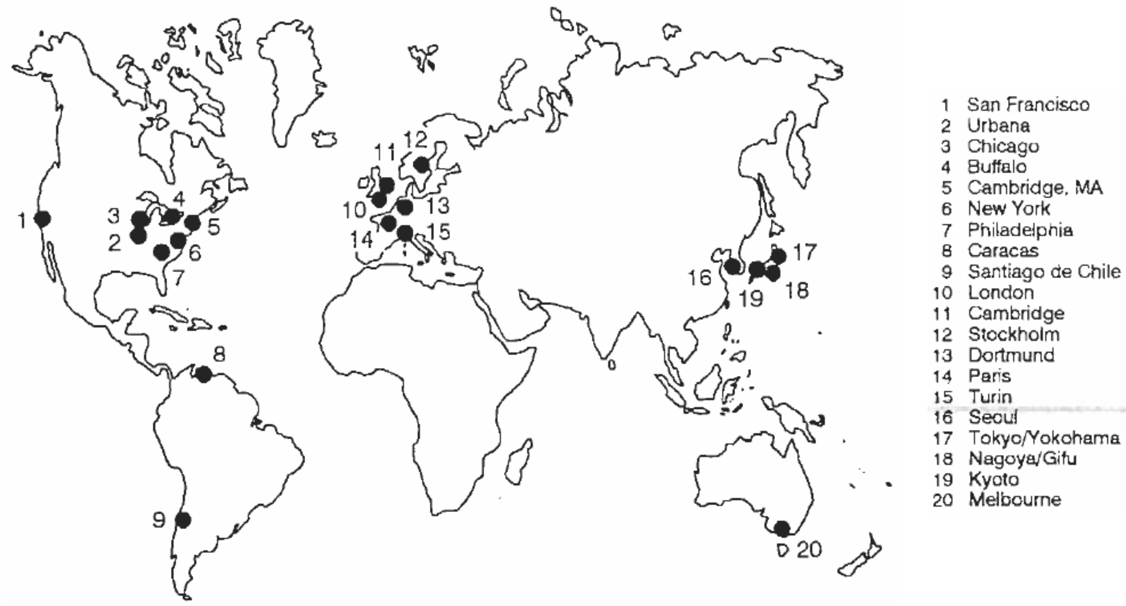


Figure 3 Map of active urban modelling centres (Wegener, 1994).

The criteria for choosing a particular group was: (1) that a mathematical model implemented on a computer and aimed to analyse past evolution and to forecast future urban scenarios should be the basis of the urban model itself; (2) that the modelling approach should have a comprehensive framework, integrating all the essential processes of urban phenomena; and (3) were excluded all those works that only presented theoretical work without any operational implementation, as valid as those works could be. It is interesting to notice that by that time, 20 years after Lee's "*Requiem…*" an important group of scientists was developing urban modelling and had constituted an informal scientific network, crucial for the strengthening and increasing gain in coherence of this scientific field (Wegener, 1994).

In the present, the number of active groups working on integrated urban models has reached the highest point, with a series of operational models implemented all over the world. Waddell, Bhat, Ruiter, Bekhor, Outwater and Schroer (2001) reports a series of ongoing work on the field: the UrbanSim framework is in operational use in the Puget Sound Region, in the state of Washington (Waddell and Ulfarsson, 2004); the Reusable Modelling Components for Land Use, Transportation, and Land Cover project is dedicated to the development of robust and modular set of modelling tools capable of being replicated in different urban and regional contexts. Miller, Hunt, Abraham and Salvini (2004) refers some integrated land use and transportation software packages that are already available, as MEPLAM (Hunt and Simmonds, 1993) and TRANUS (de la Barra, 2001), although these models are founded on aggregate approaches with strong equilibrium assumptions on several variables of the systems (Miller *et al.*, 2004). More recently, Klosterman et al updated the list of significant urban models that are currently being implemented (Klosterman and Petit, 2005). They assembled this list considering simultaneously the modelling approach of each work and the main task purpose by each model, which is summed up in Table 3 and Table 4

Table 3 Categorization of selected planning support systems (Klosterman and Petit, 2005)

Technique	Task			
	Land use/Land cover change	Comprehensive projection	3d visualization	Impact assess-ment
Large-scale urban	METROPILUS SPARTACUS TRANUS UrbanSim	METROPILUS SPARTACUS TRANUS UrbanSim		
Rule-based	CUF WhatIf?™1.1	WhatIf?™ 2.0	CommunityViz	CommunittyViz INDEX© Place^{3}S
State-change	CUF II CURBA			
Cellular automata	SLEUTH DUEM			

Complexity is also one of the major investigation areas currently under attention. Many authors identifies complexity as the key factor to understand urban phenomena (Benguigui, Czamanski, Marinov and Portugali, 2000; Batty and Torrens, 2001; Li and Yeh, 2001; Barredo *et al.*, 2003; De Keersmaecker, Frankhauser and Thomas, 2003).

Cities, like several spatial phenomena, are complex systems. “The mixture of different urban activities creates a logic of its own, but a logic nonetheless. As almost all cities presents this complexity, it is reasonable to suppose that complexity is somehow one of its essential qualities” (White and Engelen, 1993).

Complexity expresses itself through spatial scale: from local scale behaviours of individuals (vehicles or people) emerge structured and ordered patterns in aggregate large scale. In fact, the hole concept of complexity hinges on the notion of emergence (Torrens, 2000). One of the goals of complexity studies is to derive universal laws of complex systems from common principles based on simple features (Torrens, 2000).

However, many processes in the natural world may not be deduced in universal laws capable of granting a decent theoretical basis of complex systems (Casti, 1997) cited by (Torrens, 2000).

Table 4 Information sources for selected planning support systems (Klosterman and Petit, 2005)

Model	References	Web-site
CommunityViz	(Kwartler and Bernar, 2001)	http://www.communityviz.com
CUF, CUF II and CURBA	(Landis, 2001)	
DUEM	(Xie, 1996)	http://lgre.emich.edu
INDEX©	(Allen, 2001)	http://www.crit.com
METROPILUS	(Putman and Shih-Liang, 2001)	
Place^{3}S	(Snyder, 2001)	http://www.energy.ca.gov/places
SLEUTH	(Clarke *et al.*, 1997; Silva and Clarke, 2002)	http://ngcia.ucsb.edu/projects/gig
SPARTACUS	(Latuso, 2003)	http://www.fhwa.dot.gov/planning
TRANUS	(de la Barra, 2001)	http://www.modelistica.com
UrbanSim	(Waddell, 2001)	http://www.urbansim.org
WhatIf?™	(Klosterman, 2001; Klosterman, Siebert, Hoque, Kim and Parveen, 2003)	http://www.what-if-pss.com

The best way to characterize a complex system is identifying the states it can take and the conditions to take them. This can be easily understood with a simple example based on CA. Considering a system with *n* elements (for instance cells) describing a particular state, each state described by a binary existence (say developed) or otherwise for each element, then there are 2^n distinct states. If a lattice of some hundreds cells is considered, with a wider set of rules for transit between those states, then conventional theorizing can not describe the problem (Batty and Torrens, 2001).

Complex systems include two key elements (Batty and Torrens, 2001): first, "system extensiveness" along any spatial, temporal or topical dimension[3], as complex systems can not be reduced or aggregated without lost of their structure; second, process, meaning that space and time dynamics suffers unexpected changes, often followed by emergence.

2.2.3 The microsimulation approach

In Section 2.3 the evolution of urban modelling throughout the past decades was shortly described, since the early efforts from the 1950s to the new paths based on powerful and inexpensive computation. This evolution kept up with the evolution of computer science, benefiting with the new calculus resources provided by computer.

[3] The fact that complex systems are allowed to evolve without the constraints of reductionism runs directly against the traditional scientific paradigm of analysing the essence of phenomena to deduce theory.

The transition from the early large scale modelling phase, based on equilibrium assumptions and deterministic approaches gave place to the fine scale based modelling supported by new scientific approaches. Microsimulation, which was developed in the 1960's, was only applied to urban modelling later in the 1980's. Since then, the development of discrete choice modelling and the emergence of cellular automata and multi-agent simulation techniques have created a proliferation of modelling approaches (Waddell and Ulfarsson, 2004).

Waddell and Ulfarsson (2004) presents a series of preliminary model design choices that must be considered in urban modelling. These assumptions establish the difference between macrosimulation and microsimulation, as they set up a series of orientations that are thought to adjust the simulation to the reality they aim to simulate.

The first choice regards behavioural resolution. Systems can be considered working at an aggregate scale of average behaviour or they can be assumed working at a disaggregate level, based on individual agents. Secondly, the simulation must be based on deterministic or on stochastic behaviour: deterministic models are commonly used along with aggregate scale of behaviour since the average behaviours can be easily approximated with fixed rates of change. Finally, issues related with the resolution of agents, space and time must be pondered. Simulation systems range, in general, from macroscopic resolution to microscopic resolution.

Macroscopic systems have larger units of analysis, dealing with aggregate information both spatially and statistically and they are essentially static and deterministic. The low consumption of data and computational resources make macroscopic simulation one of the most widely used approaches (Waddell and Ulfarsson, 2004).

Microscopic models have, in opposition to the previous scale, small units of analysis. These scale level of modelling is the strongest beneficiary of the evolution of computation throughout the last twenty years. As computational resources evolved and become less expensive and faster, microscopic models started to increase the amount of data processed and to deepen the resolution of the models, with a consequent increase in their feasibility. These models have a stochastic behaviour as they support clearer behavioural specifications (Waddell and Ulfarsson, 2004).

In between these two modelling scales there is an intermediate mesoscopic scale. This term is essentially used to classify models that integrate characteristics of both macroscopic and microscopic models. They can present large analysis units with small time steps or they can use aggregate data for some aspects of the model at the same time they use detailed information for other aspects (Waddell and Ulfarsson, 2004).

Most geographic theories are static where rational actors were assumed to interact in a market that remains in a state of equilibrium. This is not a reasonable way to describe a city, which common sense and experience tell us is rarely if ever in an equilibrium state. Almost all cities are undergoing continual growth, change, decline and restructuring, usually simultaneous (White and Engelen, 1993).

The assumption that urban systems are better represented by dynamic, stochastic, high resolution models along with impressive developments on computation made microsimulation the most fitted approach for dealing with these issues. Applications based on microsimulation are being developed both in theoretical and operational perspectives in different areas of urban sciences. Transport systems analysis produced a series of real-time applications, based on individual agents (see Miller *et al.*(2004) for a brief list of models). Integrated land use models are also being developed in the last few years using microsimulation, such as UrbanSim, SLEUTH and WhatIf? operational models (see Table 3 on page 30).

Although microsimulation was introduced in the 1960s, it was only in the 1980s that it started to be used in urban modelling (Waddell and Ulfarsson, 2004).

The early modelling approaches were based on techniques as spatial interaction, spatial input/output and linear programming. Spatial interaction is founded on the gravity model applied to model trip destination choices or residential and employment location. These models are limited in the degree of spatial detail used and do not represent the behavioural factors influencing the phenomena they try to simulate, especially market and prices behaviours (Waddell and Ulfarsson, 2004). Spatial input/output models are an extended application of the input/output model of the US economy presented by Leontief to represent spatial patterns of location for economic activities and people and goods movements between zones (Leontief, 1966). This approach has the merit of including explicit real estate and labour markets, as well as travel demand, but it still considers different states of equilibrium for changes in the model inputs (Waddell and Ulfarsson, 2004). Another technique used in early models, although less often, is linear programming. This technique is focused on the optimization of an objective function and it is more suited to the exploration of land use alternatives that optimize some urban function (such as travel cost), than to simulate complex and realistic behavioural responses to input changes (Waddell and Ulfarsson, 2004).

Microsimulation as a modelling approach essentially implies the use of individual-level scales. Waddell and Ulfarsson (2004) describes the most important microsimulation techniques currently under use.

Discrete choice modelling is a standard method whenever the behaviour of individuals (households, people, and trips) is modelled, particularly after the publication of the Random Utility Theory by Daniel McFadden (Waddell and Ulfarsson, 2004). Models such as logit and nested logit are frequently used to predict individual choices among finite sets of alternatives, a very common goal on travel demand and mode choice modelling.

There are several land use models developed in recent years using GIS and a rule-based set of procedures to allocate population, employment, and/or land use. Examples include the CUF model (Landis, 2001) and WhatIf? (Klosterman, 2001). These applications may have a useful role in making models more accessible, but there is a risk that model users would interpret the models as having a more behavioural basis than their rules actually contain (Waddell and Ulfarsson, 2004).

Multi-agent simulation (a generalization of agent-based simulation) is another simulation method available that works on a disaggregate level. It draws upon complex systems theory (as CA do), focusing the modelling on the emergent systems behaviour arising from the interactions between agents (Waddell and Ulfarsson, 2004). This techniques has been the object of intensive research since the creation of the Swarm software environment for implementation of models of this type (Swarm, 2002). There is extensive ongoing research on this methods with promising results (Barros, 2005; Benenson *et al.*, 2005; Semboloni, 2005).

CA modelling has emerged from the field of complex systems theory as a means of representing emergent properties derived from sets of simple behavioural rules operating over a cell-based pattern. This approach was introduced to spatial problems by Waldo Tobler in his ground breaking work "*Cellular Geography*" (Tobler, 1979) and has been widely applied since then in many land use and land cover change problems (Couclelis, 1985; White and Engelen, 1993; Batty and Xie, 1997; Clarke *et al.*, 1997; Couclelis, 1997; Takeyama and Couclelis, 1997; White and Engelen, 1997; Candau, 2000; Torrens, 2000; Li and Yeh, 2001; O'Sullivan, 2001b; Silva and Clarke, 2002; Wu, 2002; Barredo *et al.*, 2003; Ward *et al.*, 2003; Benenson and Torrens, 2004).

CA presents many advantages for modelling urban phenomena. It is a decentralized approach, it provides a link to complexity theory and it makes the connection of form with function and of pattern with process. It has good visualization characteristics, it is a flexible and dynamic approach and, more important, it is based on a set of very simple elements (Torrens and O'Sullivan, 2001).

CA also have advantage when facing the traditional simulation approaches based on differential or difference equations: it is inherently spatial, with rule-base dynamics, with a much higher computational efficiency which means that dynamics can be modelled with very high spatial resolution (White and Engelen, 1993; Batty *et al.*, 1997).

But these advantages yields simultaneously its handicap: CA models are constrained by their own simplicity and their ability to illustrate real world phenomena is often diluted by their abstract characteristics (Torrens and O'Sullivan, 2001).

2.3 Cellular automata

2.3.1 The concept of cellular automata

After introducing CA as a microsimulation technique that is gaining followers for the last twenty years within the urban science community, it is important to present a brief history of CA as well as its formal mathematical structure.

CA were first introduced in the 1940s by John von Neumann, the founder of game theory, and Stanislaw Ulam, who worked in the Manhattan Project and made intensive research in the field of Monte Carlo simulation. Both researchers were dedicated to studying self-reproduction and to modelling biological life, trying to devise a mathematical formulation that could reduce the forces governing reproduction to logical rules (Torrens, 2000). The concept of automatic computation is on the very foundations of CA, as von Neumann was strongly influenced by Alan Turing's Universal Turing Machine, a theoretical concept of a computer capable of produce universal computing[4] (Wolfram, 2005). This Machine would be able of taking new actions in a given moment of time based both on a set of parameters and on its knowledge about the external world obtained from an input system of any kind[5].

An automaton can be considered a Turing-like machine that operates over a tessellated cell space. The term automaton refers to a self-operating machine that "*processes information, proceeding logically, inexorably performing its next action after applying data received from outside itself in light of instructions programmed within itself*" (Levy (1992) cited by Torrens (2000)). However, there is a great difference between CA and Turing's Machine: CA are parallel processors, with a series of rules working at the same time while the Turing Machine is a serial processor that can only handle a process after another (Torrens, 2000).

According to its first formulation for a 2D cellular automata by von Neumann, presented by Benenson and Torrens (2004)), each automaton cell processes information and proceeds in its action considering both a set of transition rules and data received from the environment. Formally, each cell A is defined by a set of cell states $S = \{S_1, S_2, ..., S_N\}$ and a set of transition rules T

$$A \leftarrow (S, T) \qquad (1)$$

Transition rules define an automaton state S_{t+1} at time step $t+1$ depending on its state at time step t, S_t ($S_t, S_{t+1} \in S$) and an input information I_t:

$$T : (S_t, I_t) \rightarrow S_{t+1} \qquad (2)$$

A grid of automata becomes cellular automata when the input is defined by the states of neighbouring cells. This definition of neighbourhood is basically the set of cells that influences the state of the cell under consideration.

$$A \leftarrow (S, T, R) \qquad (3)$$

[4] "A system may be regarded as a universal computer if, given a suitable initial program, it is capable of implementing any finite algorithm through its evolution over time, i.e., that it is capable of producing a working copy as complicated as itself, and the means to make further copies" (Torrens, 2000).

[5] In the original formulation the input was an infinitely long tape ruled off into sections where binary information could be read or wrote.

where R denotes automata neighbouring A and establishes the boundary for drawing input information I which is necessary for the application of transition rules T. Classic neighbourhoods such as the von Neumann neighbourhood and the Moore neighbourhood are commonly used in 2D cellular automata. The first one is a set of four adjacent cells usually located in the main cardinal points. The second one is the entire set of eight adjacent cells. These neighbourhoods are depicted in Figure 4.

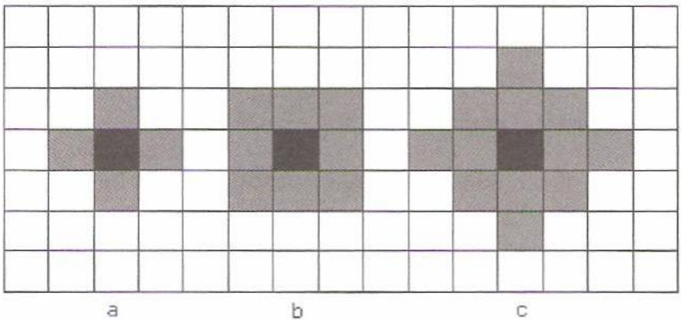


Figure 4 Classical von Neumann's and Moore's neighbourhoods for 2D CA

The complexity of cellular automata increases as the number of cells increases. Even for few states and small-size neighbourhoods, the possible number of transitions is high. Being N the number of cell states and K the number of cells in the neighbourhood, the total number of transition rules equals N^K (Wolfram, 1983). If the Moore neighbourhood (the centre cell and its eight neighbours) is considered for a set of binary states in a 2D CA, the number of different transition rules equals the expressive number of $2^9 = 512$. Von Neumann considered 29 different cell states in his groundbreaking work.

There are two main moments in the history of CA after the introduction of complex systems theory: the discussion over John Conway's Game of Life in 1970 and the work of Stephen Wolfram in the 1980s (Wolfram, 1984). Conway's Game of Life was aimed to design a simple set of rules to study microscopic spatial dynamics of population (Benenson and Torrens, 2004). This 2D CA were based on a set of two cell states, alive (1) or dead (0) and on three transition rules: Survival (Figure 5 a), Death (Figure 5 b) and Birth (Figure 5 c). The cell remained dead if none of these rules were applied (Figure 5 d).

From this simple set of rules, the Game of Life could generate extremely complex patterns of growth through simulation. Furthermore, it could be easily replicated following the concept of the universal Turing Machine. Its main achievement was to formulate a simple interdisciplinary tool for representing complex spatial systems and to modelling its dynamics (Benenson and Torrens, 2004). Conway pursued a configuration that could generate moving configurations of stable patterns. After Gardner's challenge on the pages of Scientific American for the research of such a configuration, Robert Gosper and his team at the Massachusetts Institute of Technology implemented a version of

Life-CA that generated a machine that could reproduce copies of itself that were as complicated in their structure (Torrens, 2000). Considering the Game of Life as the first widely known CA, and relating its classification as a game to represent the world evolution, it is interesting to consider CA like the world, in its ability of representing the most complex forms of evolution from simple, well-understood interaction rules (Couclelis, 1997). It is worth stating the procedures of Conway's Game of Life as they embodies the key elements of CA (Batty *et al.*, 1997).

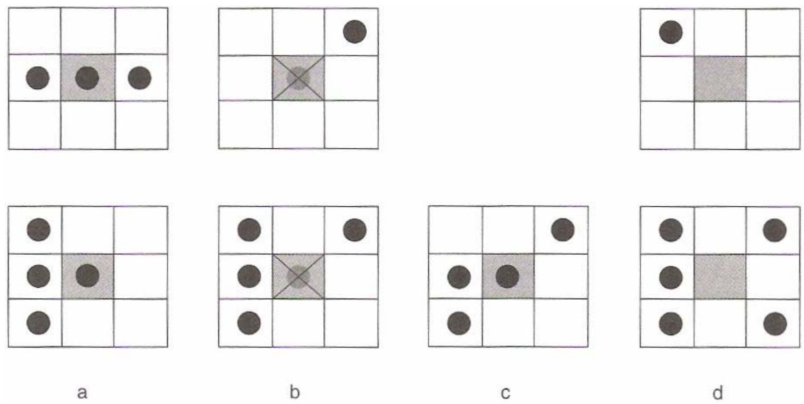


Figure 5 Conway's Game of Life set of rules (Benenson and Torrens, 2004)

The introduction of CA as a suitable tool for studying complex systems dynamic made possible through the discussion over Conway's Game of Life, the research on the limits of system's spatial patterns began. The mathematician Stephen Wolfram made extensive research on CA exploring different configurations it could evolve to from simple sets of parameters and rules. Using a simple 1D CA, where the neighbourhood is the set formed by the cell and its adjacent neighbours (a set of 3 cells), Wolfram studied limiting patterns for all the possible 256 configurations (2^8 configurations) (Wolfram, 1983, 1984).

Table 5 Wolfram's classification for 1D CA behaviour

Class	CA dynamics evolves towards	Type of system dynamics
I	Spatially stable pattern – each cell reaches the stable value of "0" and "1"	Limit points
II	Sequence of stable or periodic structures – each cell changes its states according to the fixed finite sequence of "0"s and "1"s	Limit cycle
III	Chaotic aperiodic behaviour – the sequence of cell state is not periodic, but the spatial patterns repast themselves in time	Chaotic attractors
IV	Complicated localized structures, which are sometimes long-lived and are more complex than those of Class III	Attractors unspecified

Based on numeric experiments, Wolfram demonstrated that limiting configuration of CA does not depend on initial state of cells but it is defined by transition rules. Wolfram classified CA according to their dynamic behaviour and the patters generated by them, identifying four main classes described on Table 5 and depicted in Figure 6.

Several other studies tried to devise new classifications by extending the study of Wolfram's classes to larger neighbourhoods and different set of rules and parameters. Wolfram's classification has the disadvantage of being impossible to classify a given CA based on the transition rule alone, not knowing what the spatial pattern would be (Benenson and Torrens, 2004). However, Wolfram's classification still remains the most popular one.

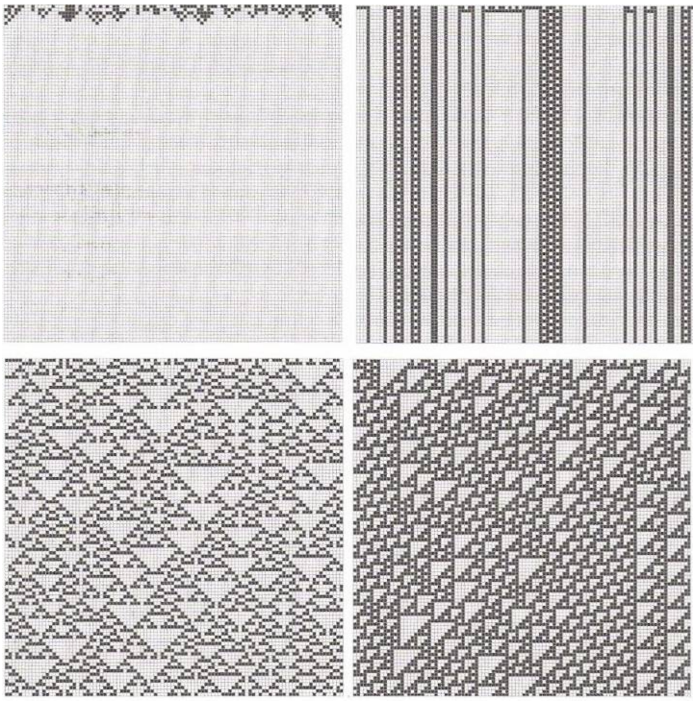


Figure 6 Typical CA behaviour for each of four Wolfram's classes (clockwise from left superior corner: Class I, Class II, Class IV and Class III) (Benenson and Torrens, 2004)

It is also important to refer the classification introduced by C. Langton which is very interesting for geography and urban studies (Benenson and Torrens, 2004). He introduces the concept of inactive cells that cannot change its state during CA evolution. It is based on a fraction λ of configurations that does not lead the cell to the inactive state (for a more detailed explanation see Benenson and Torrens (2004)). Although a series of new classifications were proposed, several studies on higher-dimensional CA show that their classifications remain similar to Wolfram's classification for 1D CA (Benenson and Torrens, 2004).

2.3.2 The use of CA in urban studies

Most geographical theories are static, assuming that the interactions of their agents take place in a market that remains in a state of equilibrium; this assumption is far from being reasonable, as all cities are continually undergoing changes. This fact makes imperative the use of dynamic models, based on the processes that occur in the territory (White and Engelen, 1993). The great attraction of CA is the fact that many classes of system dynamics can be simulated through it; another important feature of CA is its ability to give equal weight to the importance of space, time and system attributes (Batty *et al.*, 1997). The natural ability that CA have to represent complex systems with spatial/temporal behaviours from a small set of simple rules and states made this technique very interesting for geographers and urban researchers. CA are intrinsically spatial and they are used to model a wide range of phenomena due to its ability to represent spatial process, from forest fires to epidemics, from traffic simulation to regional-scale urbanization, polycentricism, gentrification, historical urbanization, urban growth, form and location (White and Engelen, 1993; Torrens and O'Sullivan, 2001). CA-based modelling also allows the integration of socio-economic and natural systems models in a detailed and realistic way (White and Engelen, 1997).

There are three main classes of urban CA models, for three different purposes, not mutually exclusive, that are a direct result of the exploration of modifications of the formal CA: (1) models designed to explore spatial complexity, (2) models designed to research themes of economical, sociological and geographical areas; and (3) models designed to produce operational tools for planning (Torrens, 2000). In this section it will be presented three main studies on the exploration of CA from a theoretical perspective both on its formulation and on spatial complexity. In Section 2.3.4 a series of other models will be presented and discussed.

From the 1960s until the mid 1980's geographers were more focused on the study of comprehensive regional models (Benenson and Torrens, 2004). The critics published by Lee (1973) in his "*Requiem…*" established a transition point where the modelling community started to question the necessity for large scale models and shifted to small scale problems. Problem complexity, even when a small set of spatial relationships was used for modelling purposes, and the lack of experimental data for model calibration generated a search for simpler approaches that could produce more reliable simulation (Klosterman, 1994). CA presented an opportunity to deal with these new modelling requirements. Tobler and Couclelis had produced the grounds for this cellular approach, in the 1970s and 1980s, when it really had a major development along with computer graphics, fractal theory, chaos and complexity (Batty *et al.*, 1997).

Waldo Tobler presented in 1970 his study on population growth simulation based on a cellular model (Tobler, 1970). In this study he invokes the first law of geography[6] to overcome complexity when dealing with the interdependence of population growth in a given location and its dependence on the population growth everywhere else in the world. Instead of using trend equations with limited set of coefficients that could surrogate the real behaviour of growth in every location, he uses an approach that considers the influence on the population growth limited to a neighbouring set of locations. Later in 1979 Tobler publishes another important study where CA are explicitly considered to the study of spatial phenomena (Tobler, 1979). Tobler makes a classification of different cell models for land use simulation, taking into consideration its dependence on spatial and temporal issues (Figure 7).

Model type I is an independent model, where the land use for a given location $g_{i,j}$ at time $t+\Delta t$ is not related with the situation at time t. Model type II is classified as functionally dependent because land use at a given cell $g_{i,j}$ at time $t+\Delta t$ depends on the land use at that location at time t. Model type III is named a historical model as land use of a given location $g_{i,j}$ at time $t+\Delta t$ depends on a series of previous states at previous time steps. Model type IV is classified as multivariate because it depends on several variables other than land use at that location i,j. Finally, model type V is classified as the geographical model because land use at a given location $g_{i,j}$ at time step $t+\Delta t$ is dependent on the land uses of a given neighbourhood at time step t.

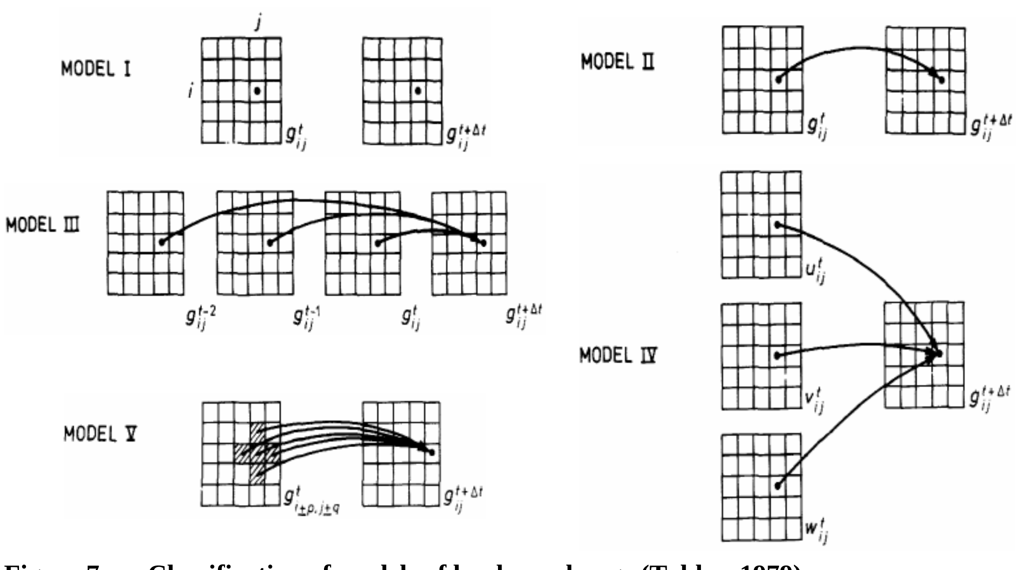


Figure 7 Classification of models of land use change (Tobler, 1979)

Tobler develops his study using a model type V which he classifies as a dynamic one, because the land use of a given location g_{ij} at time step $t+\Delta t$ is a function F of land

[6] The first law of geography states that *everything is related to everything else, but near things are more related than distant things*. This is an important concept that supports the link between CA and geography because neighbourhood is one of the main components of CA.

use at that location at time t and of a measure of all the land uses in neighbourhood n at that same location as expressed in Equation 4:

$$g_{ij}^{t+\Delta t} = F(g_{ij}^{t}, n_{ij}) . \tag{4}$$

The neighbourhood n is based on the traditional von Neumann neighbourhood and is defined as a geographical domain of influence. He considers that, because of different notions of neighbourhood different residents have, it should be possible to have dynamic neighbourhoods varying on size, shape and orientation.

Transition rules F are defined closely to formal CA definition. Tobler is more interested in studying geographical features of transition rules such as spatial isotropy and spatial stationarity. By spatial isotropy it is meant that the positioning of neighbours does not have any influence in the transition rule. By spatial stationarity it is meant that the same environment, the same neighbourhood, results in the same consequences, this is, the rules do not depend on where you are. These conclusions are, of course, very generic and considerably distant from real world geographic behaviours (Tobler, 1979).

Tobler's groundbreaking work was followed by a series of other studies that explored CA and its application to geography (Couclelis, 1985; White and Engelen, 1993). They all refer Conway's Game of Life as a paradigm of the application of CA to complex systems from a set of simple transition rules, but never as an example of a geographic system; they all were somehow far from the formal CA definition of von Neumann (Benenson and Torrens, 2004).

In the mid 1980s Helen Couclelis continues the work started by Tobler in the research of CA for urban modelling purposes (Couclelis, 1985). Her work has two main lines: the first one relates to the conceptual and theoretical linkage with the theory of complex systems; the second one relates with the exploration of possible uses in urban planning (White and Engelen, 1993). She establishes a parallelism between Conway's Game of Life transition rules and urban changing phenomena. A 'live' cell could be interpreted as an urban zone that exceeds some kind of threshold in terms of a given urban function. Transition rule 1 (survival) would then guarantee that this cell would maintain this urban function if two or three neighbouring cells also exceeded that threshold. Transition rule 2 (death) would make this cell loose that urban function if four or more neighbouring cells also exceeded that threshold because the cell would suffocate or, on the other hand, if there was only one neighbouring cell exceeding the threshold (then the cell would die from loneliness). Transition rule 3 (rebirth) would settle this urban function in a cell if there were exactly three neighbouring cells that exceeded the threshold. A cell would remain dead if none of the precedent conditions were verified.

However, the inherent simplicity of this formulation makes it inappropriate to simulate real world spatial phenomena. Couclelis identifies a series of shortcomings that limits the ability of cellular models to simulate urban phenomena. Issues regarding space dimension and boundary treatment are pointed out as the first limitations for modelling cellular worlds. The imposition of universal transition rules to an infinite regular cell space collides with an empirical interpretation of the method. A second limitation is related with the regularity of neighbourhoods. In order to be representative, neighbourhoods should be defined by different shapes and sizes for different cells. Cell regularity and spatial homogeneity within each cell are also issues that make strict CA formulation far from being geographically representative. Finally, it is also important to notice the assumption of space and time invariance for transition rules as well as the closure of the system to external perturbations (other than stochastic behaviours) (Couclelis, 1985). The solution relies on the consideration of relaxations that would enhance the ability of a cell-space approach to simulate real world phenomena without loosing both its identity and its simplicity.

Couclelis reformulates the structure of Conway's Game of Life to obtain a simple and strict geographic CA formulation, a generic structure that can be applied to a series of spatial problems (see Couclelis (1985) for further information). She also states that this formulation can be easily relaxed in order to enhance its simulation capabilities. Couclelis' formulation is independent from cell shape and size and even from the existence of a cell space itself. The cell space concept can be generalized to the point where it describes any discrete time/space model representing components and their interactions (Zeigler (1976) cited by Couclelis (1985)). It is likely that any model of interest to geographers, whether aggregate or disaggregate, continuous or discrete, deterministic or stochastic, quantitative or categorical can be expressed in the same language as the cell-space concept (Couclelis, 1985).

Another important concept was introduced by Roger White and Guy Engelen (White and Engelen, 1993): the concept of constrained CA. Standard CA are modelled with the explicit intention of maximizing their generality and as a result they present two main characteristics: (1) they are defined on a homogeneous cell space, overriding the variability of cell characteristics as they exclusively relate cells with their state values, regardless of their location on the cell grid; (2) they are unconstrained, so that the number of cells in each state is determined endogenously by the application of transition rules to the current configuration of cells (White and Engelen, 1997). In their initial work, White and Engelen (1997) make some important assumptions that can be considered an approximation of CA to more likely urban behaviours.

The first assumption is related to the consideration of a small number of hierarchically ordered cell states. There are four cell states, vacant (the lowest state), housing, industrial and commercial (the highest one). A cell in the vacant state can change to any other state but the inverse is not possible (thus the city can only grow, which is unlikely to happen). A cell in the industrial state can only change to commercial state.

Another assumption is related to the use of larger neighbourhoods than the common Moore or von Neumann ones. The interaction between cells is then dependent of a larger area of influence, breaking up with the formal concept of local neighbourhood and introducing what is commonly named action-at-a-distance. On a strict CA model the changes of state must be local, discarding any action-at-a-distance (Batty *et al.*, 1997). In this case, these relationships are assumed non-monotonous and may present positive values (meaning an attractive relationship) or negative ones (repulsive relationship); theses relationships are depicted in Figure 8.

Finally, transition rules are not based on a simple probability of change but in a composite measure of transition potential. This transition potential is a weighted sum that simulates the behavioural propensities of the actors who determine land use, as expressed in Equation 5:

$$P_{ij} = S\left(1 + \sum_{h,k,d} m_{kd} I_{hd}\right) \tag{5}$$

where P_{ij} is the transition potential from state i for state j, m_{kd} is the calibration parameter applied to cells is states k at distance d, h is the index of cells within a given distance zone, I_{hd} is equal to 1 if the state of cell h is k and 0 otherwise, and S is a stochastic perturbation. The stochastic perturbation is calculated through the following expression:

$$S = 1 + \left[-\log(R)\right]^{\alpha} \tag{6}$$

where R is an uniform random distribution in the interval]0,1[and α is a control parameter for the adjustment of the size of the perturbation. This term has a highly skewed distribution so that most values are near unity and much larger values occur only infrequently. The weighted sum in the transition potential is multiplied by S in order to simulate the stochastic behaviour of agents in each component of the transition potential.

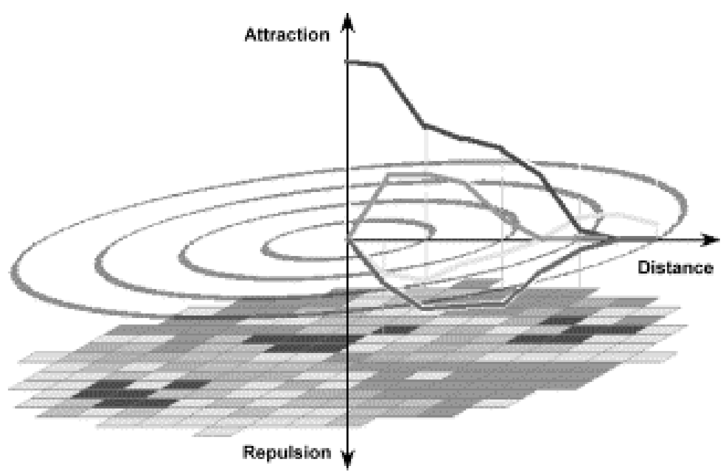


Figure 8 Neighbourhood interactions (Straatman, White and Engelen, 2004)

Another important evolution from the early CA models is related with the introduction of land use demand as an exogenous factor. In classic CA, the model evolves over time from a small set of cells to more and more complex systems through transition rules. The number of changeable cells is determined only by the change process itself. White *et al.* (1993) introduced an exogenous parameter that aims to simulate land use demand: there is a predetermined number of cells for each state that is considered in the transition procedure. If the total number of cells is not attained for a given state, there is a probability (which is affected by another calibration parameter) that converts cells in lower cell states to the given cell state until the demand is fully satisfied.

The transition potential is calculated for every cell at every time step of the simulation and is ordered in a descending manner. The set of cells with higher potential is chosen for transition considering land use demand. This approach has been developed and improved and was used in recent studies (White and Engelen, 2000; Engelen, White, Uljee and Drazan, 2002; Barredo *et al.*, 2003). Land use usually depends on three main types of factors: the inherent qualities of land itself, the effects of neighbouring land uses and the aggregate level of demand for each land use (White and Engelen, 1997). These components were first explicitly considered by White *et al.* (1993).

Another innovation introduced by this study is the consideration of fractal measures to assess model performance and simulation results. Several fractal measures are calculated for the set of theoretical cities generated by the model and then they are compared with known fractal measures from real world cities such as Cincinnati, Atlanta and Houston. The fractal dimensions obtained for the theoretical cities are in line with the values observed for the set of real world cities, which indicates CA's ability to produce realistic simulations of urban growth.

The three studies that were presented in this section are considered of great importance to the development of urban CA. They marked the scientific field because they explored CA as technique applied to geographic and urban studies. They also introduced a series of innovations that have successively enhanced the possibilities of using CA as an urban modelling technique.

There are a series of more recent applications of urban CA that are also important to the history of CA and urban modelling. In Section 2.3.4 these applications will be presented and their main characteristics and innovations will be discussed. However, it is important to first discuss the main evolutions of CA that are essentially based on relaxations of its four components. The great majority (if not all) of these applications introduced important innovations on urban CA that were aimed to improve its ability of simulating complex urban phenomena. In the next section these relaxations are discussed.

2.3.3 Main relaxations and evolutions

It was already mentioned that CA have in its simplicity one of its great attractiveness for urban modelling. From a simple set of rules operating over a simple cell structure it is possible to achieve complex forms from simple structures (Torrens, 2000). But the classic formulation of CA is limited in its own formal definition, imposing the necessity of relaxing some of its most basic components. It can be said that the notion "cellular automata" was used in geography, from the beginning, in a very broad sense and not as a rigid formal scheme (Benenson and Torrens, 2004). The formal framework of CA defined by its most known researchers as Ulam, von Neumann, Conway and Wolfram, based on a very simple formulation, is far from being able to represent real cities; many, if not all, urban CA bear little similarity with classical CA. It is questionable whether urban CA still are evolutions of classical CA or are just cellular-based models (Torrens and O'Sullivan, 2001). There are four major adaptations of the strict CA concept: (1) most applications to urban systems relax the neighbourhood effect to incorporate action-at-a-distance (considering these as a global consequence of local spatial diffusion, thus reinforcing the application of strict CA); (2) it is hard to identify a scale for urban systems where everything is reducible to one activity in one cell; (3) the need of CA to meet plausible values of change rates; and (4) the use of GIS and map algebra (Batty *et al.*, 1997).

Helen Couclelis pointed out two main characteristics that CA-based models must possess: interactivity and realism (Couclelis, 1997). Interactivity is an essential property that CA-based models must observe. They must allow the evaluation of small changes in the model conditions, to make sensitivity analyses possible (moreover being CA models of complexity, in which a small change in initial conditions or transition rules may produce major changes in the results). This type of models should also incorporate good visualization techniques, not only in the graphic output but also in the statistical characterization of the results. The integration with GIS is pointed out as the next big step, as a result of the natural affinity between CA and raster images, and because of the potential of graphical visualization provided by GIS. However, the more complex is the CA-model structure, the more visualisation becomes difficult.

Another important feature of CA-based models is realism. Couclelis argues that no model based on classic assumptions of CA – homogeneity, uniformity, universality – can claim good performance when applied to real world problems (Couclelis, 1997). There are two main dimensions of model realism: realism with respect to data and realism with respect to model structure. The linkage between CA and GIS once again can be very profitable in order to guarantee data realism. Several CA-based models are already supported by GIS features (Takeyama and Couclelis, 1997; Wagner, 1997; Batty *et al.*, 1999; Putman and Shih-Liang, 2001).

Structure realism is achieved through the adaptation of the classic CA structure to the specific needs of simulating complex urban behaviours. CA ability to generate complex patterns from simple cell configurations and through small sets of rules is one of its main attractions, as it was already stated. But in order to generate feasible urban landscapes it is necessary to adapt the formalism of CA to the behaviours that are the subject of this specific area of simulation.

Couclelis (1997) identifies a series of relaxations that can be implemented in order to enhance the ability of a CA-based model to correctly acquire the behaviours of urban areas in general. Figure 9 depicts a series of relaxations that can be implemented for all the four main components of CA.

The majority of CA models presented so far are based on regular square cells forming an orthogonal cell lattice. The main reason is related with land use data availability from remote sensing maps. In order to enhance spatial representativeness it is possible to forget this formality and to use irregular cells as the spatial unit for CA. Tobler refers that there are some analytical advantages in considering the irregular spatial division of political jurisdictions (Tobler, 1979). However, Tobler states that the basic difficulty, of topological type, relies on the fact that these irregular cells do not all have the same number of adjacent cells, thus its neighbourhood can not be defined by any simple notational scheme. There are very few studies based on irregular cells (one of the main proposals of the present study). The model developed by Vandergue *et al.* (2000) is cited by Ménard and Marceau (2005) as the only CA model that uses census tracts as cells. Semboloni (2000) used Voronoi polygons to model urban growth allowing that a spatial partition of a given cell could occur, which invalidates the use of regular cells. O'Sullivan (O'Sullivan, 2001b; 2001a) developed another CA-based model integrating CA and graph theory. Benenson and Torrens (2004) refers the existence of CA theoretical studies that used triangular and hexagonal cells (Gerling, 1990; Eloranta, 1997). These attempts showed that no significant improvements were introduced in the performance of CA model.

The cell state set can also be modified to simulate different land uses in different regions of the territory under consideration. This modification has some implications not only in the definition of neighbourhoods but also in the establishment of the set of transition rules.

The concept of neighbourhood is also susceptible of being modified. There is a problem or representativeness in the first place. The importance of the neighbourhood is that it defines the geographical domain of influence (Tobler, 1979). The fact that there are different concepts of neighbourhood must be taken into account: a resident of an urban area has a different idea of neighbourhood from a resident in a rural area, therefore it is possible to consider the shape, size and form of a neighbourhood as a function of the location of its central cell, giving a special attention to neighbourhoods in boundary areas (Tobler, 1979). Neighbourhoods such as von Neumann's and Moore's are particular indicated to physical phenomena; human interactions are more likely to be explained by wider areas of influence (White and Engelen, 2000). Neighbourhood structure may also vary in time. There are some studies in the field of tissues and leaves growth that considered the creation of new cells within a given neighbourhood (Lindenmayer, 1968). Semboloni (2000) also assumed that cells could be divided, thus creating new neighbourhoods.

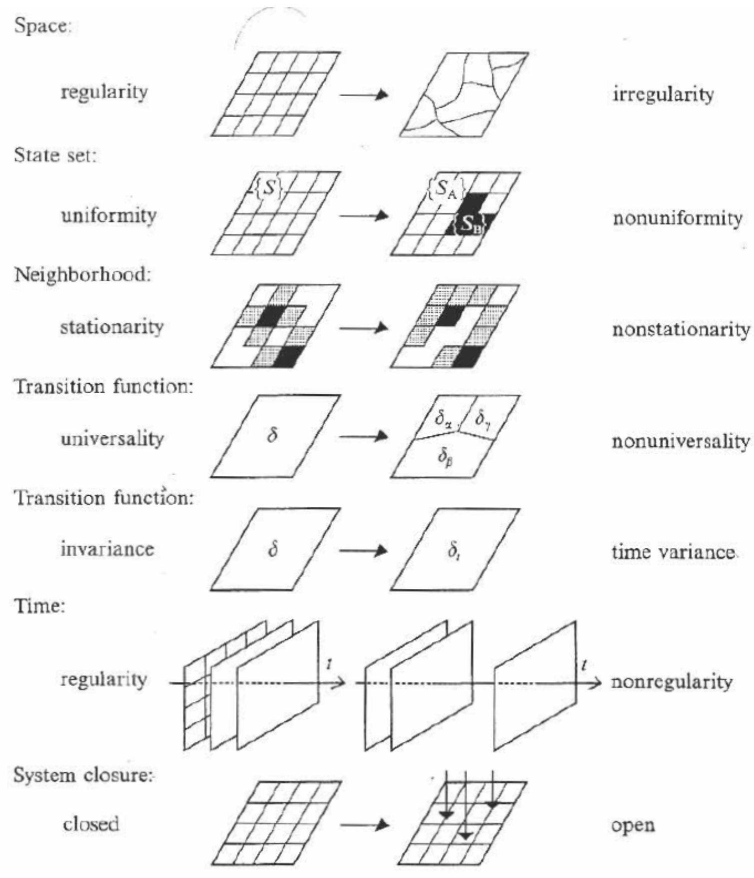


Figure 9 Possible CA relaxations (Couclelis, 1997)

Ménard and Marceau (2005) produced an interesting study on spatial scale sensitiveness in geographic CA. They gathered information about the neighbourhoods used in several CA models from 1993 to 2003. They defined CA spatial scale as a set of three components: spatial extent, cell size and neighbourhood configuration. They focused their study on the analysis of different cell sizes and neighbourhood configurations for a simple two state problem. Small variations in cell size can produce significant variations in results when a given scale threshold is exceeded. The authors state that CA models are not sensitive to variations of neighbourhood configuration. Spatial scale sensitivity affects CA models were the cell actually represents a portion of the geographic space. Ménard and Marceau (2005) also state that although scale issues are supposed to be considered as a design choice that will not drastically vary during the simulation process, modellers must pay more attention to the subject because some of its components are effectively dependent on scale choice.

Transition rules also underwent major developments from their initial classic formulation. The main component of CA is the transition rules (Torrens, 2000). The first and more significant evolution is related with the incorporation of stochastic perturbations in the transition rules. Deterministic behaviours are not suited for the simulation of complex urban phenomena because they can be considered a result of complex interactions between agents. For this reason, many CA models have incorporated stochastic perturbations in their transition rules (White and Engelen, 1993; Clarke *et al.*, 1997; Barredo *et al.*, 2003). Transition rules can also be considered non-static, varying through time. Their universality, an assumption of the classic CA formulation can also be relaxed, as different areas and land uses present different change behaviours. There may be also a distinction regarding the way transition rules operate over time. They can be applied sequentially, updating cell states one after another or in parallel, updating all cells at the same time (Benenson and Torrens, 2004). Von Neumann's self-reproducing CA presented an asynchronous behaviour with the cells being updated considering the previous updates generated by the automata. Conway's and Wolfram's approaches were based on synchronous behaviours, with the entire cell set being updated at the same time (Benenson and Torrens, 2004). There are two main processes of asynchronous updating. The cell set can be updated at a moment in time after a predetermined order according to some cell characteristic. Another method is based on a probability of change in a given moment that is a function of the cell's waiting time (Benenson and Torrens, 2004).

Finally, formal CA can be considered as a black box that processes input data towards an output without any interference from the outside world, that is, it can be considered as a closed system. Urban systems are too far from being close; in fact, an urban system is one of the most opened systems. Therefore, the assumption that an urban system modelled by CA can experience exogenous interference – say stakeholders interference or political decisions – enhances its ability to simulate urban complexity and, consequently, its representativeness. Many models deal with this relaxation by introducing stochastic perturbations that only occur when certain exogenously defined thresholds are overcome (Clarke *et al.*, 1997).

Relaxations are needed to improve CA's ability to simulate complex urban phenomena. The use of real, disaggregate data made possible the relaxation of any or all the assumptions of classical CA, allowing the models to better fit into the complexity of cities (Couclelis, 1997). However, excessive relaxation of traditional CA assumptions may increase exponentially the difficulties to understand the outcomes of a model, in a comeback to widely criticised large-scale simulations (Couclelis, 1997).

2.3.4 Applications of CA in urban change problems

The use of CA in geography and in urban studies was introduced in Section 2.3.2. A group of three groundbreaking studies were presented in order to illustrate the capabilities CA-based models have to simulate complex urban phenomena. The evolution from the classic formulation of CA to modern urban CA-based models was also presented in Section 2.3.3, where the main relaxations and evolutions were discussed. In this section a group of representative CA-based models developed during the last decade or currently under development will be presented. However, the list of CA-based models is by far larger than the one presented in this section.

One of the most widespread models is CA-based SLEUTH developed by Keith Clarke to model and predict urban growth (Clarke *et al.*, 1997; Candau, 2000; Clarke, 2002; Silva and Clarke, 2002). This model is aimed to create a high resolution simulation tool for modelling urban growth (Benenson and Torrens, 2004). Its name is an anagram for Slope, Land use, Exclusion, Urban, Transportation and Hill Shade.

SLEUTH has two main modules (Clarke, 2002): first an urban growth model; second, an embedded land use model that uses information from urban growth model.

The urban growth model requires five GIS-base inputs that are used in image format: urbanisation, land use, transportation, areas excluded from urbanisation, slopes and hill shading for visualisation. Urban extents are required for four different time periods for calibration purposes. Urbanisation results from an urban seed file and at least two road maps that interact with a slope layer to allow the generation of new urban centres.

It considers a regular grid space, a neighbourhood of eight cells and only two cell states: urban and non-urban. It operates five sequential rules of transition: (1) Diffusion, (2) Breed, (3) Spread, (4) Slope resistance, and (5) Road gravity. The growth rate is a sum of five different factors. The diffusion factor determines the overall dispersion of the distribution of single grid cells and of the movement of new settlements outward through the road system. There is a breed factor which is a coefficient which determines how likely a newly generated settlement is to begin its own growth cycle. The spread factor is a coefficient that controls how much outward “organic growth” expansion takes place within the system. Slope resistance is a factor that influences the likelihood of settlement existence on steeper slopes, and Road gravity is a factor of attraction of new settlements onto the existing road system if new areas fall within a given distance of a road (Clarke, 2002). These coefficients generate four different types of subsequent urbanisation (Benenson and Torrens, 2004). First, spontaneous growth: any non-urban cell can be urbanised according to a probability inversely proportional to cell slope. Second, generation of new diffusion centres: each spontaneously urbanized cell can become a new spreading centre if it has a given number of neighbouring urban cells and reaches a probabilistic threshold defined as a model parameter. Third, diffusion at the edges of urbanised areas: there is a fixed probability (another parameter of the model) that allows an edge cell to become urbanised given a certain number of urban neighbouring urban cells. Fourth, road-influenced diffusion: a new spreading centre is chosen given its distance to the road network. It can be dislocated along that road in a randomly selected direction for a given distance (another parameter of the model) for a new location where it can “paste” development. Two randomly chosen neighbouring cells would then change state.

An important improvement of this model is the consideration of self-modification rules aimed to modify the model’s behaviour over time in order to allow it to integrate historical conditions by calibration. Intensive growth periods or periods of little or no growth are identified and the model can stochastically overcome these behaviours.

After the first phase of urban growth modelling, SLEUTH will assign land uses to the new urban areas obtained from the first phase of the model. This assignment is made through an embedded model named Deltatron. The main assumption is to consider each cell as a Deltatron, an urban entity that is associated to only one cell state (or land use) (Benenson and Torrens, 2004). These entities evolve during simulation in a way similar to the one described for urban growth, through four stages.

Firstly, cells are selected at random as candidate locations for land use change on the basis of how much urban growth has taken place. Each newly urbanized cell is assumed to induce a potential change in land use and, as a result, determines whether the selected cell will keep the same or change to another land use; this is made through the consideration of a probability that depends on historical change and cell slopes. If a transition occurs and the cell is not a Deltatron already, a new Deltatron is created. Cluster dynamics are defined as an aggregation process of these new Deltatrons and the associated land use transition. The newly transitioned cell acts now as the land use aggregation centre. This process behaves closely to "organic" growth described above. Age is also considered to characterize Deltatrons: as time goes by, Deltatrons get older and are eliminated when their age reaches a given threshold.

SLEUTH has the credit of introducing the concept of evolving transition rules with the aim of improving the model's ability to retrieve past behaviours. It is also a straightforward model that can be easily calibrated to different urban and regional areas with a small set of calibration parameters. There are already an important number of applications not only in the United States but also in Europe, Africa and South America.

Another innovative approach was proposed by Li *et al.* (Li and Yeh, 2001) They formulated a CA-based model that uses artificial neural networks (ANN) to simulate urban change. It has been used in the Pearl River Delta, China. The set of parameters is obtained automatically by the ANN method, which is made more robust because it uses a back-propagation training procedure.

The authors consider that the traditional CA formulation has serious problems in obtaining consistent parameters values. The method is believed to be suited for dealing with complex interactions between dependent variables without user involvement, a major issue in urban modelling. The model is capable of analysing relevant data, and to eliminate noise and other redundant data. Another important development of this model is related to transition rules: they are obtained from the ANN training process and not from user definition. The model only needs to be fed with training data. It is also capable of dealing with uncertainty by simulating alternative developments through network training, integrating potential exogenous interventions in the change process whose deduction is not possible from historical data. However there still are some shortcomings that need to be overcome. Neural networks proved well for training samples but gave poor results for unseen data sets. Basic model structure is depicted in Figure 10.

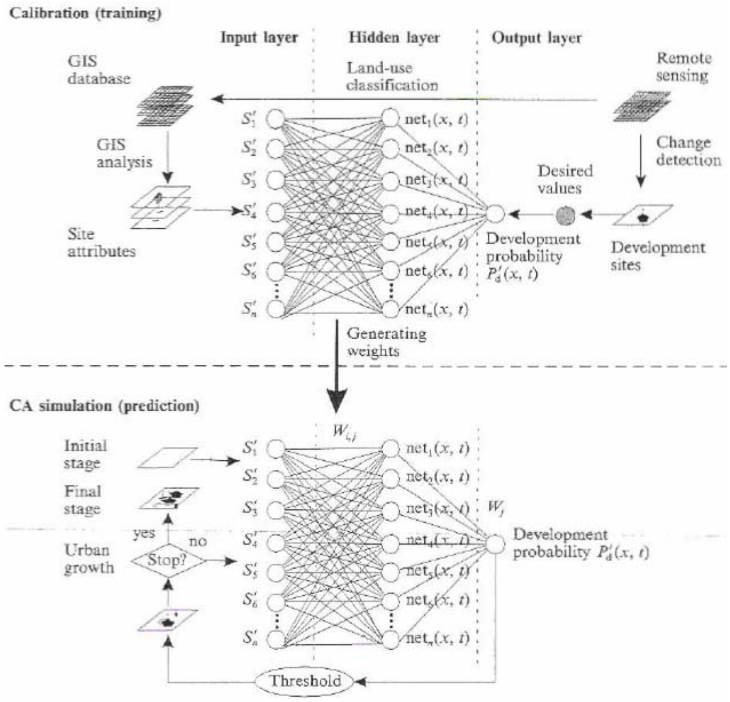


Figure 10 CA model based on artificial neural networks by Li *et al.* (Li and Yeh, 2001)

David O'Sullivan (2001b) proposed a CA-based model that integrates graph theory to enable research into relationships between spatial structure (represented with graphs) and urban dynamics (simulated by CA), as depicted in Figure 11. The model is based on a partition that need not be space-filling and overlapping spatial elements might be used. In order to study urban phenomena, the entities forming the basis of urban morphology (buildings, blocks, streets, census tracts or administrative/planning zones) are a natural set of spatial elements to use as the graph vertices. Edges in the graph represent some sort of relationship between vertices, so that any relationships relevant to the model being developed might be used. The neighbourhood of each vertex consists of a set of adjacent vertices. Cell states may be defined in a suitable way, considering the land use classes as large as necessary to correctly simulate reality.

The graph/CA model was applied to study gentrification in small local neighbourhoods in central London. Data availability and quality is considered to be a problem for using this approach to micro-scale modelling. However, it is the dynamic behaviour of complex models that constitutes a major shortcoming to the application of micro-scale modelling to gentrification problems.

Following the studies of White and Engelen (1993), Barredo *et al.* (2003) developed a constrained CA model. They used the concept of transition potential introduced in the former to build a model based on a regular cell grid with a radial neighbourhood of 172 cells. They also considered a neighbourhood effect materialised through a user-defined weighting factor matrix, establishing quantitative values for attraction and repulsion.

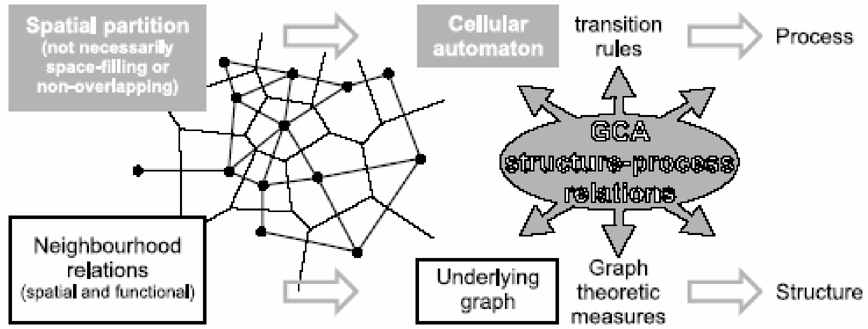


Figure 11 Graph-CA model concept (O'Sullivan, 2001b)

The model was applied to Dublin considering a thirty-year long historic period (1968 to 1998). Using fractal measures and contingency matrices (discussed in the next section) as goodness of fit functions, the model was able to achieve interesting results. Simulated fractal measures for each land use were quite similar to real values for Dublin and the *kValue* can be considered satisfactory. However, although the simulation was able to reproduce urban patterns at a macroscopic level, a quick analysis of the graphic output reveals that the model was unable to match land uses for a large number of areas.

2.3.5 Measurement of CA models performance

Any modelling activity implies the consideration of a series of measuring procedures to evaluate model performance and to establish the boundaries to its applicability in real world situations.

CA have, as a result of their spatial characteristics, a set of well designed measures that are often applied as performance measures and model fitness functions.

The first and most simple method to evaluate the performance of a CA model is based on the visual comparison between maps from the reference situation at the initial, intermediate and final moments considered on the one side, and maps resulting from the application of the model on the other (Clarke *et al.*, 1997; Barredo *et al.*, 2003; Ward *et al.*, 2003; Herold, Couclelis and Clarke, 2005). Although this method is used as a qualitative procedure to calibrate a CA model to certain conditions as it allows to quickly assess its ability to replicate growth conditions (Clarke *et al.*, 1997), it can not assure a reliable quantitative measure for comparison. Therefore, it can only be used in early phases of the calibration process, with an exploratory purpose. Consequent quantitative measures are needed to correctly evaluate CA-based model performance. Statistical relationships must be derived from a set of model data in conjugation with reference data.

A well known measurement procedure is based on the use of contingency matrices for mapping comparisons (Barredo *et al.*, 2003; Couto, 2003; Benenson and Torrens, 2004). This technique has its origin in image processing science, and is applied in a wide group of different areas of knowledge, from geography to medical imaging.

The basic procedure of this technique is to establish a set of comparison indicators between two maps, one resulting from a modelling process and another serving as a reference for the modelled reality. These two maps can be raster images with a pixel structure or any pair of maps that share the same spatial partition, based on regular or irregular units (or cells). The comparison is made through a measure of agreement between the pair of maps. Both maps must be classified by the same finite set of classes, each class associated to a unique land use. The number of cells in each class is then accounted for both maps, and the comparison matrix is assembled. A series of measures is then calculated in order to estimate an indicator of agreement named *kappa* value.

The contingency matrix is computed confronting the model outcome with the corresponding reference layout, accounting the total number of cells n_{ij} that were modelled in a particular state i when they are referenced as being in state j. These two classifications, modelled situation and reference one are independent one from another. The matrix is usually assembled considering the modelled map represented in rows and the reference map represented in columns, as showed in Table 6.

Table 6 Contingency matrix and its notation

Modeled map row i	Reference map column j				Total row i
	1	2	…	k	
1	n_{11}	n_{12}	…	n_{1k}	n_{1+}
2	n_{21}	n_{22}	…	n_{2k}	n_{2+}
…	…	…	…	…	…
k	n_{k1}	n_{k2}	…	n_{kk}	n_{k+}
Total column j	n_{+1}	n_{+2}	…	n_{+k}	

n – total number of cells
k – total number of cell states
OA – Overall accuracy
PA – Producer's accuracy
UA – User's accuracy

Where:

$$n_{i+} = \sum_{i}^{k} n_{ij}, j = 1,...,k \qquad (7)$$

$$n_{+j} = \sum_{j}^{k} n_{ij}, i = 1,...,k \qquad (8)$$

$$OA = \frac{\sum_{i}^{k} n_{ii}}{n}, i = 1,...,k \qquad (9)$$

$$PA = \frac{n_{jj}}{n_{+j}}, i = 1,...,k \qquad (10)$$

$$UA = \frac{n_{ii}}{n_{i+}}, i = 1,...,k \qquad (11)$$

The contingency matrix is an effective way to assess map accuracy, not only globally but also considering each cell state (Couto, 2003). The evaluation of the overall accuracy represented as the sum of the main diagonal of the matrix (Equation 9) is complemented with the evaluation of the producer's accuracy (Equation 10) and the user's accuracy (Equation 11). However, these measures must always be presented along with the contingency matrix, so their values can be compared with the distribution of cells by each pair of modelled/reference states.

Despite the group of measures presented, the main comparison test that can be produced from the data displayed in the contingency matrix is the *kappa* value, from now on referred to as *kValue*. This measure indicates the degree of agreement between the modelled map and the reference map. The *kValue* is based on the difference between the actual agreement in the contingency matrix, i.e. the agreement between modelled and reference situations indicated by the main diagonal, and a chance agreement, indicated by the total of rows and columns.

The closest the *kValue* of a given contingency matrix is to one the highest the similarity between the modelled and the reference maps is. A *kValue* of one means an absolute agreement between the two maps.

The procedure consists in distributing n cells through a $k \times k$ matrix where k is the total number of cell states. Each cell is associated to one of the k cell states within the simulation (usually placed in rows), and to one of the k cell states within the reference data (usually placed in columns). n_{ij} is the number of cells in cell state i ($i = 1, 2, \ldots, k$) in simulation and cell state j ($j = 1, 2, \ldots, k$) within the set of reference data.

The proportion of cell i,j from the total number of cells $p_{ij} = n_{ij} / n$ is the actual agreement p_0 of the matrix, is given by

$$p_o = \sum_{i=1}^{k} p_{ii} \tag{12}$$

and the chance agreement is given by

$$p_c = \sum_{i}^{k} p_{i+} p_{+j} \tag{13}$$

where $p_{i+} = \sum_{j=1}^{k} p_{ij}$ and $p_{+j} = \sum_{i=1}^{k} p_{ij}$. Therefore, the *kValue* is given by

$$\hat{K} = \frac{p_o - p_c}{1 - p_c}. \tag{14}$$

or

$$\hat{K} = \frac{n \sum_{i}^{k} n_{ii} - \sum_{i}^{k} n_{i+} n_{+i}}{n^2 - \sum_{i}^{k} n_{i+} n_{+i}}. \tag{15}$$

Another important measure that may also be computed from the contingency matrix is the conditional *kappa* value. Conditional *kValue* represents the agreement of each cell state inside the matrix. Its value is calculated by the expression

$$\hat{K}_i = \frac{nn_{ii} - n_{i+}n_{+i}}{nn_{i+} - n_{i+}n_{+i}}, i = 1,2,...,k \tag{16}$$

Because of its strong intuitive character, the use of *kValue* and its associated statistics is widely used as an assessment of accuracy for image similarity and spatial based model performances.

Another way of measuring simulation quality is using form indexes that compare simulated and reference urban forms. An example is the Lee Sallee form index used for example by Clarke *et al.* (1997). It is a simple relationship between the common area of two different forms and the total area occupied by those forms. Being *A* an object with a given form and *B* another object with a form similar to *A*, the Lee Sallee index *LS* of similarity between *A* and *B* is given by the following expression:

$$LS = \frac{A \cap B}{A \cup B} \tag{17}$$

Lee Sallee index is equal to one if both objects are equal and is equal to zero if both objects do not have any area in common. The higher the index value is, the better the simulation is.

Finally, it is important to mention the use of fractal dimensions as goodness-of-fit functions for model performance evaluation, commonly applied as measures for CA modelling (White and Engelen, 1993; Benguigui *et al.*, 2000; Longley and Mesev, 2000; Herold *et al.*, 2005; Ménard and Marceau, 2005). Complexity associated to CA soon provided the grounds for the use of this type of measures. Cities can be understood as complex systems with complex rules of evolution, integrating the influence of several interdependent factors that are very difficult to model and to control. Several studies published in the 1990s established a series of values for fractal dimensions that where aimed to classify city's morphology (Frankhauser, 1991; White and Engelen, 1993; Batty and Longley, 1994; De Keersmaecker *et al.*, 2003).

The use of fractal dimensions as measures of performance for CA models depends on the comparison of fractal measures obtained for simulated and for reference maps. White and Engelen (1993) uses sensitivity analyses for different sets of calibration parameters to obtain different urban patterns. Fractal measures are calculated and compared to known values for previously studied cities. The aim is to evaluate the ability of the CA model to generate probable urban patterns. Barredo *et al.* (2003) uses fractal measures to compare CA-based simulation for Dublin metropolitan area with real values assessed from reference maps of Dublin. Calibration is then achieved when a set of parameters that generates the most similar urban pattern to reality is found.

Benguigui *et al.* (2000) presented a method to determine the fractal dimension of an object, based in the correlation between the side of a square in a squared grid, l, and the total dimension of the grid that covers the entire object, as depicted in Figure 12. The number corresponding to the surface of the square, N, is equal to the relationship between the total area of the grid, L^2, and the area of the square unit, l^2, $N=L^2/l^2$.

If the object turns out to be fractal, the relationship can be described by the expression $N=(L/l)^D$, where D represents the object's fractal dimension. The logarithmic plot of $ln(N)$ versus $ln(l)$ yields a straight line with slope ($-D$) (Figure 13). As an example of the use of this fractal dimension applied to the case of the Tel-A-Viv metropolitan area, Benguigui *et al.* (2000) conclude that because of the existence of different fractal dimensions for different regions of the study area, that is, different areas presenting different slopes for the relationship described above, it is considered non fractal, therefore heterogeneous.

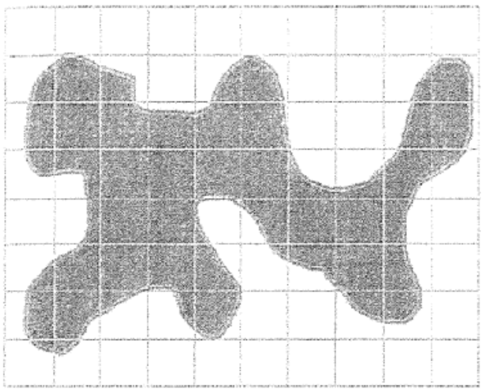


Figure 12 Grid used for a two dimensional object (Benguigui *et al.*, 2000)

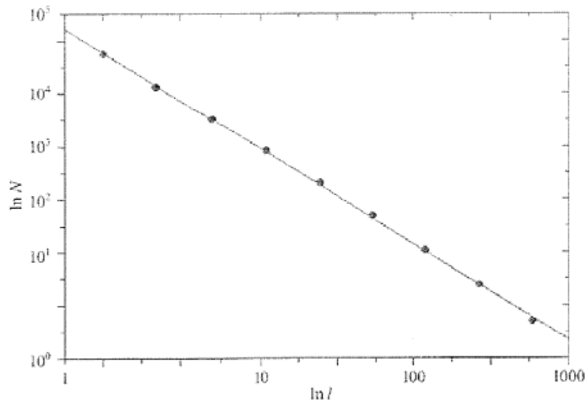


Figure 13 Linear relationship for the fractal dimension of a region (Benguigui *et al.*, 2000)

Another method to assess the fractal dimension of an object is presented by White *et al.* (1993) (inspired by the work of Mandelbrot (1983)). This method is based on the evaluation of an object's dimension (expressed, for example, by the number of cells) versus its radius. Cities consist of scattered distributions of urban activities disposed within the territory they occupy. The use of fractal dimensions for cities can be perceived from the parallel established with the Sierpinski carpet problem[7] depicted in Figure 14. For such objects, it has been shown that

$$B^i = q^{-iD} \tag{18}$$

[7] Cities, of course, present a stochastic behaviour that is totally absent of in the Sierpinski carpet problem.

where B is the number of cells occupied by the original object (it takes the value of 5 for the Sierpinski problem), i is the step number (i=1 in Figure 14a, i=2 in Figure 14b and i=3 in Figure 14c), q is the scale reduction factor (it takes the value of 1/3 in Figure 14) and D is the fractal dimension. Solving this Equation 18 for D it can be obtained

$$D = \frac{\log(B)}{\log\left(\frac{1}{q}\right)}. \tag{19}$$

As it can be seen in Figure 14, as the object expands in cell space, the number of cells composing it grows less rapidly than the number of cells in the square area necessary to contain it, so that the object becomes more sparse (White and Engelen, 1993). In particular, the length L of a side of the figure is

$$L = \left(\frac{1}{q}\right)^{i}, \tag{20}$$

and the total number of cells B_T is given by

$$B_T = B^i . \tag{21}$$

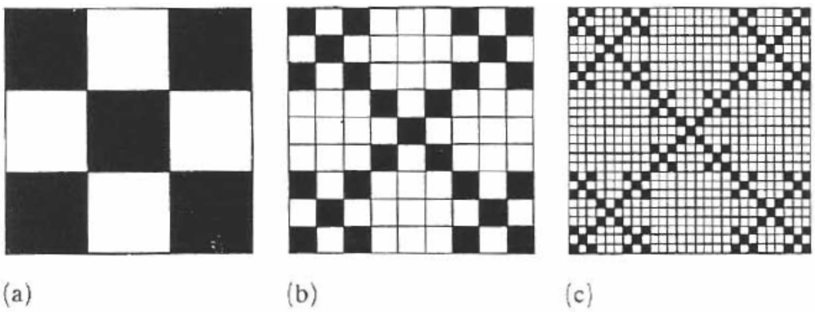


Figure 14 The Sierpinski carpet

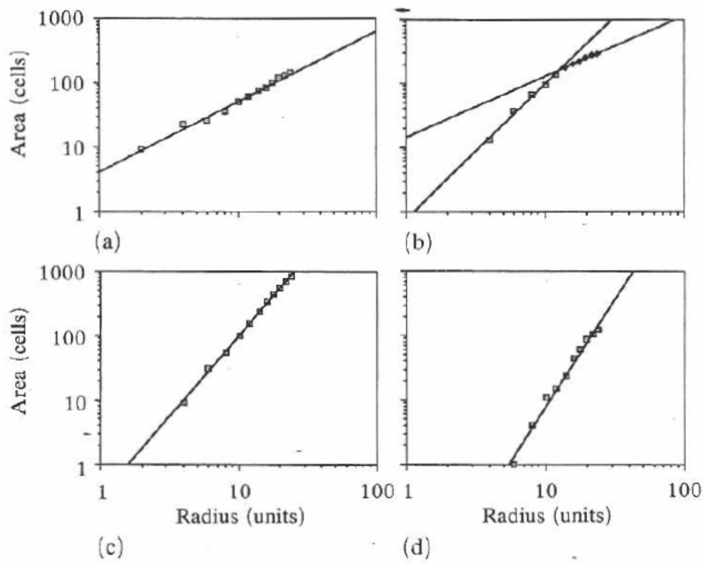
The relationship between the size of an object (measured in the number of cells) and its diameter (fractal dimension) is given by the expression

$$B_T = L^D , \tag{22}$$

which can be transformed in a linear relationship by the application of logarithm

$$\ln B_T = c + D \ln r . \tag{23}$$

where c is a constant and r is the radius of the object. Equation 23 provides a useful way of estimating the fractal dimension D of a city (considering its stochastic spatial distribution) as it can be obtained from a regression analysis between the variables B_T (total number of cells in a given cell state) and r (the radius of its distribution), being its slope the fractal dimension value for that cell state. Note that this measure is location specific, as it depends on the choice of the point of origin for measuring the objects radius. Figure 15 depicts the graphical output of the relationship for the city of Cincinnati for the four land uses considered by White and Engelen (1993).



(a) commerce (b) industry (c) housing (d) other land uses

Figure 15 Area-radius relationships for the city of Cincinnati (White and Engelen, 1993)

The measure of the area-radius relationship yields important information on the urban morphology of a city. The regression equation represented in Figure 15 (b) for industrial land use has a bi-linear behaviour, indicating that the city of Cincinnati has a bi-fractal behaviour for industry, that is, there are two distinct concentric regions of the city with two different values of the fractal dimension. White and Engelen (1993) also present the values for the radial dimensions of a series of conceptual spatial distributions generated by their CA model and for a set of four US cities, shown in Table 7.

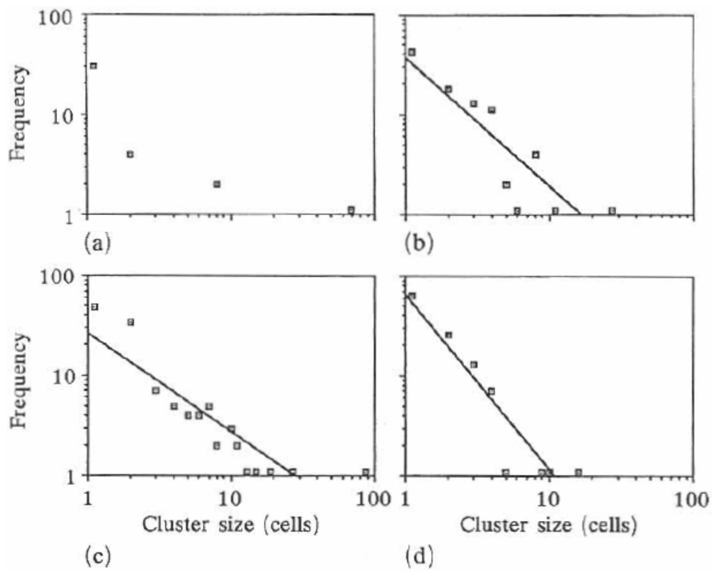
This radial dimension measure, as it was referred to by Frankhauser and Sadler (1991), is based on the consideration of a single centre, ignoring the fact that city growth generates a series of new city centres.

Another fractal measure presented by White and Engelen (1993) is the cluster size frequency spectrum. Considering the set of cells occupied by a particular land use in a city as an object formed by a number of clusters of various sizes, it can be found a fractal behaviour of the object if there is no characteristic cluster size, this is, if every change of scale maintains the frequency ratio of clusters that differ in size by a given factor. This behaviour indicates that the object is self-similar. The log-log plot of frequency of occurrence of each size against cluster size will be linear if the object is fractal. In Figure 16 is depicted this relationship for commerce for a set of four US cities. Note that Atlanta does not present a linear relationship; hence commerce has no fractal behaviour in this case.

Table 7 Radial dimensions for a set of conceptual cities (Cellular group) and US cities (White and Engelen, 1993)

	Cellular group			US city			
Land use	1	2	3	Atlanta	Cincinnati	Houston	Milwalkee
Commerce	1.09	1.17	1.03	1.00	1.10	1.24	1.27
Industry	1.85	2.41	2.72	1.97	2.11	1.51	1.83
Housing	2.29	2.96	2.77	2.12	2.51	2.76	2.38
Other				2.52	3.42	2.77	2.17

Clusters are commonly defined in the literature only by horizontal and vertical adjacencies (White and Engelen, 1993). However, as a city grows, all land uses grow along, making this cluster size frequency measure iteration specific.



(a) Atlanta (b) Cincinnati (c) Houston (d) Milwaukee

Figure 16 Cluster size frequency spectra for commerce in four US cities, in 1960 (White and Engelen, 1993)

These authors also present perimeter-length scaling as a fractal measure used for assessing urban growth. It is based on a property of fractal lines: its measured length increases as the length of the unit with which it is measured decreases. The slope in a log-log relationship between the number of steps required to make one trip around the object's boundary and the length of the step used yields the fractal dimension of the boundary. The close range of values obtained by White and Engelen (1993) for a conceptual city for different moments of the city evolution makes evident that cities are quickly organized into a structure which is maintained as they grow (see Figure 17).

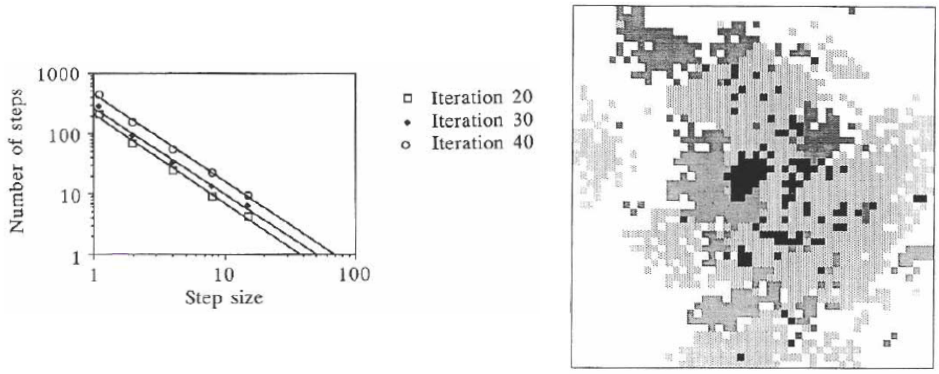


Figure 17 Perimeter-length scaling for a conceptual city (White and Engelen, 1993)

2.3.6 Calibration of CA

As important as the conceptual development of a model, calibration is aimed to ensure the necessary connection between simulation and reality. No matter how capable a model approach is to simulate real world phenomena – and CA are a very physical method of simulation – it will not produce good simulation without proper calibration.

In the past section a series of measures commonly used to assess CA performance were presented. From visual methods to performance measures and fractal analysis, there are several approaches that are usually mixed to develop calibration procedures. In the present section these approaches will be presented for the major CA models that were developed or are currently under development.

There are two main approaches to calibration: one more dependent on user intervention and another based on optimization techniques. The first one uses visual comparison of simulation and reference maps to determine initial parameter values and sensitivity analysis to improve these values. The second one is based on different techniques aimed to identify the optimal set of parameters without user intervention. The main purpose is to allow the model to establish the correct set of parameters that can simulate complex urban interactions.

SLEUTH calibration (Clarke, Hoppen and Gaydos, 1996; Clarke *et al.*, 1997; Candau, 2000; Silva and Clarke, 2002) is based on a two-step procedure: first, a visual calibration oriented for a broad parameter definition and debugging, based on an user-friendly, real time application that allows the user to stop the simulation and to modify a given parameter, analysing model's sensitiveness to each parameter; second, a brute force calibration procedure where multiple runs are produced in order to generate enough model data to statistically compare reference data. Visual calibration is considered useful to establish meaningful ranges of values for parameters. A set of statistical measures are displayed in real time to allow the user to understand how the model reacts to a given parameter change. The second stage is much more complex and is based on a sequential procedure of consecutive runs of the model aiming to obtain a good set of parameters at three levels of calibration (coarse, fine and final). Considering the results obtained from a set of thirteen metrics, each phase is fed with the best scoring results for the set of control parameters, weighted with a set of user-defined control parameters (for a detailed descriptions see (Clarke *et al.*, 1996)).

The Li and Yeh (Li and Yeh, 2001) CA-based model used artificial neural networks (ANN) not only to simulate urban change but also for calibration purposes. The use of ANN for model calibration is presented as a solid alternative to other methods that are user-dependent and computationally expensive. The model uses the concept of transition probability that is multiplied by a stochastic factor aimed to deal with uncertainty. As it can be seen in Figure 10, the ANN model uses empirical data obtained from GIS and remote sensing to train the network in order to establish values for development probabilities for each land use. It relates multiple cell attributes (obtained from GIS) and growth history (obtained from remote sensing) to calculate a set of calibration parameters that are subsequently used by CA-based model to simulate urban change. Calibration is produced with selected sample data sets, using a stratified sampling method to assign a certain number of observations for each category to be evaluated. Results showed that the method succeeded in achieving logical results for site attributes. However, training process must be controlled in its duration in order to stop the process when the network begins to over-train.

The Barredo *et al.* (2003) model was calibrated through a simple process of sensitivity analysis of the calibration factors considered. After a visual analysis of simulation results, calibration factors were carefully altered and new simulations were produced and their results assessed. Fractal measures were calculated and compared with observed values for the case study. This process can be considered excessively simple because it does not consider the combined effects of the entire set of parameters. Urban phenomena are a result of several interactive behaviours and the analysis of each of these behaviours by itself considering the others as fixed may produce significant distortions and reduce the representativeness of the model.

3 Methodological Approach

In the previous chapter an attempt was made to present structured information on urban modelling and on the use of microsimulation approaches to deal with a wide set of urban phenomena. The literature survey was mainly focused on the use of cellular automata (CA) for modelling urban change problems, particularly those related with the growth of built areas. This chapter is dedicated to the presentation of the methodology adopted in the present study.

The reasons for the use of CA will be discussed in Section 3.1, based on the theoretical foundations for the use of this modelling technique in urban science presented in the previous chapter.

The discussion over the classical formulation of CA and its relaxations takes place in Section 3.2. Issues regarding all the five main aspects of CA –cell space, neighbourhood, cell states, transition rules, and time – are discussed, presenting the theoretical grounds for the CA model designed specifically for this study.

Section 3.3 introduces the theme of the model calibration and validation. An optimization approach based on Particle Swarm algorithm (PS) is used and it is introduced in this section.

Finally, it will be presented in Section 3.4 the CA model developed.

3.1 The use of CA

One of the most important design choices to be made when modelling urban phenomena is related with the choice of suitable modelling approach. It is important to consider several aspects of the problem at hand: which type of problem is set to be studied; which characteristics of the problem are selected for modelling; what is the appropriate scale of analysis; what type of measures would be selected to correctly evaluate the model performance; how can the model be validated and calibrated considering the available data; and finally, in a theoretical perspective, if a new modelling approach that have not yet been used prove to be effective.

CA are a very intuitive and straightforward technique and it is easily applied to spatial problems of any kind. It is based on a set of rules that intentionally operate over a spatial (cellular) structure thus reproducing spatial phenomena that evolve over space and over time quite well.

But why should one choose CA instead of other micro-simulation techniques that are also capable of reproducing intrinsically spatial behaviours in different types of problems, such as rule-based modelling or agent-base simulation? What characteristics do CA have that make this approach more suitable than other approaches mentioned?

This study was set to study urban growth phenomena considering two main aspects that were not well studied by the time it was produced: how could the growth of small urban areas be studied following a scientific approach based on urban modelling; and how could the model be tailored to address a reality that is intrinsically simple, no matter what modelling approach would be chosen.

Small urban areas tend to constitute intrinsically simple problems when comparing with larger ones. This results mainly from the fact that these small urban areas usually comprise a smaller set of urban land uses and functions than the large ones, which reduces substantially the number and the scope of the interactions and relationships that concur to their evolution.

For this reason, the modelling technique chosen should be able to replicate this simple set of land use interactions in an intuitive manner, reproducing those interactions from a simple set of change rules. Both CA and agent-based simulation are able to do the job, but then another important factor must be taken into account: data availability.

The question of data availability is crucial in the Portuguese planning context (or any other planning context where information is poor). The implementation of an agent-based simulation model requires more data, as the consideration of individual agents such as households, land parcels or even individuals increases the resolution of the behaviours modelled. Agent-based simulation is oriented for simulations where individual behaviour (understand as individual decision-making) is the driven force of the system interactions.

In the present study, the relationships that are modelled have, at least implicitly, its foundations on the individual behaviour of a small set of agents, say land owners/developers, households and administration. The relationships that result from these agents are strongly oriented for land use rather than the individual behaviour of each one of them.

But, considering that all the available information is disaggregated, in the Portuguese case, to the census tracts, and that it is very difficult to obtain reliable information about individual agents, particularly for these small urban areas outside the main urban centres[8] which are the subject of the present study, the choice of an agent-based approach appears to be less indicated than the choice of a CA-based approach.

Another important aspect that led to the choice of a CA approach is related with the context in which the Portuguese urban areas evolved for the last two decades. As it was mentioned before, planning in Portugal evolved from an under regulated situation in the early 1980s to a sophisticated legal framework aimed at updating planning prac-

[8] Even for large urban areas information disaggregated to the individual is commonly unavailable; essentially, there is only information on mobility patterns for the metropolitan areas of Lisbon and Porto.

tice in Portugal. At the same time, benefiting from large sums of structural funds provided by the European Commission along with the explosive growth of the housing construction market, there was a strong growth of the urbanized areas.

In this context of rapid change, CA models appears as a good approach for modelling urban growth, benefiting from the fact that CA are oriented to the occupation of land rather than to modelling individual behaviour of agents. In fact, with no historical data on market behaviour, individual housing choice or mobility patterns, it could become very difficult to validate and calibrate an agent-based model.

Finally, there was a great interest in applying CA techniques to problems that meet a certain group of conditions.

The rapid growth of urban areas under changing legal conditions can benefit with the support of decision models that are able to simulate urban phenomena based on intrinsic land characteristics and neighbouring relationships. This has been an important area of CA research, with several studies on its application to underdeveloped areas and rapid growth contexts (Xie, 1996; Li and Yeh, 2001; Wu, 2002; Fragkias and Seto, 2005).

The application of CA models to problems with smaller scale is another important issue, as CA are usually applied to regional/metropolitan areas with a wider set of urban land uses and functions.

The use of an irregular cell fabric based on spatial information units is another challenge, because it has not been well studied yet. Several issues are raised by the consideration of irregular cells, such as neighbourhood form and land demand assessment.

3.2 The formulation of the CA model

Geographic applications of CA are adaptations of the classical formulation of CA presented in Section 2.3.1, with a set of frequent relaxations that were discussed in theoretical terms in Section 2.3.3.

The present application, used to study small urban areas, also required a series of relaxations to the classical of CA. However, it was set as a main goal for the formulation of this model the respect for the backbone of the classical formulation, mainly characterized by its simplicity. Because of this, the model can be classified as a classic constrained CA model.

The CA model developed in this study has a simple structure that derives from the classical formulation of CA with the consideration of constrained land use demand (after the pioneer works of Tobler (1979) and White and Engelen (1993)). Based on a cell defined by the combination of census tracts and urban areas, the model has a set of six major aggregate cell states that are intended to represent the whole system of urban land uses and functions. Besides cell state, a set of other cell parameters are defined, such as cell suitability for each land use and cell accessibility (measured as a weighted value of the distance to major urban centres). A neighbourhood dimension is defined as a model parameter to establish the boundaries of urban interactions. Land use dynamics will be a result of the application of a set of transition rules based on the assessment of a transition potential. The simulation will evolve over ten years time steps, based on information from two consecutive demographic census.

The model comprises a series of calibration parameters. These parameters are used to calibrate the model to specific conditions. Simulation results depend on the concurrent effect of their application.

The values for the calibration parameters will be established through an optimization process based on the Particle Swarm algorithm considering as a fitness measure the *kappa* value associated to a contingency matrix.

3.2.1 Cell space

The first component of CA to be discussed is cell space. The structural base of CA is the cell unit and the spatial structure created by the entire set of cells.

The classical formulation of the model is based on an orthogonal system of quadrangular cells, each one assuming a single state within a given neighbourhood in each time period.

In geography applications, the cell lattice used is usually obtained from remote sense maps, which are raster images with a certain level of resolution, depending on the satellite source chosen. The cell is set to be the pixel of that raster image, with its dimension indexed to pixel resolution. The usual range of resolutions used on CA models varies between 100 and 25 meters.

One of the main goals of this study is to assess the use of irregular shaped cells instead of those traditional square ones. Therefore, it is important to clarify the reason for making this choice and to present the main characteristics of the spatial unit chosen.

The first reason that led to the choice of irregular cells is related with the traditional structure of the Portuguese urban centres. The evolution of urban areas was always strongly related with the adaptation of urban form to the natural conditions of the occupied land. This continuous process has its origins back to when the Romans founded the first cities, occupying the previous settlements usually located in suitable lands with high agricultural potential and strategically located close to road infrastructures. The Arab occupation also left an important mark in the traditional urban structure: the complex system of small streets that is a pattern for all the main urban centres in Portugal. The Romans set the location of the majority of the Portuguese cities and the Arabs gave them their form.

The medieval city was based on the Arab urban structure. The construction of larger churches and monasteries was the first important moment of urban expansion. The old city inside walls had no available land left for such large constructions. These monasteries occupied large estates outside the walls, and lower social strata started to occupy contiguous areas founding new neighbourhoods that last until today.

Only in the transition from the 19th to the 20th century new urban design trends practiced in central Europe started to influence the Portuguese urban planners, with the expansion of the major cities (particularly Lisbon and Porto) to new areas, drawn to accommodate new urban functions within a more monumental setting. New avenues were drawn, according to an orthogonal pattern, thus changing the traditional urban form. This process evolved to the mid 20th century with the rapid growth of coastal urban areas.

The evolution of the typical Portuguese urban form, briefly summarised in the past paragraphs, generated a highly heterogeneous urban pattern[9], often centralized by an old historical centre, a heritage from the medieval city, surrounded by a series of expansions that took place in the past 150 years, accompanying the growth of population.

Considering the scale of analysis proposed for this study, centred in small urban areas, the consideration of regular square cells could became a problem, not only because of the resolution of the cell, but also because of the data availability for this spatial partition.

If a high resolution was considered, say a 25×25 square meters cell, the model would became closer to agent-based simulation than to CA with cell size becoming similar to land parcel size. The problem of gathering available data could become critical, along with the need for changing the approach.

If the choice was for a low resolution for the cell, around 100×100 square meters or higher, the cell would become less representative of the urban form, particularly in the definition of the urban areas.

For these two reasons, cell representativeness and data availability, irregular cells were chosen. But what kind of cells? And with what shape?

The Portuguese statistical information system is based on census tracts designed with the primary purpose of distributing the effort of data collection during the major demographic census. The process that leads to the final configuration of the census tracts is conducted by INE, the national bureau of statistics, and includes the participation of municipalities within the administrative limits of their own territory. As a result of this cooperation, the census tracts came out to be polygons whose shape is strongly related with the spatial distribution of the built areas, as it can be seen in Figure 18. For this reason, census tracts are quite homogeneous with regard to indicators such as population, population density, number of households, number of homes, etc..

These census tracts represent quite well the urban structures of cities, because of the way they are designed. At the same time, they hold reliable information about demographics, households, constructions, and employment. This combination of facts qualifies these spatial units as good cell unit used by this application of CA.

[9] There are only two cases in Portugal of cities drawn from the beginning with an orthogonal structure: the reconstruction of Lisbon's downtown after the major earthquake of 1755, the first Portuguese planning operation, and the city of Espinho, a 19th century city that was built based on an orthogonal style strongly influenced by important contemporary urban expansions like Barcelona's *eixample* and down town Manhattan.

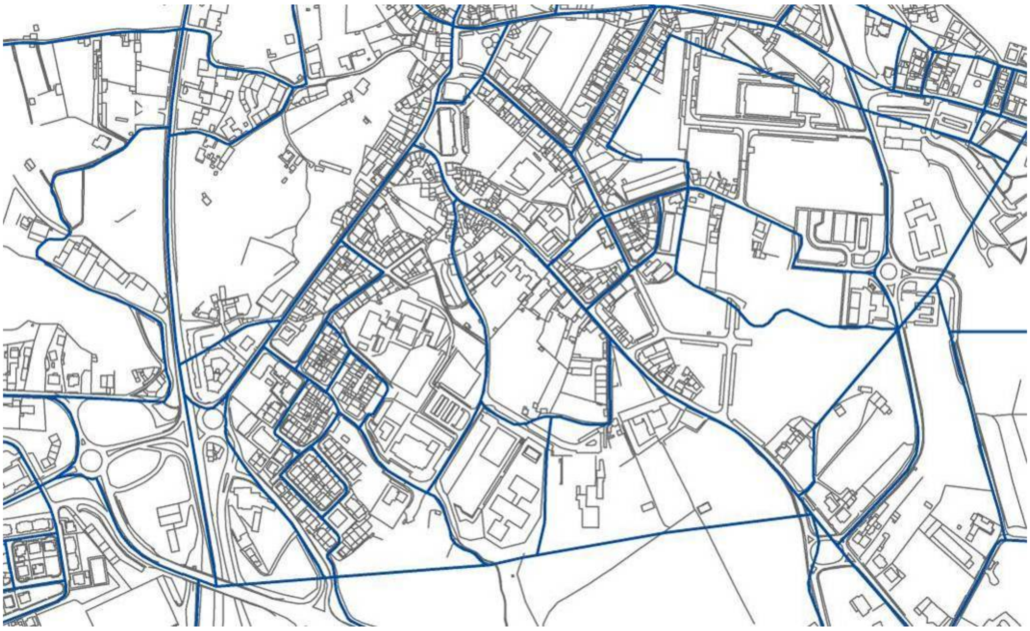


Figure 18 An example of spatial distribution of census tracts

But the census tracts do not represent completely the urban structure of the territories that are under discussion in this study. In the present case of small urban areas, considering the Portuguese context, urban structure is technically defined by the urban limits of the localities.

Therefore, to define the cell for the application of CA, it is necessary to intersect information between urban areas, which contributes with the definition of the urban structure, and census tracts, which yield reliable statistical information. The cell will be a portion of the structured territory characterized by good statistical parameters that will allow the automaton to grant a good degree of representativeness.

3.2.2 Cell states

After discussing the pertinence of using an irregular cell space based simultaneously on spatial information units and urban areas it is important to discuss how the information will be structured and used by the automaton in the model.

The cell in the CA model is classified in a particular state that is aimed to represent, in geographic applications, the type of land use occupation. In a simple formulation, commonly used in the early models based on CA, only two states were assumed to be possible: occupied and unoccupied. This assumption is very close to the strict formulation of CA.

More sophisticated models expanded the notion of occupation, in order to bring the models closer to the real behaviour of land use occupation. A series of land use states were commonly referenced in several studies, ranging from different types of agriculture to a set of density-based residential states, industry and commerce, among others (Clarke *et al.*, 1997; Semboloni, 1997; Barredo *et al.*, 2003; Ward *et al.*, 2003)

There is a common distinction to be made between these types of occupation: there are states called active which influence and are influenced by other states; and states called passive which influence other states but are not influenced by them. The former include all the urban land uses that are the subject of demand, that is urban and industrial land uses; the latter include all the restrictive land uses such as agriculture, ecological reserve and forestry.

In the present case, the establishment of a set of land use states followed the main goal of keeping the model simple, as close to classic CA formulation as possible. The problems under consideration involve small urban areas, which are characterized by simple urban structures and small diversity of land uses and functions.

For this reason only six aggregate cell land uses were considered, each one with the corresponding cell state. These cell states are: (1) urban low density (ULD), (2) urban high density (UHD), (3) industry (IND), (4) non-urbanized urban areas (N-UUrb), (5) non-urbanized industrial areas (N-UInd) and (6) areas where construction is highly restricted (Rest).

In addition to these six cell states, a transitory condition was imposed for each cell chosen for a transition of state (this process is described later in Section 3.2.4). For one time step immediately after the state transition the cell is considered as undergoing an occupation phase and is put aside the urban change process.

All the urban land uses and functions were considered in an aggregate form under the designation of urban land use, varying only the construction density (states ULD and UHD). This option is founded on the fact that commerce and small industries are, for the problems under consideration, typically located within the urban area, mainly in the ground-floor of residential buildings, independently of their dimensions (single family houses or apartment buildings). The cell unit described in the past section is independent from the road infrastructure, and public facilities are, at this scale, also included in the urban states.

Urban areas were classified according to construction density. Unlike some studies referred to before, in which a wide range of urban construction densities was considered only for residential land use, in this study the choice was for a classification that could represent the traditional structure of Portuguese small towns. This classification is also close to the planning rules commonly applied by municipal master plans, which are based on construction coefficients that are site dependent.

The consideration of all urban functions aggregated under a general urban land use is thought to incorporate those areas occupied by infrastructures and public facilities, as their demand is generated by the increase on population. In the simulation, the variation of urban construction density reflects the increasing public demand for public areas, facilities and infrastructures over the years, as it represents an increase in population and in the levels of quality of life. As a consequence, the construction densities decreased over the simulation, as explained later in Section 3.2.4.

The consideration of both census tracts and urban areas for the establishment of the CA is of great importance to increase the representativeness of the cell regarding the declared aim of this study of using a CA model based on a "natural" spatial lattice. Urban areas are defined by the municipality's master plan in order to accommodate both the existing urban structure and future expansions, and are revised at each revision of the plan, legally within a ten years period.

Industrial areas are, in these territories, usually located outside urban areas, in specific locations predetermined by the municipality's master plan. The industrial areas are occupied exclusively with industrial activities, and they are located usually near a main road to ensure good levels of accessibility. The importance of the industrial dynamics and its influence over the spatial structure of the territory imposed the consideration of this land use, both its previous existence (state IND) and possible expansion (state N-UInd).

It is also important to mention the set of land use constraints that were considered which influences land use dynamics but are not influenced by it. The Portuguese planning system has a series of land uses that are treated as constraints to the development of urban areas. There are national constraints, such as the National Agricultural Reserve (RAN) and the National Ecological Reserve (REN), and local constraints, such as protection areas surrounding historical or natural landmarks and public facilities or infrastructures. The consideration of a specific state for this group of land uses is extremely important, as it will represent the main constraints to the expansion of urban and industrial areas.

There is a set of other cell parameters that are assessed and are aimed to characterize different aspects of the territory, such as land suitabilities, zoning and accessibility.

Land suitability parameter intends to assess the main geophysical characteristics of a given parcel of land, such as solar exposure and land slope. These characteristics will influence the cell's change potential, playing an important role in the land use dynamics.

Another important parameter is the zoning status of the cell. Zoning is imposed by the administration via the municipal master plan. It is a result of a series of evaluations that take into account a series of parameters that are directly related to land suitability and urban pre-existences.

In fact, zoning parameter is, by definition, a classification for each location for a specific land use, making unfeasible its occupation by any other land use within the duration of the planning regulation. For this fact, and considering once more the aim of keeping the model as close to the classical formulation of CA as possible, zoning and land suitability where considered as one single parameter called suitability, assuming a binary value of 1 if the cell is suitable for a given land use or 0 otherwise.

The last cell characteristic considered in this model is related with the accessibility level of each location. Accessibility is of great importance for the evolution of a territory, as the proximity to road infrastructures generates higher attraction levels both for industrial/commerce activities as for residential location. For this reason accessibility is one of the components, with suitability and the neighbourhood effect (which will be discussed in the next section), that constitute the urban transition potential of a cell.

Small urban areas tend to have less urban activities – public services, schools, commerce – than large urban areas and these activities are usually located on the municipality main town. This suggests the necessity of measuring accessibility not as an aggregate measure that takes into account the whole territory but as a relationship between the distance from the cell and the location of main urban functions.

The relationship used to assess the accessibility measure is expressed in Equation 24,

$$AccMeas_i = \alpha_{acc} \times D_{i,MunicipalMainTown} + \beta_{acc} \times D_{i,CivilParish} + \gamma_{acc} \times D_{i,IndustrialPerimeter} \quad (24)$$

where $AccMeas_i$ is the accessibility measure for cell i, α_{acc}, β_{acc}, and γ_{acc} are calibration parameters, $D_{i,MunicipalMainTown}$ is the distance (in minutes) from cell i to the municipality main town, $D_{i,CivilParish}$ is the distance (in minutes) from cell i to its civil parish and $D_{i,IndustrialArea}$ is the distance (in minutes) from cell i to the major industrial area located within the municipality's territory.

The accessibility measure is then based on the assessment of the level of access that is granted by a location considering both the location of the major urban functions, particularly public services, education, health attendance, security services, commerce and, more generally, employment and the road network that supports these connections.

The road network of the territory was classified in three levels: main roads, secondary roads and local accesses. Every cell is served by the network and has a mobility level associated with the existence of a connection to only one of the hierarchical levels. If the centroid of a cell is served by a main road then its mobility level is classified as 1 (the highest value), if it is served by a secondary road then its mobility level is classified as 2 and if it is served by a local access then its mobility level is classified as 3 (the lowest value).

The final accessibility value must be normalized for comparison reasons, being this procedure performed through the following expression

$$Acc_i = 1 - \frac{AccMeas_i}{\left\| \sum_{i \in C} AccMeas_i \right\|}. \quad (25)$$

Final accessibility Acc_i is then the difference to one of the normalized value for the accessibility measure $AccMeas_i$. This transformation is necessary as this accessibility measure is calculated by the weighted sum of the distance to urban functional centres. Consequently, the higher this sum is, the less accessible the location is. Accessibility levels are considered constant during the entire period modelled.

3.2.3 Cell neighbourhood

One of the most distinctive characteristics of CA is its dependence of the interactions generated between each cell and its neighbours, as these interactions have a great influence in the occupation or vacancy of cells under certain conditions, for example, the proximity of two repulsing states such as residential and industrial.

As it was discussed in Sections 2.3.3 to 2.3.6, there is a group of commonly used neighbourhoods that range from the classical formulation of CA to more relaxed adaptations of the concept.

The classical formulation considers that the neighbour effect is related only to the set of cells that are directly connected to the one in the centre, assuming different configurations as the Moore or the von Neumann neighbourhood represented in Figure 4. These configurations were used specially in early CA models, when the degree of relaxation was yet far from those considered in more recent models, as the complexity of problems started to increase.

The first law of geography states an important principle, particularly in urban sciences (where spatial relationships are a major component of analysis, as it was presented back in Chapter 2): *everything is related to everything else, but near things are more related than distant things* (Tobler, 1970). The concept is embedded in the relationship described by the gravity model, where the interaction between two entities varies inversely to the distance (or travel cost) between them. This relationship is valid for land uses and for urban functions: two locations with similar or different land uses generate attraction or repulsion that varies with the distance between them.

Therefore, the strict concept of CA that imposes the consideration of local neighbourhoods, essentially relating the cell with its direct neighbours, can be too much of a simplification of reality, reducing the model's ability to represent real world phenomena. This is to say that action-at-a-distance must be considered, in an effective way of incorporating human behaviour in the models' dynamic.

Because of the necessity of effectively evaluating the spatial relationships of the land uses that were considered in the model, it was considered that action-at-a-distance relationships should be integrated in the evaluation of the cell urban potential. For that, it was necessary to define neighbourhood dimensions, depending on the model's dynamic.

The model's neighbourhood should meet the natural, therefore real, perception of the neighbourhood. But this concept has two different perspectives. In larger urban areas, neighbourhoods tend to be perceived locally, because functions are more distributed throughout the territory granting higher levels of service[10]. In smaller urban areas urban functions are provided by a smaller set of facilities, commonly located all over the city. For this reason, the concept of neighbourhood tends to include the entire urban area.

The choice was for considering neighbourhoods as a circular area defined by a value for its radius. This choice is also considered in several other models based on CA: instead of using the traditional neighbourhoods (von Neumann's or Moore's), it has been used a circular area within a given radius measured by a constant number of cells (White and Engelen, 2000; Barredo *et al.*, 2003). In these cases, the number of cells was determined considering, in an arbitrary way, a distance commonly perceived as a natural neighbourhood.

In the present case, the neighbourhood distance was considered as a model parameter, that is, its value was determined by the calibration process in order to establish a final value that would influence and be influenced by all other parameters of the model. This is thought to be a more appropriate approach for the problem, as the notion of neighbourhood relates to factors which can not be dissociated from each other. For example, the opening of a new shopping centre in a mid-size city will affect not only the way that people do their shopping; it will also modify travel patterns, business location trends, and housing locations. This will have a significant impact on the notion of neighbourhood, far from been concentrated on the natural local neighbourhood.

3.2.4 Transition rules

Transition rules are, as it was discussed in Chapter 2, the most important aspect of a CA. They are a set of rules that will guide the automaton in the dynamic process of cell state change.

The components of CA described earlier – cell space, cell states and cell neighbourhood – constitute the support structure over which the automaton will operate through the transition rules. These rules define a set of conditions that will produce the change of states for each cell in each time step. Their application takes into account all the characteristics of each cell by the form of a measure of the transition potential, a weighted value of accessibility, suitability and those interactions between neighbours that occur within the predetermined neighbourhood, named neighbourhood effect.

Both accessibility and land suitabilities where assumed as cell (or land) characteristics in Section 3.2.2. The last component of the transition potential is the neighbourhood effect, defined as a measure of the level of interaction between two locations within a certain distance. It is the result from the attraction or the repulsion generated by two cell states (equal or different), considering the distance between them. Neighbour-

[10] A good example is the location of public facilities. In larger urban areas, the number of public facilities of each type (civic centres, libraries, municipal swimming pools, etc.) is considerably higher than it is in smaller ones, creating neighbourhoods that are well served, thus reducing its perceived dimension.

hood effect was considered as an interaction value that decays as the distance between two land use locations (or cells) increases. This value is set to -1 if the interaction between two land uses is characterised by total repulsion, 0 if they do not interact, and 1 if the interaction between two land uses is characterised by total attraction. Repulsion means that two land uses have their transition potential reduced of a given value that depends on their distance. On the contrary, attraction means that two land uses have their transition potentials increased of a given value that depends on their distance.

It is important to notice that all the six states influence the transition potential of each other. On the contrary, only the five active states, all but *Rest* state (areas where construction is highly restricted), are influenced by the other states. This neighbourhood effect is then a relationship that must be determined for each pair of land uses that are considered in the analysis of urban change dynamics.

These relationships were considered linear, as depicted in Figure 19. The fact that these relationships are very difficult to assess, as they depend on several interdependent factors such as land value, housing demand, public facilities location among several others, suggests their consideration as calibrations parameters. Two parameters were considered for each neighbourhood effect relationship: (1) the ordinate at the origin and (2) the abscissa for the ordinate with value 0, the point where the interaction between the two land uses becomes inexistent. The total number of parameters equals 60, 5 active states multiplied by 6 states (the entire set of cell states) multiplied by 2 parameters per relationship. The relationships are independent from the order of its items: the relationship between cell state *n* and cell state *m* is the same of that between state *m* and state *n*. Only half of these relationships are needed for calibration, summing 30 calibration parameters.

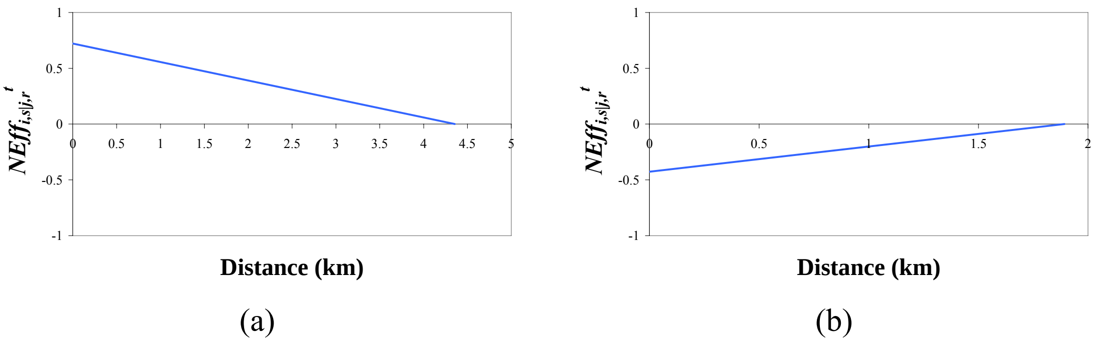


Figure 19 Neighbourhood effect relationships for two pair of states: (a) UHD-Rest, (b) N-UInd/N-UUrb

The basic principle of a geographic CA model is to simulate human behaviours that influence the evolution of the land use pattern over some territory. Thus, a deterministic approach is not the best way to simulate urban phenomena, as they strongly depends on the uncertain nature of human decisions. A stochastic approach is therefore preferable, introducing a stochastic perturbation to the assessment of the transition potentials for the cells.

These transition potentials are measured through a procedure that weights the influence of accessibility, land use suitabilities and neighbourhood interactions for each cell in each simulation time period. This transition potential can be expressed as follows:

$$Pot_{i,s}{}^{t} = f(Acc_i, Suit_{i,s}{}^{t}, NEff_{i,s}{}^{t}, \xi), i \in C; s \in CS \tag{26}$$

where, for each cell i from cell set C and for each state s from cell state set CS, $Pot_{i,s}{}^{t}$ is the transition potential for cell state s of cell i at time step t, Acc_i is the accessibility measure for cell i (constant during the entire simulation period), $Suit_{i,s}{}^{t}$ is the suitability value for cell state s of cell i at time step t, $NEff_{i,s}{}^{t}$ is the neighbourhood effect for cell state s of cell i at time step t considering its neighbourhood N_i, and ξ is the stochastic perturbation parameter.

Neighbourhood effect is the sum of interactions between land use in cell i and all the land uses of the neighbouring cells j that belong to neighbourhood N_i considering the neighbourhood distance parameter. It is calculated through Equation 27

$$NEff_{i,s}{}^{t} = \sum_{j \in N_i} NEff_{i,s|j,r}, i \in C; j \in N_i; s, r \in CS\,. \tag{27}$$

Interactions are obtained from the neighbourhood effect relationships (Figure 19 above). The value for interaction is determined by these relationships and depends on the distance between cells: considering the relationship depicted in Figure 19(a) for a UHD/Rest interaction, a cell in Rest cell state within a distance of, say, 3 km from a cell in UHD cell state will produce an attraction of about 0.20. These interactions are only limited by neighbourhood distance: the attraction or repulsion will only be considered if the cell is located within the neighbourhood.

The potential function is a weighted value of those components identified in Equation 26. The calibration factors are used to evaluate the relative importance of each of the components in the formation of the transition potential. In practical terms, Equation 26 is rewritten as

$$Pot_{i,s}{}^{t} = \left(\chi_{Pot} \times Acc_i + \nu_{Pot} \times Suit_{i,s}{}^{t} + \theta_{Pot} \times NEff_{i,s}{}^{t}\right) \times \xi \tag{28}$$

where χ_{Pot} is the calibration parameter for accessibility, ν_{Pot} is the calibration parameter for suitabilities and θ_{Pot} is the calibration parameter for the neighbourhood effect. ξ is the stochastic perturbation and is calculated through Equation 6, back in Section 2.3.2, now rewritten as follows:

$$\xi = 1 + \left[-\log(rnd)\right]^{\alpha_{pot}} \tag{29}$$

where *rnd* is an uniform random distribution in the interval]0,1[and α_{Pot} is a control parameter for the adjustment of the size of the perturbation (White and Engelen, 1993; Barredo *et al.*, 2003). Again, this term has a highly skewed distribution so that most values are near unity and much larger values occur only infrequently. The main goal is to introduce the stochastic behaviour of agents in each component of the weighted function.

The transition potential value is calculated for every cell state for each cell, being the final cell potential the highest value for the set of cell states. The transition potential value must be normalized for comparison reasons, being this procedure performed through the following expression

$$Pot_i^{\,t} = \max\left\{ \frac{Pot_{i,s}^{\,t}}{\left\| \sum_{s \in CS} Pot_{i,s}^{\,t} \right\|} \right\}. \tag{30}$$

where Pot_i^t is the transition potential of cell *i* at time step *t*.

This transition potential will determine which cell state the cell may transit to in the next time step. At each time step, the entire set of cells is ordered by its new transition potential so the model can be able to choose the highest ranked cells to apply the transition.

The cell state transition procedure consists of a verification of a set of conditions that will identify which cells are better positioned for moving to a new cell state, until land use demand is satisfied.

Land use demand can be defined in different ways. It can be defined exogenously or endogenously to the model. Integrated models are aimed to use spatial models along with several other types of models that are structured over non-spatial logics, such as econometric models for demand, demographic models, and transportation models. The purpose is to calibrate these models under the same and concurrent set of conditions[11], working together with a common workflow and several degrees of interdependence.

When defined exogenously, land use demand can be represented with the number of cells that have changed state in the problem modelled, what can be called a cell oriented demand approach. This method is widely used in geographic CA models (White and Engelen, 1997; Barredo *et al.*, 2003). This method is particularly suited for regular grid-based CA because regular cells have constant area, which means that an increase

[11] One of the main differences between sensitivity analysis and an optimization approach for calibrating a model is that optimization takes into account the interdependence of models parameters; on the contrary, sensitive analysis does not consider that interdependence between variables when varying a particular parameter.

on the number of cells in a given state is proportional to the increase on population or built area.

The model presented in this study deals with land use demand in a different perspective than common CA models. Irregular cells have different areas, with different values for population as well as for construction density. If the demand was assessed by the number of cells that had changed state, the sum of the newly occupied areas would not match the increase in population that generated the demand.

Land use demand is proportional to the increase of population, as well as to the variation of construction density: the accounting of new public facilities and spaces that the new built areas offers, in result of more demanding planning regulations, usually leads to a decrease on construction density.

For these reasons, the model is based on a population density oriented demand approach. The model calculates the increase of population during the reference period from reference data and distributes that population over the territory under consideration. Values for maximum construction densities (directly related with population density) are used to constrain the demand for cells. The occupation of a given cell in a given state must not be higher than the maximum value for population density for that state. Another constraint was considered to deal with the existence of different urban land uses. There is a value for the increase of area for all cell states in the reference period that can be assessed from the reference maps. The reference ratio between cell states UHD and ULD must be observed. The goal is to make the model verify the final values for this ratio.

After ranking the cell set by the higher transition potential, the model will assign a fraction of the population increase to the first best cell that is not under the construction transitory condition and that is admissible to the corresponding cell state. This procedure goes on until the entire population increase is totally assigned to the territory. The work flow of the entire CA process is depicted in Figure 20.

This process is highly representative of the land use dynamics: housing location is a decision commonly based on location characteristics that are taken into account in this transition potential function: accessibility, land suitability and neighbourhood interactions.

Industrial land use is modelled considering employment location. The assumption[12] that the industrial use is concentrated in dedicated industrial areas leads to the formulation of industrial demand to be based on employment density and not on population density. The process of calculating the transition potential is identical to the one described for urban land uses. Values for employment density were assessed based on field observation and estimated values for similar territories.

[12] As it was explained before, the Portuguese planning regulation and practice determines the constitution of industrial areas located outside the urban areas; these industrial areas are dedicated exclusively to industrial land uses. There are some exceptions though, such as retail parks and warehouses.

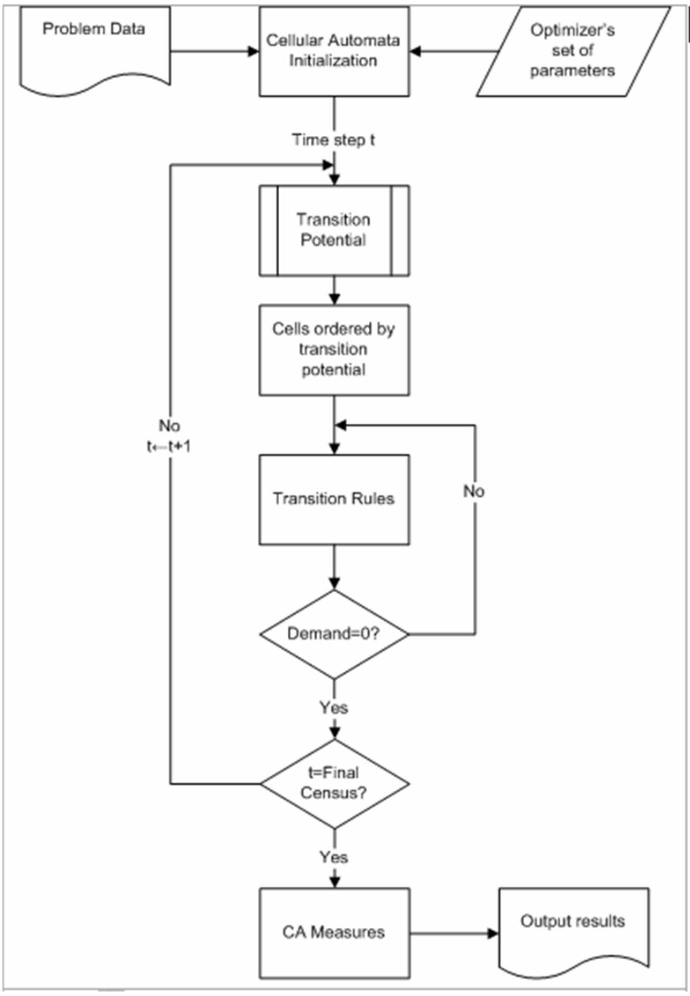


Figure 20 Workflow of the transition rules set

Four transition rules were considered. Generically, a cell in state *n* can transit to state *m* after verifying three of conditions: (1) the cell is not in the construction transitory condition; (2) there still is population to distribute over the territory; and (3) the cell is in state *n* and its highest transition potential in time step *t* is associated with state *m*.

Cells are chosen from the sub-set of cells that present the highest transition potential and are available for transition. Cells can transit from state 1 (ULD) to state 2 (UHD) and from state 4 (N-UUrb) to states 1 and 2. Non-occupied industrial cells (state 5, N-UInd) can transit to state 3 (IND), with the criterion changing from population to employment.

3.2.5 Time

The final issue regarding the CA model behaviour is related with the definition of the time step and consequent evolution throughout time.

The application of a CA model to the Portuguese context implies the consideration of planning regulations and commonly observed urban phenomena. This aspect was already considered in the definition of cell suitabilities, for example. It is extremely important that the model may simulate behaviours as close as possible to reality, otherwise there could be some lost of representativeness.

For this reason, modelling time step was 10 years, equal to the length of the period covered by the historical data available.

The choice is based on two main issues: data availability and land use dynamics. The use of spatial census units is founded on the existence of good and reliable demographic data for these moments, in the present case for two censuses. Between these two moments there is no official demographic information (only projections). On the other hand, land use changes take place in a particular way in these small municipalities. As the available land area is usually large when compared with the occupied land, land use change takes place through the occupation of new urban areas, rather than the transformation of existing urban areas. In fact, when a location is occupied by a UHD state, the probability of its transition for another land use such as ULD is almost (if not) null. So, when the cell changes for a specific state is highly improbable that in the next ten years (and probably for a much longer period) the cell will change to another admissible state.

Commonly the time step used for CA modelling is one year long. This is a very short time period, inappropriate for simulating urban change for these small municipalities in Portugal. However, the model is prepared to use different time steps of 1, 5 and 10 years.

3.2.6 Measuring cellular automata performance

Back in Section 2.3.5 the issue of CA model performance was discussed in a theoretical perspective. Different measures and their range of applicability were presented. After presenting the formulation of the CA model designed for this study in the previous sections of this chapter, it is important to present the measurement procedures that were implemented.

Due to the configuration of the problems that are under study the use of fractal dimensions such as radial dimension or perimeter-length scaling is not suited. These small urban areas are usually structured by a set of small localities surrounding the municipality main town. Unlike large metropolitan areas, where the territory is usually highly concentric with the city centre, the territories under study are not concentric with the municipality centre. Therefore, the use of fractal measures that are location dependent would not produce good and comparable fitness measures.

The choice was for contingency matrices and corresponding *kValue* measurements. The comparison of the simulation map with the reference map through a measure of similarity is a good comparison measure because it is oriented for the analysis of the entire territory as a distributed structure and not only as a centralized urban layout. But there are urban land uses (cell state Rest in the present model) that were not considered in the changing dynamics. Cells in this state will not experience any state change, remaining unaltered in the end of the simulation. Therefore, the consideration of the entire set of cell state for the calculation of the *kValue* value would produce a distortion on its significance. The number of cells at inactive states – cell state 6 (Rest) – is usually large, which would guarantee by itself a high value for the *kValue*.

To avoid this distortion, a modification of the *kValue* measure was considered, named *ModkValue*. This modification consisted of considering only the cell states that take part in the urban change dynamics, that is, not considering any cell state that is forced to remain unaltered throughout the simulation (and during real world evolution). The purpose is to eliminate cells that would concur for a large – though meaningless – agreement between simulation and reference situations.

Along with the analysis of the *ModkValue* measure, other measures were taken into account. The conditional *kValue* is a measure of the accuracy of a single cell state within the contingency matrix.

Another measure was produced to assess the quality of the simulation. The Lee Sallee form index was used only to allow the comparison with other studies where this form index was also used. This form index was calculated only for the real world case study.

3.3 The calibration process

Calibration of a model can be achieved through two main approaches: (1) performing a sensitivity analysis of each parameter's behaviour considering the other parameters fixed; or (2) running an optimization procedure for searching the calibration parameters that optimize the fitness function chosen for the model.

The first approach is based on a group of procedures that take place after the simulation. Visual comparison can be used to quickly assess the similarity between modelled and reference maps. After this initial procedure, a series of sensitivity analyses can be performed to evaluate how each parameter varies when the other parameters are controlled. This calibration approach becomes difficult to apply as the number of parameters increases[13].

The main reason for using an optimization approach is to ensure an extensive search for the parameters, leading to the best possible values for the parameters given the fitness function of the CA model.

In the present study the fitness function chosen to assess the quality of the model results was the modified *kappa* value, *ModkValue*, as described in the past section. The value of this measure should be as close to 1 as possible. This characteristic of the fitness function indicates the optimization approach as a good method to calibrate the model.

The number of calibration parameters is considerably high: there are three accessibility parameters (α_{acc}, β_{acc}, and γ_{acc}), four transition potential parameters (α_{Pot}, χ_{Pot},

[13] Each parameter represents a phenomenon that influences land use dynamics. The higher the number of parameters is, the more complex the problem gets.

v_{Pot} and θ_{Pot}), the neighbourhood distance parameter and the 30 parameters of the neighbourhood effect relationships.

For the problem at hand, the optimization method chosen was the Particle Swarm (from now on referred to as PS). This is a new optimization algorithm that has been given promising results for complex optimization problems.

PS has its origins in the simulation of social behaviours, in the study of the synchronized movement of bird flocks and fish schools. It is an optimization paradigm that simulates the ability of human societies to process knowledge (Kennedy, 1997). The member's movement in those groups is the result of the individual effort to maintain an optimum distance between him and his neighbours in the group (Parsopoulos and Vrahatis, 2002). The movement of the swarm takes place in a multi-dimensional solution space. The individual's successes influence their searches and those of their peers (Kennedy, 1997). Interaction in these groups enhances progress towards a solution, as particles benefit from their own knowledge and from their neighbours' knowledge. In this technique the aim is not the survival or the fittest but the joint effort of the swarm in finding the best solution.

Each particle has a memory of its past search history, usually called the cognitive component. It represents the natural tendency of individuals to return to environments where they experienced their best performance. Formally, it is the distance that the particle is from is personal best position. Each particle also knows the search results of the swarm, called the social component. It represents the tendency of individual to follow the success of other individuals. Formally, it is the distance that the particle is from its neighbours' best position (van den Bergh and Engelbrecht, 2005).

The formulation of PS is quite simple. It is based on a swarm of *p* particles that will fly through the search space during *n* iterations. The number of particles varies: it usually ranges from a few up to 60 particles (but there is no upper limit). The larger the swarm is, the better the search space is searched.

Each particle has *D* dimensions: in this application of PS to the calibration of the CA model each calibration parameter is represented by a PS dimension. Hence, there will be 38 dimensions for each particle.

Two vectors of particle data are necessary: one vector to store the particle's position and another to store its velocity (velocity represents the position change in each iteration). Considering that the search space is *D*-dimensional, then the i^{th} particle of the swarm can be represented by its positional, D-dimensional vector $X_i = (x_{i1}, x_{i2}, ..., x_{iD})^T$ and by its velocity, *D*-dimensional vector $V_i = (v_{i1}, v_{i2}, ..., v_{iD})^T$. The best previous position for the i^{th} particle is another D-dimensional vector, $P_i = (p_{i1}, p_{i2}, ..., p_{iD})^T$ and the global best of the swarm is represented by the last *D*-dimensional vector, $P_g = (p_{g1}, p_{g2}, ..., p_{gD})^T$, where *g* is the index for the best particle in the swarm. The velocity vector is updated according to Equation 31:

$$v_{i,d}^{t+1} = \omega v_{i,d}^{t} + c_1 r_1^t \left(p_{i,d}^{t} - x_{i,d}^{t}\right) + c_2 r_2^t \left(p_{g,d}^{t} - x_{i,d}^{t}\right) \tag{31}$$

where $v_{i,d}^t$ is the velocity of particle *i* on the *d*-dimension at iteration *t*, $x_{i,d}^t$ is the position of particle *i* on the *d*-dimension at iteration *t*, $p_{i,d}^t$ is the best individual position of particle *i* on the *d*-dimension at iteration *t*, $p_{g,d}^t$ is the best swarm position of best particle *g* on the *d*-dimension at iteration *t*, ω is the inertia factor, c_1 is the cognitive parameter, c_2 is the social parameter, r_1^t and r_2^t are random numbers uniformly distributed in [0,1]. The position vector is updated according to Equation 32:

$$x_{i,d}^{t+1} = x_{i,d}^{t} + v_{i,d}^{t+1}. \tag{32}$$

The flowchart for the PS algorithm is depicted in Figure 21. Notice that CA are an embedded process that is called as many times as the number of PS iterations multiplied by the number of particles.

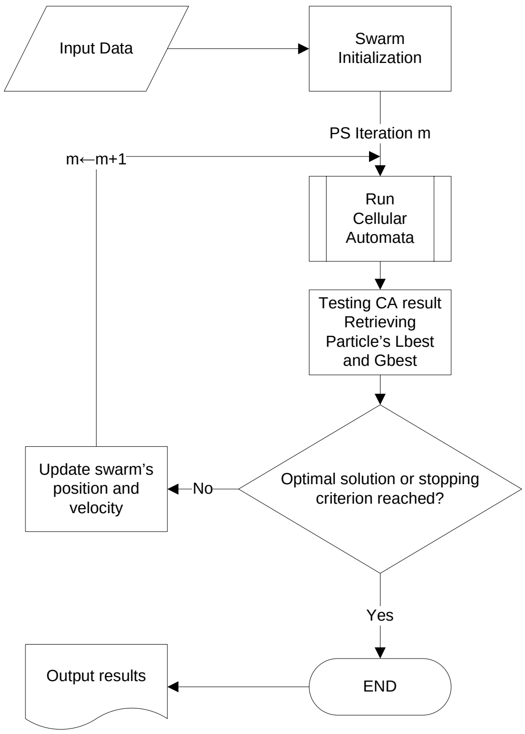


Figure 21 PS algorithm flowchart

The values for the inertia factor and for the cognitive and social parameters were studied by many researchers with the goal of establishing good sets of parameters in order to guarantee PS convergence. Van den Bergh and Engelbrecht (2005) and also Trelea (2003) present a series of applications of different parameter sets to a wide group of mathematical testing functions in order to study PS convergence. The purpose of the present study is far from being related with the calibration of a PS algorithm. For this reason a set of values was chosen from the literature with application purposes. The inertia factor was set at 0.729 and both the cognitive and the social parameters set at to 1.494 (Kennedy, 1997; Parsopoulos and Vrahatis, 2002; Trelea, 2003; van den Bergh and Engelbrecht, 2005).

The number of iterations was set at 150, but a stop criterion was established in order to save computational effort after the model had stabilized in a good solution. The model would stop when none of the particles could reach a solution that produces an increase in the individual best *ModkValue* of 0.1% for the last 5 consecutive iterations.

The method converges to optimum or near-optimum solutions after an initial exploratory search of the search space. Tests were made with different sets of particles (5, 30 and finally 60) and the larger the swarm is, the quicker it reaches good solutions. In many cases the optimum was reached in the first 20 to 30 iterations but there were other cases where the algorithm was still improving solutions after 50 or 60 iterations. Figure 22 depicts a typical trajectory for a particle for a given dimension *x(t)*.

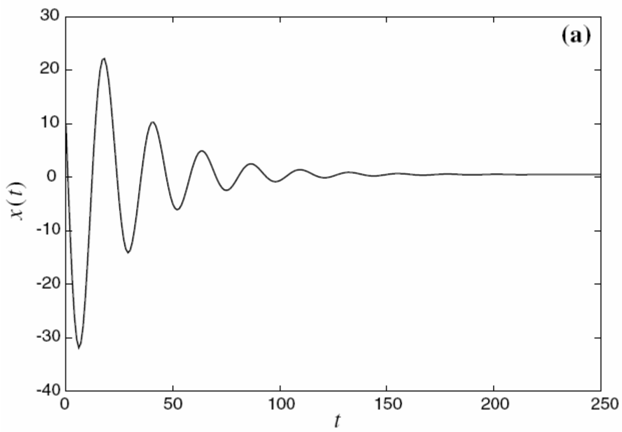


Figure 22 A typical particle trajectory in PS (van den Bergh and Engelbrecht, 2005)

This optimization process is based on the consecutive application of sets of calibration parameters that will generate different CA solutions. The CA model is applied for as many times as the number of particles multiplied by the number of iterations. Then, a large number of slightly different problems are solved, thus enlarging the search space.

3.4 Computational application

After discussing the methodology followed to implement a CA-based model it is important to present its computational application. The Windows application developed for this study was named SmallUrb|CA, as it is oriented for the study of small urban areas through a CA-based modelling approach.

The application was programmed in Visual Basic 6.0. A user friendly interface was developed to simplify the use of the model. The interface is designed to allow the introduction of information related with the parameters used in the calibration process and some major choices for the CA model, such as the type of demand, the fitness function or the duration of the time step. The user's interface is presented in Figure 23.

Because of limitations of the Visual Basic language in the field of image processing, maps must be generated outside the application by GIS software. SmallUrb|CA produces an output text file that is later used by GIS with this purpose.

The application is prepared to acquire input data from specially designed text files that define the problem: generic information on the problem (number of cells, number of states, etc), detailed information on the cells characteristics and the shortest paths matrix for the entire set of cells.

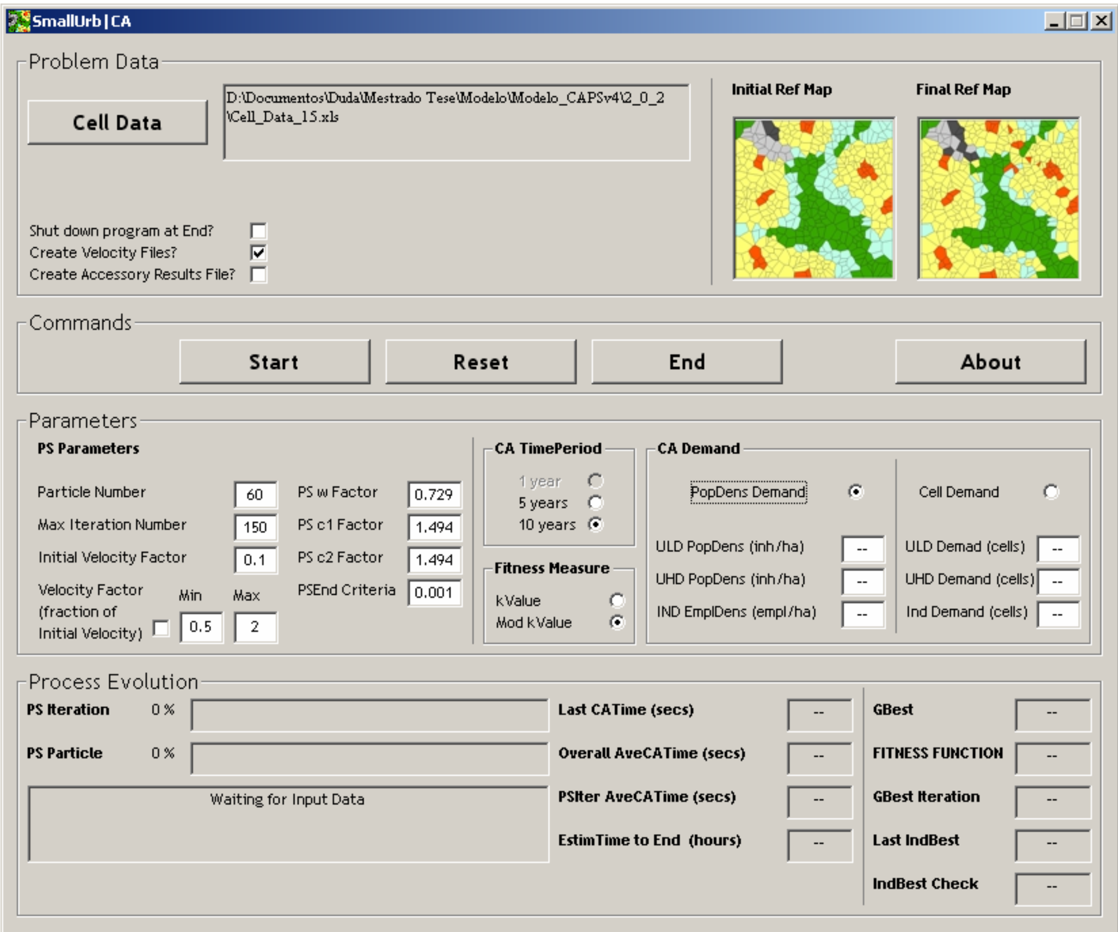


Figure 23 SmallUrb|CA interface

The application generates a series of output data files in text format. Different types of information are published in a series of some tens of files:

- a global results file with information regarding the optimum solution for the problem as well as all the measures of performance considered;

- a cell data file with the final characteristics of the cells; this file is used later by a GIS software to generate the modelled maps;
- a history file for registering the evolution of the fitness function value over the simulation;
- a calibration history file for each PS particle; these files allow the user to control the evolution of the calibration set of parameters for each particle in the swarm.

The application was used on a set of 20 theoretical test problems and on a real world case study. The generation of those test problems, as well as the model results, will be presented in the next chapter. The case study is presented in Chapter 5.

4 Test Problems

In the last chapter, the formulation of the CA model developed in this study was presented, where considerations about design and computational choices were discussed. This chapter is dedicated to present the set of theoretical test problems to which the model was applied to.

One of the goals of this study is to develop a new modelling approach based on CA theory for the analysis of the growth of small urban areas. To achieve this goal, it was considered necessary to evaluate the performance of the model for a group of theoretical test problems, generated in a randomly manner under a set of conditions that could replicate real world spatial structures, before using it to process any real world problem.

The use of these theoretical test problems was mainly oriented to exploring the performance of a CA model designed for simulating urban change phenomena in small urban areas. Before applying the model to real world case studies it is believed to be important the understanding of how such model would behave when dealing with these problems.

Another objective of using theoretical test problems regards the study of calibration procedures. Calibration parameters were not only related with CA, but also with the optimization algorithm used to calibrate CA. This last set of optimization parameters is extremely important for future application of the model.

After this first phase of calibration based on the study of the model performance, through its application to the mentioned theoretical test problems, it will be applied to a real world problem, in order to achieve the final aim of assessing the feasibility of a CA-based microsimulation approach for real world small urban areas. After analysing the main characteristics of a group of Portuguese municipalities, considering several aspects as urban structure, urban growth, location and dimension, the municipality of Condeixa-a-Nova was chosen. This case study will be presented in full depth in the next chapter.

4.1 Definition

The group of test problems generated for this study was produced by an algorithm that simulates spatial structures considering a set of conditions for the occupation of each cell. The main goal of this procedure is to generate spatial structures that can be considered similar to an average small Portuguese municipality, both in scale and in number of cells. Cells represent the census tracts that are currently used as the geographic unit of the statistical system. The algorithm produces land use occupation starting from a predetermined group of initial settlement centres[14]. However, the transition of state throughout the years is based on proximity to the main network and on a probability of transition based on a predetermined neighbourhood distance. This is to say that these problems are basically founded on a probabilistic procedure considering a small set of conditions (neighbourhood size and road infrastructure).

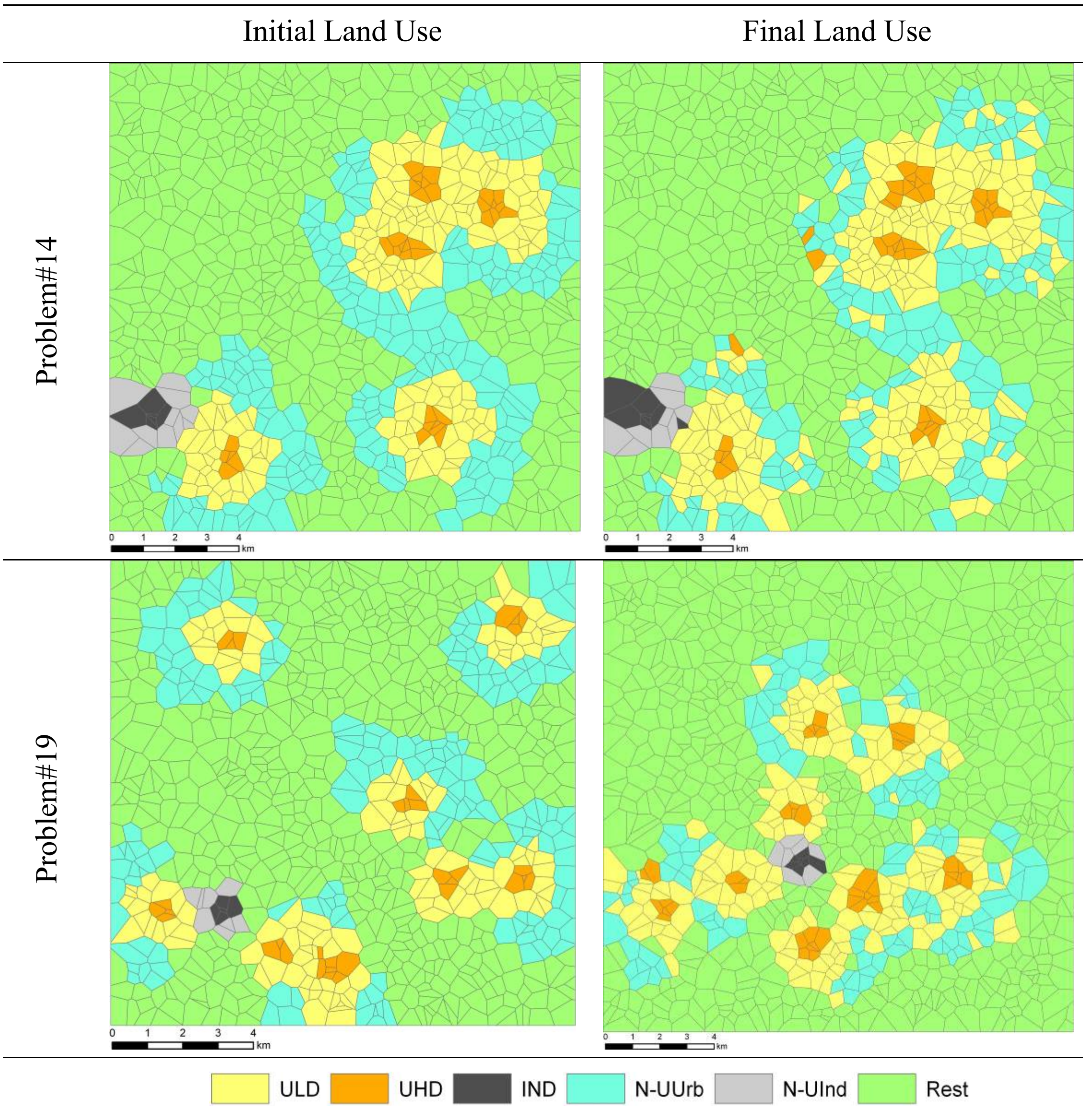


Figure 24 Two examples of theoretical test problems

The algorithm has three main phases. From that set of initial centres, randomly chosen between 4 and 10 settlements, which could assume cell state 2 (UHD) or, for

[14] A settlement centre is an initial seed for an urban area in the theoretical test problem.

only one case, cell state 3 (IND), the algorithm establishes the land use of the neighbouring cells based on a probability related to its distance to the settlement centre.

The process is repeated until all the cells have taken a cell state. Figure 24 depicts two examples of these theoretical test problems, both in their initial and final land use maps.

These problems where generated considering a maximum size of the territory (the measure of the side of each square territory) that ranges from 10 to 20 km. The number of cells varies between 800 and 1200. The conjugation of these two criteria made possible the generation of different patterns, considering both the dimension and the density of cells. There are some problems with high cell density and small size territory as well as high cell density for problems with a large maximum dimension. This diversity is expected to allow a series of prospective analysis over the model performance. Table 8 in page 90 summarizes the main characteristics of the set of problems. Average cell dimension ranges from 9.6 up to 47 hectares and territory size varies from 114 to 380 km^2. The increase of total ULD area varies from 14% up to 46%, while the same value for UHD ranges from 7% and 42%. There is no direct relationship between the variations of these two sums. The change to state UHD is made from cells that are both in state UHD and in state N-UUrb, while the change for state ULD is only possible for cells in state N-UUrb. This last state presents always a decrease on its total area, as expected. The total increase of built urban area (the sum of ULD and UHD areas) varies from 15% up to 44% with an average 27% increase.

Fact sheets for the entire set of test problems are presented in Appendix: Test Problems Fact Sheets. These fact sheets contain reference and simulation maps and main information about problem characteristics and results.

Table 8 Test problems characteristics

Problem	TotalCells	Max Dimension (m)	Average Cell Area (ha)	Area (ha)	1991 Population Densities				2001 Population Densities				Area variation 1991/2001					
					ULD	UHD	IND	N-UUrb	ULD	UHD	IND	N-UUrb	ULD	UHD	IND	N-UUrb	N-UInd	Rest
1	933	10681	12.23	11408	12.97	21.89	10.00	1.00	7.19	13.36	10.00	1.00	14%	22%	34%	-20%	-10%	0%
2	1033	15703	23.87	24658	11.69	23.80	10.00	1.00	7.23	11.95	10.00	1.00	26%	20%	32%	-26%	-5%	0%
3	1082	18214	30.66	33175	13.54	29.75	10.00	1.00	6.15	11.24	10.00	1.00	27%	37%	31%	-25%	-14%	0%
4	1132	10725	10.16	11503	10.40	25.71	10.00	1.00	7.27	14.93	10.00	1.00	20%	11%	12%	-40%	-3%	0%
5	1157	16981	24.92	28835	13.83	28.68	10.00	1.00	5.63	14.58	10.00	1.00	28%	7%	67%	-33%	-10%	0%
6	1182	13236	14.82	17519	12.26	21.66	10.00	1.00	6.19	14.22	10.00	1.00	20%	32%	10%	-36%	-3%	0%
7	807	19492	47.08	37994	10.69	24.64	10.00	1.00	6.75	13.87	10.00	1.00	30%	29%	85%	-20%	-24%	0%
8	832	15747	29.80	24797	14.12	27.61	10.00	1.00	7.32	13.51	10.00	1.00	19%	19%	37%	-24%	-9%	0%
9	1045	13875	18.42	19252	10.83	29.10	10.00	1.00	6.50	11.14	10.00	1.00	20%	25%	37%	-31%	-6%	0%
10	857	12003	16.81	14407	12.55	20.59	10.00	1.00	5.68	13.16	10.00	1.00	25%	10%	16%	-22%	-5%	0%
11	1069	10130	9.60	10262	14.26	22.08	10.00	1.00	7.06	15.18	10.00	1.00	17%	36%	20%	-16%	-6%	0%
12	882	18258	37.80	33335	10.98	23.57	10.00	1.00	6.24	12.81	10.00	1.00	17%	27%	35%	-12%	-14%	0%
13	1094	16386	24.54	26850	12.69	25.05	10.00	1.00	7.62	14.83	10.00	1.00	22%	42%	70%	-26%	-12%	0%
14	907	14514	23.23	21066	14.41	26.54	10.00	1.00	6.80	12.45	10.00	1.00	36%	23%	70%	-26%	-23%	0%
15	1119	12641	14.28	15979	11.12	28.03	10.00	1.00	5.98	14.47	10.00	1.00	19%	36%	36%	-37%	-6%	0%
16	932	10769	12.44	11597	12.84	29.52	10.00	1.00	7.36	12.10	10.00	1.00	26%	34%	24%	-28%	-8%	0%
17	1038	14833	21.20	22002	13.69	25.26	10.00	1.00	5.85	15.31	10.00	1.00	46%	24%	18%	-22%	-4%	0%
18	1144	18897	31.21	35710	14.55	21.01	10.00	1.00	6.54	14.12	10.00	1.00	35%	19%	35%	-40%	-12%	0%
19	850	12961	19.76	16799	10.41	26.75	10.00	1.00	7.23	12.93	10.00	1.00	27%	17%	35%	-20%	-12%	0%
20	950	17025	30.51	28985	11.27	22.50	10.00	1.00	5.72	11.74	10.00	1.00	30%	7%	52%	-25%	-11%	0%

4.2 Model results for test problems

The model was applied to the set of test problems described in the past section. Results included a series of data files containing both performance and calibration data and final cell state configuration in order to produce the final simulation maps, as described back in Section 3.4.

The analyses produced for evaluating the set of test problems results focus on the performance of the model and on the assessment of model behaviour considering the characteristics of test problem. Results were obtained from model runs on 2.9 GHz Pentium processor machines with 1 Gbyte of RAM memory using Windows XP operating system. The average running time was about 15 hours because the PS algorithm solved as many CA problems as the number of particles of the swarm multiplied by the number of iterations. Table 9 presents a summary of the results for the set of test problems.

Most of the problems achieved an optimum solution within the first 50 iterations, after which the algorithm started to be unable to improve individual best (and consequently global best), forcing the application to end when the sum of individual improvements higher than 0.1 percent for the last 5 iterations reached 0. The total increase achieved by the PS algorithm for the performance measure *ModkValue* was between 17 percent and 51 percent with an average 27 percent. These results show the ability of the PS algorithm to search the solution space.

Table 9 Main results for the set of test problems

Problem	*ModKvalue*	Min *ModkValue*	*ModkValue* Increase	*kValue*	Overall Accuracy	Number of active cells	Proportion of active cells	Number of changed cells	Number of matching cells	Similarity (%)
1	0.827	0.666	24%	0.907	0.897	594	64%	56	25	45%
2	0.770	0.621	24%	0.896	0.856	465	45%	57	24	42%
3	0.726	0.593	22%	0.868	0.842	609	56%	73	24	33%
4	0.770	0.642	20%	0.878	0.884	795	70%	93	50	54%
5	0.801	0.529	51%	0.927	0.874	222	19%	27	14	52%
6	0.784	0.580	35%	0.921	0.867	278	24%	34	15	44%
7	0.792	0.614	29%	0.912	0.869	306	38%	30	10	33%
8	0.841	0.672	25%	0.934	0.899	276	33%	27	11	41%
9	0.802	0.629	28%	0.926	0.881	285	27%	40	20	50%
10	0.803	0.664	21%	0.909	0.880	417	49%	40	17	43%
11	0.811	0.703	15%	0.910	0.882	533	50%	43	12	28%
12	0.865	0.723	20%	0.948	0.914	233	26%	21	8	38%
13	0.745	0.622	20%	0.859	0.848	716	65%	86	27	31%
14	0.735	0.556	32%	0.879	0.842	450	50%	72	34	47%
15	0.718	0.564	27%	0.834	0.857	853	76%	110	46	42%
16	0.811	0.638	27%	0.925	0.870	193	21%	21	10	48%
17	0.746	0.552	35%	0.903	0.841	277	27%	39	15	38%
18	0.791	0.578	37%	0.929	0.879	240	21%	34	21	62%
19	0.757	0.602	26%	0.893	0.848	356	42%	44	19	43%
20	0.781	0.624	25%	0.909	0.863	344	36%	38	15	39%

Global *ModkValue* results for the entire set of problems are depicted in Figure 25. These results can be considered good for a simulation process: 50 percent of the problems achieved a *ModkValue* around 0.800 or higher and 75 percent of them exceeded 0.750. As it was explained back in Sections 2.3.5 and 3.2.6, *ModkValue* is a measure of agreement between modelled and reference maps that does not take into account inactive cell states[15].

Figure 25 also presents the variation of the absolute *kValue* measure for the set of test problems. For 65 percent of the problems, the agreement exceeded 0.900 and 95 percent exceeded 0.850. These values are commonly accepted as good agreement between modelled and reference situations (Barredo *et al.*, 2003)[16]. Overall accuracy (the proportion of correctly classified cells, that is, the sum of cells located in the main diagonal over for the total number of cells) for the *ModkValue* measure also exceeded 0.850 for 75 percent of the cases. This is to say that the model showed a considerable capacity to simulate land use dynamics when dealing with theoretical small urban areas. This suggests that it can also deal properly with real world problems.

Another measure used to assess the performance of the simulation is the number of cells that have changed for the same state both in simulation and on reference maps. This value can be referred to as a proportion of the total number of cells that have changed state in simulation. For 80 percent of the problems achieved a matching proportion higher than 35 percent with 20 percent of the entire set achieving more than 50 percent. Although these proportions present low values, they can be considered a good indicator of the model's capacity to simulate urban change phenomena. The model was unable to match a large number of state changes. However, it was able to choose cells that were close to the ones whose change was not matched. These contiguous cells have similar values for transition potential because of similar accessibility and suitability conditions.

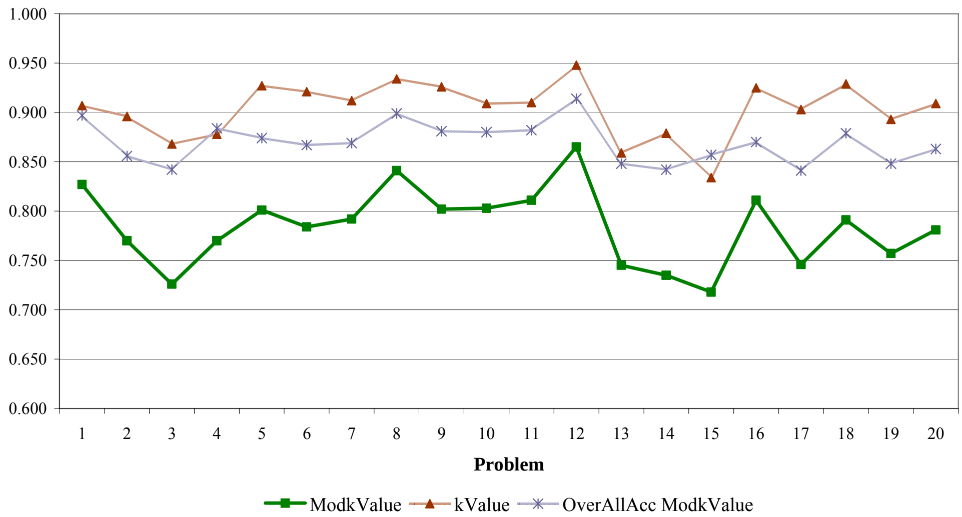


Figure 25 Global *ModkValue* and *kValue* results for the set of test problems

---

[15] As mentioned in section 3.2.6, the consideration of inactive cell states produces a distortion in the measurement of the model, because these cell states do not participate in land use dynamic and they concur for the agreement between modelled and reference maps.

[16] For example, in remote sensing the value of 0.850 for the *kValue* measure can be considered a good agreement between observed and reference data.

Another measurement of good agreement can be assessed from the conditional *ModkValue*, the chance of agreement for each cell state within the contingency matrix. The conditional *ModkValue* for urban cell states (ULD, UHD and N-UUrb) is depicted in Figure 26. It is notorious a variability of the values not only along the set of problems but also within cell states.

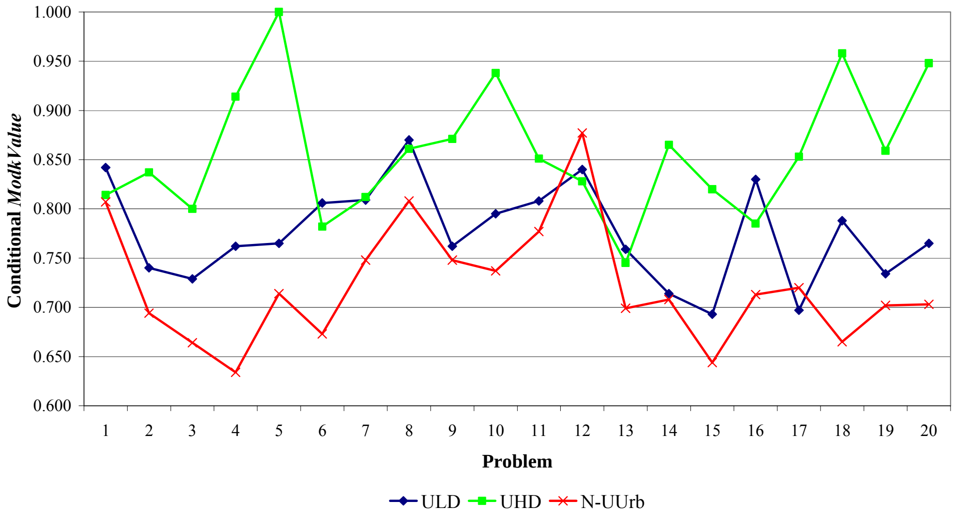


Figure 26 Global conditional *ModkValue* results for the set of test problem

Good results were obtained for UHD cell state, with 80 percent of the problems with more than 0.800 for the conditional *ModkValue*. For the other cell states lower values were obtained, particularly cell state N-UUrb, for which only 35 percent of the problems achieve values of conditional *ModkValue* higher than 0.800. It is interesting to note that 40 percent of the problems have differences for ULD and UHD conditional *ModkValue* smaller than 0.050 which indicates a better distribution of the cells that changed state.

For industrial land uses the results obtained show higher values of agreement (Figure 27).

The simplicity of the problems with regard to industrial land uses, contributes to these good results. For 85 percent of the problems IND conditional *ModkValue* exceeded 0.850 and the entire the set also exceeded this value for N-UInd conditional *ModkValue*. And 25 percent of the entire set achieved total agreement for the two industrial land uses, with both values of conditional *ModkValue* equalizing 1, which means that there was total agreement between simulation and reality for both land uses.

Another important parameter for model evaluation is the relationship between modelled and reference areas for each active cell state. They were compared by ratio Θ_S that is calculated through expression 33 and aims to assess how different the final model outcome is from the reference map in terms of total area occupied for each lad use. It is important to assess this ratio because land use demand is considered as a function of the population and the model is oriented for the distribution of population throughout the territory. The model tries to simulate land use through population density rather than through a predetermined number of cells for each cell state, regardless of their area.

$$\Theta_S = \frac{\left(\sum_{i \in C, s=S} \Omega_i^{Modelled} - \sum_{i \in C, s=S} \Omega_i^{Reference} \right)}{\sum_{i \in C, s=S} \Omega_i^{Reference}} \times 100 \tag{33}$$

where Θ_S is the ratio between areas for state S, $\Omega_i^{Modelled}$ is the sum of the areas of every cell i in state $s=S$ in the simulation, and $\Omega_i^{Reference}$ is the sum of the areas of every cell i in state $s=S$ in the reference map.

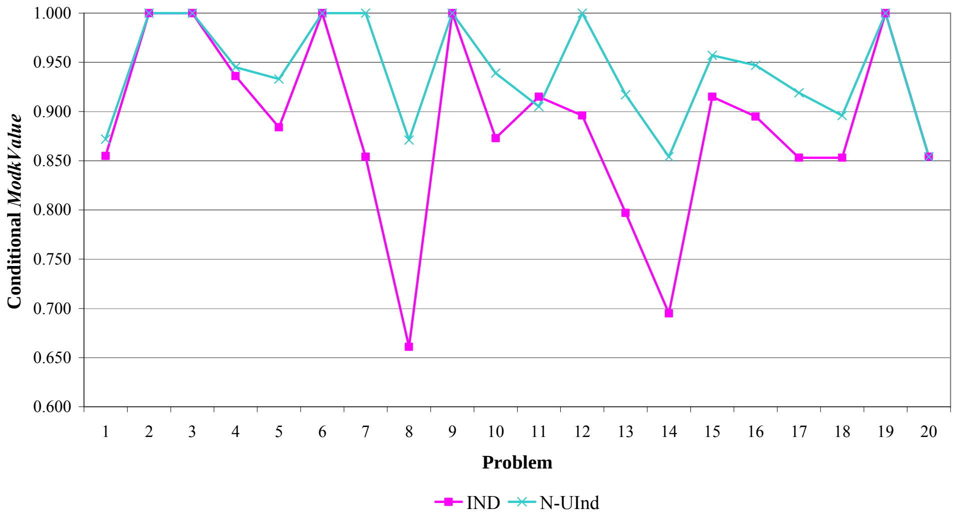


Figure 27 Global conditional *ModkValue* results for the set of test problems

The total area was determined for each cell state both for modelled and reference maps. The values for this ratio Θ_S are depicted in Figure 28 and in Figure 29 for every active land use for the set of test problems.

The variation of total area for ULD cell state takes values between 0% and +4% with an average of +1%; for UHD cell state this variation takes values between -13% and 0%, with an average -5%; for N-UUrb cell state this variation takes values between -4% and +1% with an average -1%. These values show that the model is capable of evolving to a situation similar to reality[17] in terms of total occupied area. Differences between modelled and reference maps results from the existence of similar values of transition potential for neighbouring cells as a result of similar accessibility conditions and neighbourhoods. The model chooses cells near or even directly connected to cells that have changed in the reference map but not in the simulation.

[17] Test problems involve what might be called “fake reality”: they are prototypes of spatial structures that aim to imitate real-world territories covering a wide range of characteristics (scale, cell size, cell density, population density).

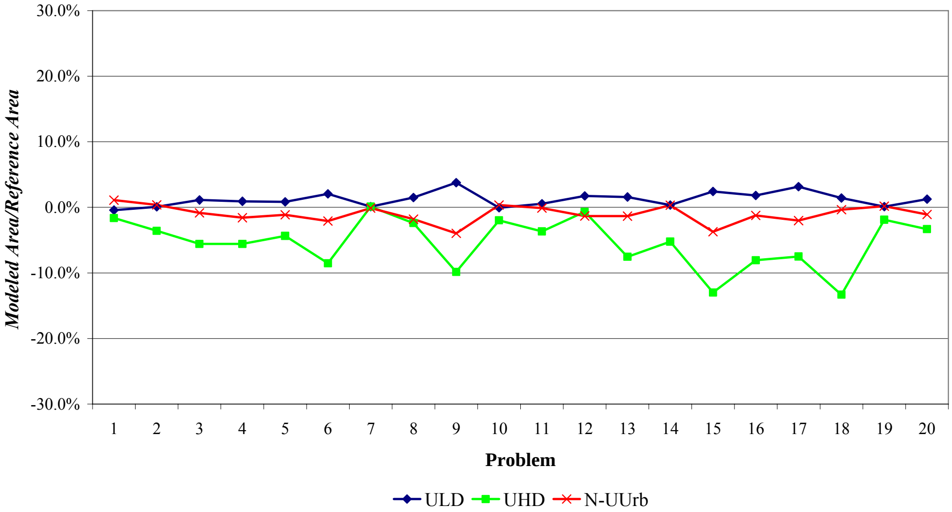


Figure 28 **Ratio between modelled and reference area by cell state, Θ_S**

The variation of total area for industrial land uses is significantly higher than the values obtained for urban land uses. The values for IND cell state vary from 0% up +49% with an average of +12%. For N-UInd cell state the variation of total area takes values between -20% and 0% with an average -5%. This behaviour for industrial land uses may be explained by an excessive simplicity of the test problems because there are few options for change in these land uses. Therefore, a small difference between modelled and reference maps (one or two cells) may result in a significant difference in the correspondent area ratios.

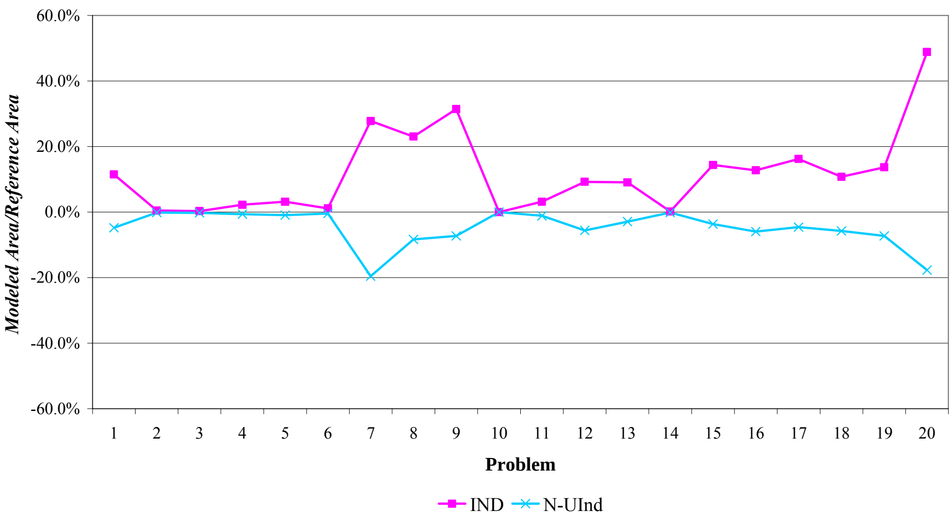


Figure 29 **Ratio between modelled and reference area by cell state, Θ_S**

Another important aspect to analyse regards the values obtained for the calibration. These values are presented in Table 10.

Neighbourhood distance varies between 2.1 and 8.0 km with an average 3.9 km. The accessibility calibration parameters show some variability in their results. For 45 percent of the problems α_{acc} (calibration parameter for the distance to the municipality main town, see Section 3.2.2) was the maximum value among all the accessibility parameters. The other parameters, the calibration parameters for the distance to the civil parish β_{acc} and to the industrial area γ_{acc} took the maximum value among them all for 35 percent and 20 percent each, respectively. This indicates a possible trend that may evidence the importance that the distance to the main functional centre (the municipality main town) has for the formation of the accessibility measure.

Table 10 Calibration parameters results

Problem	*ModKvalue*	Neigbourhood Distance (km)	α_{acc}	β_{acc}	γ_{acc}	χ_{pot}	v_{pot}	θ_{pot}
1	0.810	3.0	0.618	0.429	0.154	0.739	0.536	0.377
2	0.757	3.7	0.580	0.545	0.277	0.396	0.596	0.150
3	0.715	3.0	0.710	0.207	0.480	0.929	0.490	0.439
4	0.754	3.7	0.350	0.517	0.500	0.161	0.047	0.894
5	0.808	6.1	1.104	0.949	0.455	0.589	0.469	0.760
6	0.757	2.7	0.861	1.049	0.395	0.604	0.540	0.454
7	0.776	8.0	0.607	0.508	0.470	1.071	0.681	1.051
8	0.847	3.7	0.457	0.706	0.482	0.282	0.737	0.520
9	0.798	2.3	0.668	0.841	0.193	0.418	0.390	0.654
10	0.796	3.2	0.844	0.173	0.624	0.366	0.429	0.388
11	0.791	3.7	0.608	1.069	1.433	1.006	0.924	0.695
12	0.858	5.1	0.452	0.447	0.900	0.476	0.402	0.394
13	0.711	5.6	0.621	0.566	0.840	1.318	0.930	0.583
14	0.708	3.0	0.481	0.487	0.245	0.972	0.569	0.584
15	0.680	4.4	0.735	0.917	0.300	0.479	0.309	0.581
16	0.809	2.1	0.332	0.360	0.716	0.618	0.584	0.291
17	0.742	4.0	0.674	0.696	0.598	0.166	0.124	0.483
18	0.797	4.9	0.739	0.502	0.293	0.608	0.272	0.544
19	0.761	2.4	0.760	0.493	0.475	0.383	0.252	0.285
20	0.780	4.3	0.818	0.142	0.665	0.432	0.099	0.761
Highest Parameter			9	7	4	11	3	6

Regarding the transition potential calibration parameters (see Section 3.2.4) that aim to establish the importance of the three components of transition potential – accessibility, suitability and neighbourhood effect – the existence of a trend seems to be more evident. Both v_{pot} and θ_{pot}, calibration parameters for suitabilities and for neighbourhood effect respectively, took the maximum value among the transition potential parameters for only 15 percent and 35 percent of the problems, respectively. The majority of the problems – 55 percent – took its maximum value for the accessibility calibration parameter, χ_{pot}. This distribution suggests the existence of a trend that turns evident the importance of accessibility in the account of the transition potential. However, it is believed that more sophisticated measures would improve significantly the distinction between the influence of transition potential components, thus improving the representativeness of the simulation.

Finally, it is important to evaluate the relationship between performance and problem size in order to assess the model's ability to simulate urban change for small urban areas. Two measures were considered: the number of cells and the proportion of active cells over the number of cells (see Table 9). The relationship between the measure of performance of the model *ModkValue* and the number of cells is depicted in Figure 30.

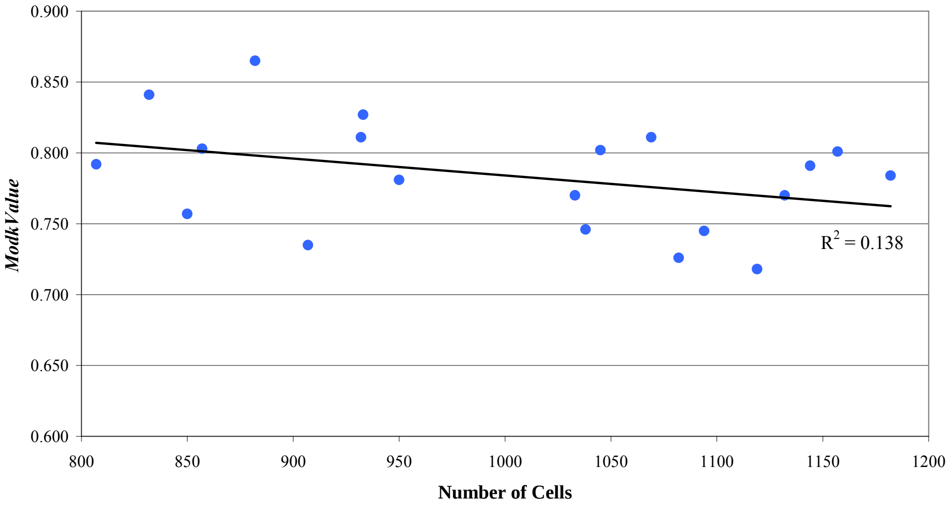


Figure 30 Relationship between *ModkValue* and number of cells

There is a small correlation factor R^2 of 0.138 which corresponds to a Pearson correlation ρ of -0.372. This value indicates the existence of a tenuous linear relationship between these two variables: smaller values of the number of cells are related to smaller values of performance.

The relationship between the performance measure and the proportion of active cells from the total number of cells is depicted in Figure 31. The proportion of active cells is considered to be a good parameter for assessing problem size because it only considers the number of cells that participate in land use dynamics. It presents a R^2 factor of 0.221 that corresponds to a Pearson correlation ρ of -0.470 indicating a significant linear correlation between these two variables.

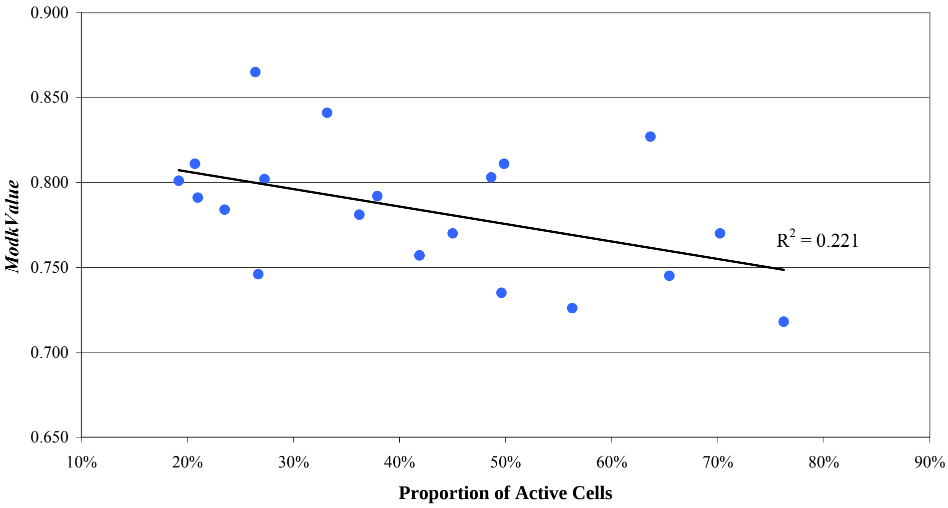


Figure 31 Relationship between *ModkValue* and the proportion of active cells

5 Real World Application

In addition to using the CA microsimulation approach on a set of theoretical test problems, it was applied to the municipality of Condeixa-a-Nova. This municipality was chosen because it has grown very fast in recent years.

5.1 Case study of Condeixa-a-Nova

Condeixa-a-Nova is a municipality located in the Portuguese central region, in the Baixo Mondego NUT3[18], near Coimbra, the region's largest city (see Figure 32). The municipality occupies about 140 square kilometres and had, according to the latest census (INE, 2001), a population of 15340 inhabitants, has his administrative centre located in the town of Condeixa-a-Nova.

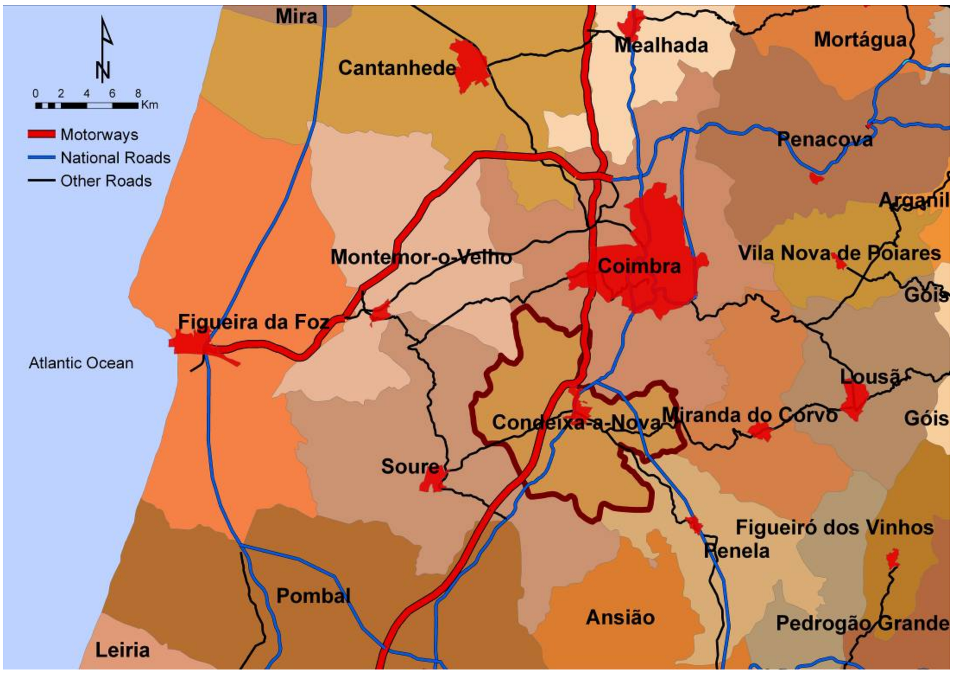


Figure 32 Location of Condeixa-a-Nova

[18] Level 3 Territorial Unit Nomenclature – a spatial sub-region with aggregate statistical data.

The population growth rate for the municipality in the period 1991/2001 was 17.8% – one of the highest rates within the national context – much larger than the population growth rate in Baixo Mondego NUT3 (population and its growth rate is represented in Table 11 for all the Baixo Mondego NUT3 municipalities and its variation for Condeixa-a-Nova is depicted in Figure 33). This value has even more significance considering the fact that the natural growth rate (birth rate minus mortality rate) has a negative value of -3.7% over the period between the censuses of 1991 and 2001 (see Table 12 and Figure 34).

Table 11 Population in NUT3 Baixo Mondego

Municipality	Population (inh)		Growth Rate (%)
	1991	2001	
Cantanhede	37140	37910	2.1
Coimbra	139052	148443	6.8
Condeixa-a-Nova	13027	15340	17.8
Figueira da Foz	61555	62601	1.7
Mira	13257	12872	-2.9
Montemor-o-Velho	26375	25478	-3.4
Penacova	16748	16725	-0.1
Soure	21704	20940	-3.5
Total	328858	340309	3.5

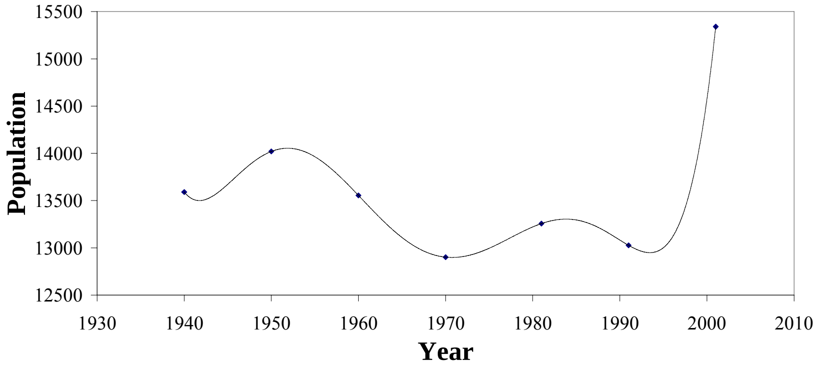


Figure 33 Evolution of the population for Condeixa-a-Nova

The town of Condeixa-a-Nova is located within 12 kilometres of the city of Coimbra central city. These two urban centres are linked by two high capacity roads, a main highway (IC-2, tool free) and a motorway (A-1, with a toll fee of €0.35, see Figure 32), which grants this connection high accessibility levels. Figure 35 represents the main road network that serves Condeixa-a-Nova.

The land use distribution in Condeixa-a-Nova territory is depicted in Figure 36. The majority of the municipality territory is occupied with National Ecological Reserve (REN), National Agriculture Reserve (RAN), forestry and other agricultural land uses of different types.

Table 12 Population indicators for NUT3 Baixo Mondego municipalities

Municipality	Birth Rate (%)	Death Rate (%)	Natural Growth Rate (%)	Internal Migration Rate (%)	Total Growth Rate (%)
Cantanhede	9.7	12.8	-3.1	5.2	2.1
Coimbra	12.3	10.6	1.7	5.1	6.8
Condeixa-a-Nova	11.4	15.1	-3.7	21.5	17.8
Figueira da Foz	10.2	13.3	-3.0	4.7	1.7
Mira	9.9	11.9	-2.1	-0.9	-2.9
Montemor-o-Velho	9.2	13.3	-4.0	0.6	-3.4
Penacova	9.3	13.0	-3.7	3.6	-0.1
Soure	8.1	14.7	-6.6	3.1	-3.5
Total	10.8	12.2	-1.4	4.9	3.5

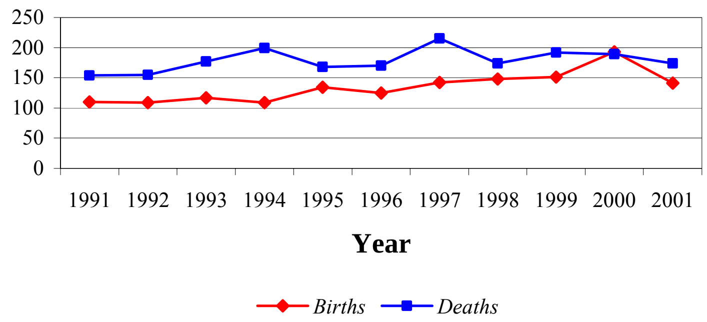


Figure 34 Evolution of births and deaths for Condeixa-a-Nova

Industry land use is located in predetermined industrial areas. Legally, it is impossible to build industrial facilities outside these areas. Only a reduced number of activities also considered as industrial such as automobile repairing and small warehouses can be located within urban areas.

Urban areas registered an impressive 48% increase between 1991 and 2001. In 2001 they occupied about 10% of the totality of the territory, a significant value considering the relative dimension of Condeixa-a-Nova when compared with other municipalities across the country where high growth rates were observed.

Commerce and other services are located entirely within the urbanized areas, occupying the ground-floors of centrally located buildings. There are no major commercial areas that could impose the consideration of a cell state for this use.

A series of assumptions had to be made to formulate the problem for applying the CA model. The first issue regards the definition of the cell and of the cell lattice. As it was already deeply discussed, the irregular cell CA model is based on the intersection of spatial census units, which holds the demographic information, with the urban areas in 1991 and in 2001. From this intersection results a cell lattice which has its grounds in the combination of urban structure and demographic information and is depicted both in

Figure 37 and in Figure 38 (a detailed view of the centre of the town of Condeixa-a-Nova).

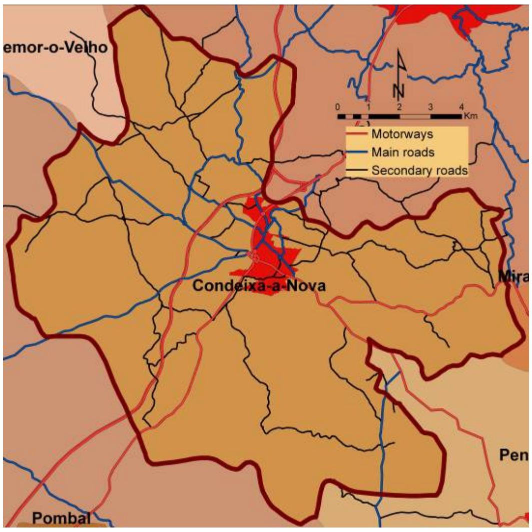


Figure 35 Road network of Condeixa-a-Nova (2001)

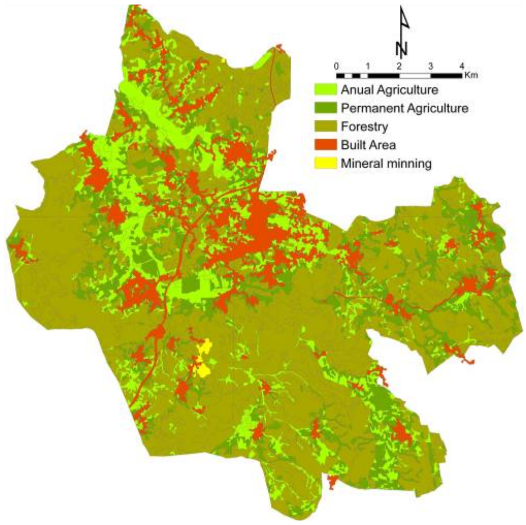


Figure 36 Aggregated planning map of Condeixa-a-Nova (2001)

Land use classification was aggregated in order to meet the set of six aggregate classes that support the model, as described in Chapter 3. This aggregation was made according with the aggregate set of land uses considered by the SmallUrb|CA were classification was based on construction densities. Figure 39 and Figure 40 depicts land use maps both for 1991 and for 2001 that are the result of the aggregation of the more detailed set of cell states.

However, as a result of the existence of available data for population densities, it was necessary to convert values of built density to population density. For this, it was considered values of 2.5 inhabitants per household for 1991 and 2.2 inhabitants per household in 2001. It was also considered values for urban high density (state 2 UHD) of 10 dwellings per hectare for 1991 and of 6 dwellings per hectare for 2001. For cell state 1, urban low density (ULD) these values were assessed in 5 houses per hectare in 1991 and 3 houses per hectare in 2001. The reduction of population density in the period (also considered for ULD cell state) is a result of two factors: (1) there was a reduction of the average number of members per household (from 2.5 to 2.2 inhabitants per household); and (2) built areas in 2001 included much more public facilities and public space areas than in 1991, thus decreasing population density in 2001. These public areas increased mainly due to more demanding standards of quality of life. These new standards imposed the consideration of more and better public facilities and spaces as well as a decrease in construction density.

From 1991 to 2001 the variation of total area for UHD cell state was enormous, achieving an impressive 351 percent as it is showed in Table 13. ULD cell state registered a small increase of 2 percent in the total state area. This difference between UHD and ULD cell states is related with the change in population density over the period.

Final UHD population density (in 2001) is close to initial ULD population density (in 1991).

Table 13 Condeixa-a-Nova's problem main characteristics

Number of cells	Maximum Dimension (m)	Avgerage Cell Area (ha)	Area (ha)
1433	15000	9.63	13803

Population Density by State 1991 (inh/ha)	ULD	UHD	IND	N-UUrb
	12.50	25.00	10.00	1.00

Population Density by State 2001 (inh/ha)	ULD	UHD	IND	N-UUrb
	6.60	13.00	10.00	1.00

Total Area by State 1991 (ha)	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	914	33	24	526	130	12176

Total Area by State 2001 (ha)	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	936	147	29	389	125	12176

Area variation 1991/2001	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	2%	351%	23%	-26%	-4%	0%

Figure 37 Cell space for Condeixa-a-Nova

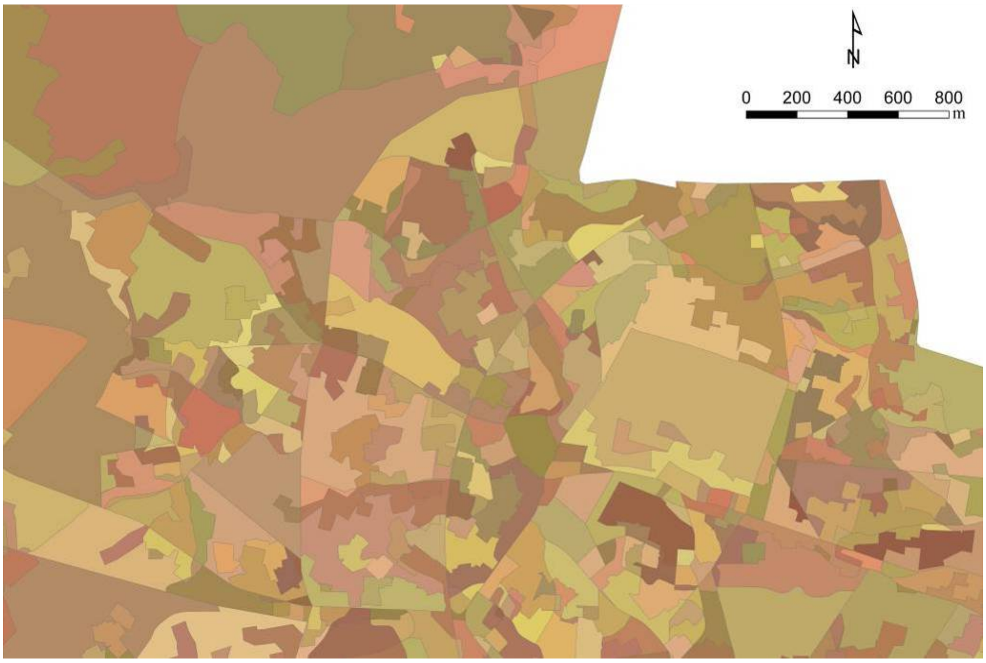


Figure 38 Detail of the cell space in the town of Condeixa-a-Nova

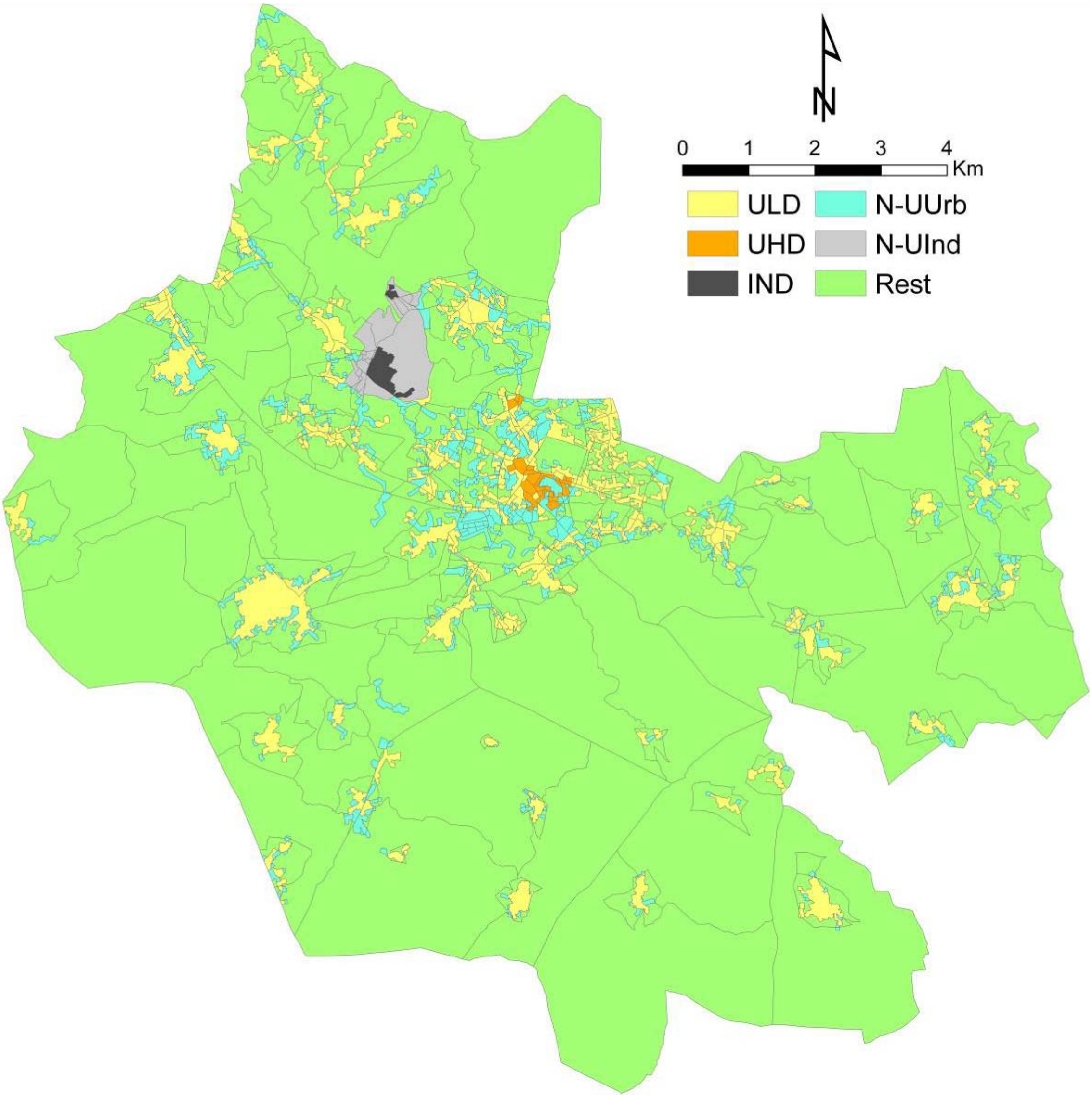


Figure 39 Reference land use map in 1991

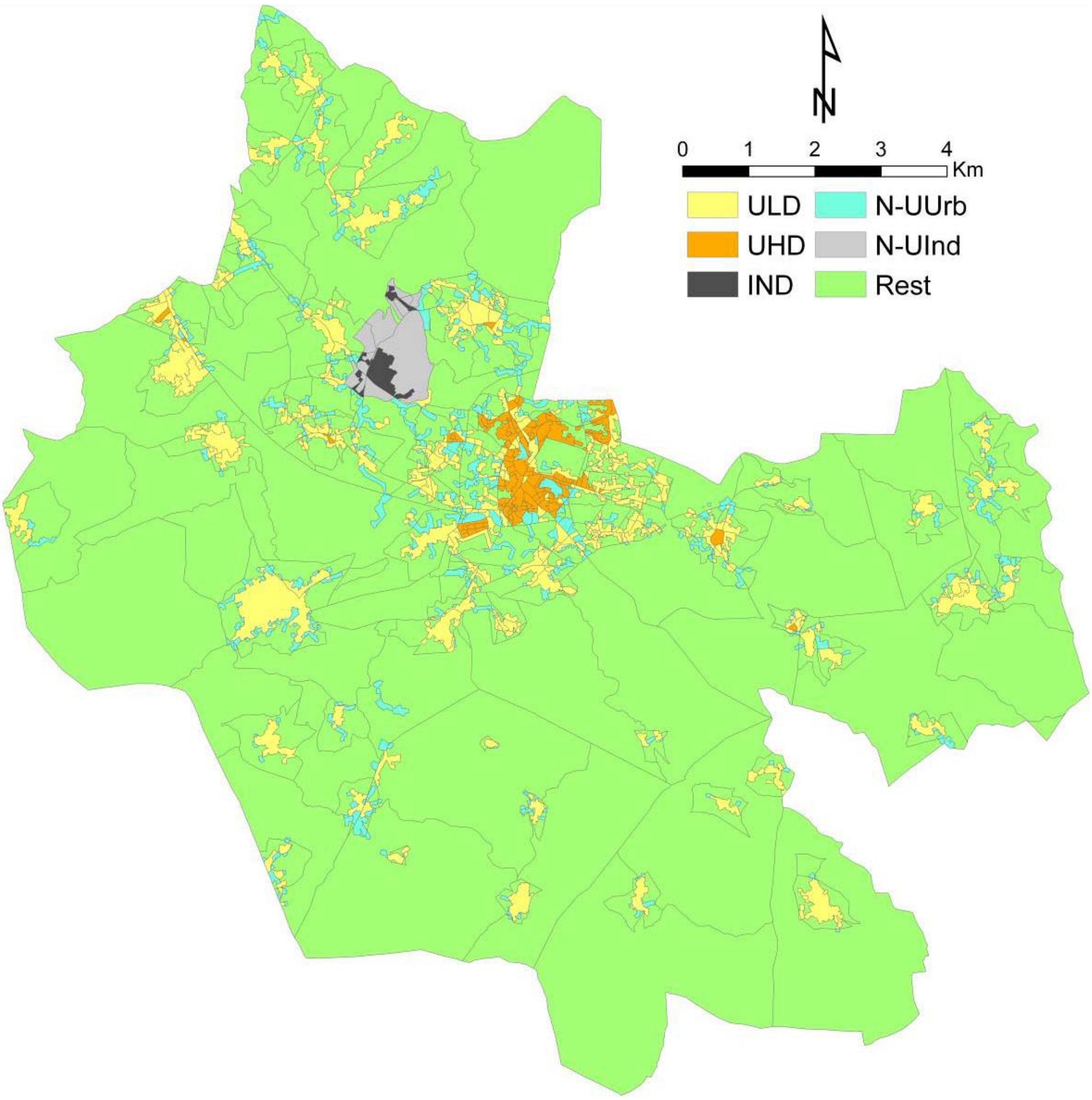
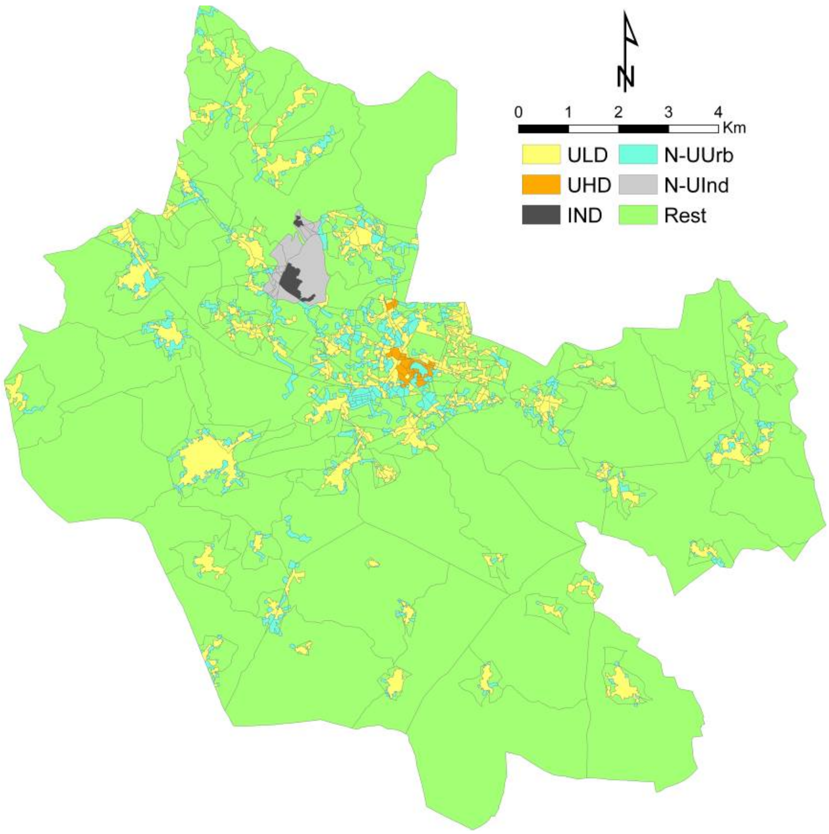
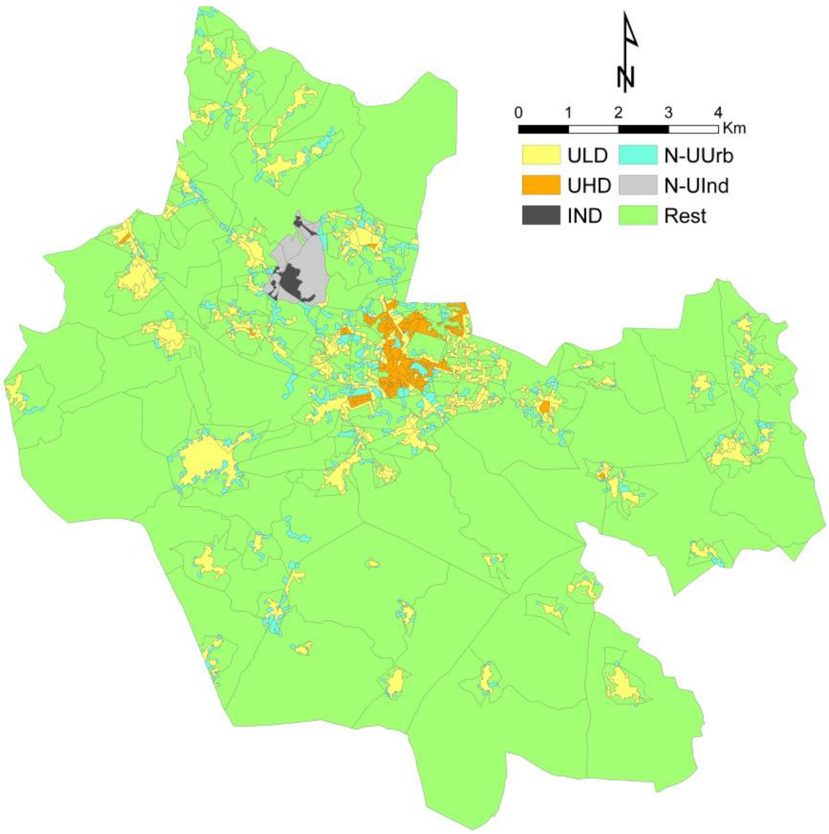


Figure 40 Reference land use map in 2001



(a)



(b)

Figure 41 Evolution from reference land use map in 1991 (a) to reference land use map in 2001 (b)

5.2 Model results for the case study

5.2.1 Calibration procedure

The model achieved an optimum solution for the case study after 54 iterations, when it became unable to improve individual best (and consequently global best), thus forcing the application to end. The optimization algorithm was able to fully exploit the search space and improved *ModkValue* by 184 percent reaching the final value of 0.621, as showed in Table 14. This value is quite lower than the average value for the set of test problems. *kValue* achieved the value of 0.774. The ratio between *ModkValue* and *kValue* can be interpreted as an indicator of how good the agreement for active cell states is: the closest *ModkValue* is to *kValue* the better the general agreement of the simulation. For the set of test problems this ratio had an average value of 0.867 while for the case study it decreased to 0.802. It means that the model was not able to achieve an agreement for active cell states as good as it achieved in average for the set of test problems.

Table 14 Simulation results for Condeixa-a-Nova

ModKvalue	*ModkValue* Increase	*kValue*	Overall Accuracy	Number of active cells	Proportion of active cells
0.621	184%	0.774	0.771	1015	70%

Number of changed cells	Number of matching cells	Similarity (%)
94	19	20%

Conditional *ModkValue*	ULD	UHD	Ind	N-UUrb	N-UInd
	0.718	0.362	0.598	0.610	0.831

Θ_S (1991/2001)	ULD	UHD	Ind	N-UUrb	N-UInd
	-10%	33%	42%	11%	-10%

Neigbourhood Distance (km)	α_{acc}	β_{acc}	γ_{acc}	χ_{pot}	v_{pot}	θ_{pot}
4.0	0.445	0.238	0.568	0.130	0.341	0.933

Table 15 shows the contingency matrices determined with reference to both cell types and cell areas. The model was calibrated using the *ModkValue* associated to the first matrix. But it is interesting to present the correspondent contingency matrix based on area units (it was calculated along with the Lee Sallee form index). The main difference occurs for the conditional *ModkValue*. The results are considerably worse for UHD cell state in the second matrix, once again as a result of the distortion induced by large urban cells. However, the value for *ModkValue* is similar for both matrices, being slightly higher (0.643) when using cell areas are considered. The Lee Sallee index for the case study was 0.778, which indicates good agreement between reference and modelled urban forms.

The proportion of active cells over the total number of cells is also below the expected value obtained from the relationship[19] illustrated in Figure 31. However, it is in line with the trend of the test problems, which suggests that the result obtained for the case study can be improved after the introduction of small changes to the model.

It was mentioned in the past section that there was an enormous difference in the variation of total area per cell state from 1991 to 2001 between ULD and UHD cell states; at the same time, the difference between the values for ULD population density in 1991 and UHD population density in 2001 is very small (less than 1 inhabitant per hectare). This small difference may have some influence on the final simulation results because it reduced the capacity of some cells in state ULD to receive new population.

Table 15 Contingency matrices for the case study (a) considering cells types; (b) considering cell areas

	Reference Map						
Modeled Map	ULD	UHD	IND	N-UUrb	N-UInd	Σ	Conditional *ModkValue*
ULD	328	43	0	20	0	391	0.718
UHD	33	40	0	18	0	91	0.362
IND	0	0	3	0	2	5	0.598
N-UUrb	74	40	0	402	0	516	0.610
N-Uind	0	0	2	0	10	12	0.831
Σ	435	123	5	440	12		
OverAll Acc	0.771						

Unit - cell

(a)

	Reference Map						
Modeled Map	ULD	UHD	IND	N-UUrb	N-UInd	Σ	Conditional *ModkValue*
ULD	735.9	66.6	0.0	27.9	0.0	830.4	0.732
UHD	125.0	45.7	0.0	22.0	0.0	192.7	0.161
IND	0.0	0.0	25.4	0.0	5.7	31.1	0.809
N-UUrb	75.4	34.6	0.0	339.5	0.0	449.5	0.678
N-Uind	0.0	0.0	39.9	0.0	119.2	159.1	0.728
Σ	936.30	146.90	65.30	389.40	124.90		
OverAll Acc	0.778						

Unit - hectare

(b)

Another cause to this simulation results can be imputed to an inappropriate problem definition. The criterion used to define cells was based on the intersection of census tracts with urban areas for 1991 and 2001. The cell space for the centre of Condeixa-a-Nova (Figure 38) can be considered as a good representation of the urban structure: cells are strongly disaggregated and their size is more homogeneous. Outside the town centre, particularly in the other urban areas, cells are larger and, in some cases, there is

[19] These two variables, *ModkValue* and proportion of active cells, present a correlation value of -0.470. It is believed that through a series of small improvements a better correlation could be obtained.

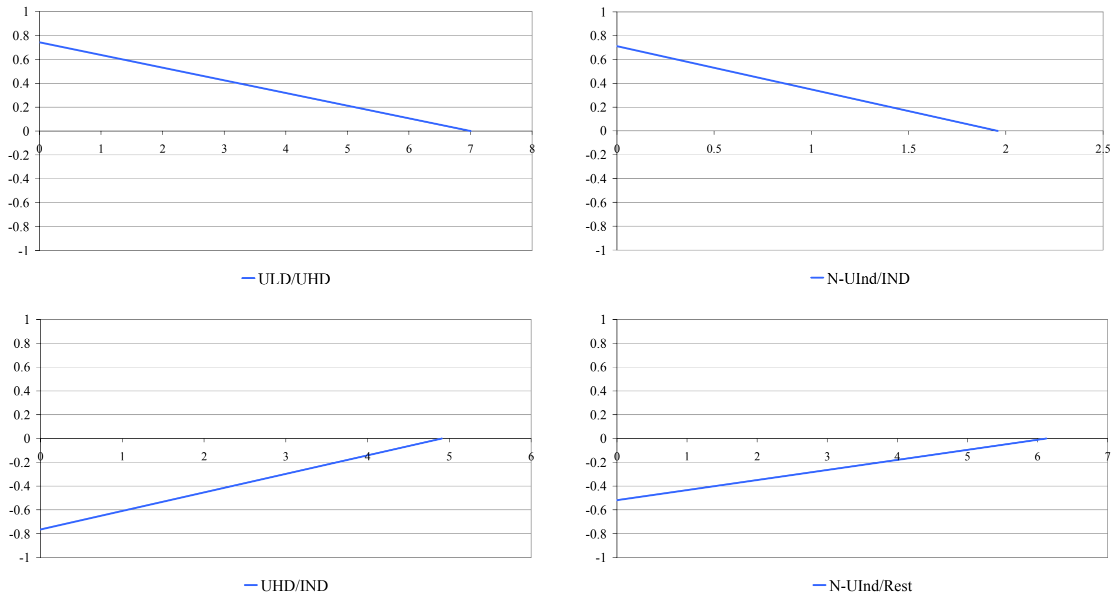
only one large cell representing the entire village centres (see Figure 37). These cells are particularly well located both in terms of accessibility and neighbourhood, usually neighboured by cells in cell state Rest (agriculture, forestry and other natural spaces). Due to these conditions they tend to have high transition potentials during simulation, granting them good possibilities for changing state. Their larger areas ensure that a large amount of population may be assigned to them without infringing the UHD population density limit for 2001. The same behaviour produced the result for IND cell state where a large cell was chosen making the model able to allocate a large number of employments.

This issue may also be considered for Rest cell state. Cells in Rest state tend to be larger than other cells, firstly because census tracts are larger in non-urban areas and secondly because the model deals with aggregate cell sates. The attraction effect that Rest cell state produces on urban cell states is normally high. Therefore, large Rest cells that enfold small urban-located cells may alter results by granting the later a high transition potential, thus enhancing their possibilities of being chosen for transition.

To tackle this situation cell dimension must receive more attention in problem definition. When facing large cells in small urban areas, it is believed that they must be divided into a number of cells with average area values, considering the remaining set of cells. This problem will be addressed in the concluding discussion.

Another consequence of what was expressed above is the larger difference for the ration Θ_S the case study presented when compared to the results obtained for the set of test problems. The large value of +30 percent registered for UHD cells along with the negative value of -6 percent for ULD cells and +3 percent for N-UUrb cells is a direct result of the choice by the model of larger cells. This choice was made mainly for ULD cells: in simulation about 127 hectares (a total of 34 cells) changed from ULD to UHD cell state, while in reality only 68 hectares (48 cells) changed this way. On the opposite direction goes the change from N-UUrb to UHD cell state: simulation assigned only about 32 hectares (30 cells) to this change while in reality about 46 hectares (48 cells) were changed from the former cell state to the latter. The same behaviour is observed in the change from N-UUrb to ULD cells: the simulation assigned 44 hectares of 27 cells to this change while in reality there were 91 hectares that changed between these to states for a total of 85 cells. These values corroborate the idea that the model is hungry for well located large cells in ULD cell state: the model assigns a large amount of population in large ULD cells and only a small amount is assigned to smaller ULD and N-UUrb cells.

The number of cells whose transition is well determined by the simulation is also low. From a total of 94 cells that suffered state transition during simulation only 19 saw its state transition matched between simulation and reality, corresponding to 23 hectares (summing all the urban state changes, excluding industrial uses) in about 205 hectares that changed in reality (again, the sum of all urban state changes). The matched area is only about 11.2 percent of the total area that changed for urban cell states. This is a distortion clearly related to the appetite of the model for large ULD cells.



Vertical axis – $NEff_{i,S|j,r}$ (Section 3.2.4) Horizontal axis – Distance (km)

Figure 42 Neighbourhood effect relationships for Condeixa-a-Nova

In Figure 42 a group of four neighbourhood relationships is depicted. These values also present logical behaviours, not only in terms of attraction/repulsion but also regarding the distance of influence they establish.

The results for the set of calibration parameters were not in line with the results for the group of test problems. The neighbourhood distance was set to 4 kilometres, a value that can be accepted as a perceived neighbourhood distance for a municipality with the size of Condeixa-a-Nova.

Regarding accessibility, the most important parameter is related to the accessibility to the industrial area, γ_{acc}, whereas accessibility to the municipal main town, α_{acc}, was the most important parameter for the test problems. Regarding the calibration parameters for the transition potential, the highest value was obtained for θ_{pot}, related to the neighbourhood effect. This result is also different from the one obtained for the set of test problems where a trend giving more importance to accessibility conditions was detected.

The final simulation land use map is depicted in a larger scale in Figure 43. This map is also compared to the initial reference land use map (Figure 44) to present how the simulation evolved from initial conditions. It is clear that the model produced a quite peculiar result. Rather than reinforcing urban concentration as the driven force of urbanization for Condeixa-a-Nova, as it can be seen in Figure 41, the model favoured the emergence of surrounding urban centres with high transition potentials[20].

It is also necessary to compare the reference map of 2001 with the simulation results (Figure 45). It is notorious that the model failed to reproduce the growth of Condeixa-a-Nova's town centre.

Some issues can be pointed out as causes for this result. Firstly, the distortion produced by the inappropriate definition of the problem regarding large cells in central areas. Secondly, the use of better accessibility and suitabilities measures capable of better differentiating close cells. Although more studies must be made, it seems that there is a trend associated to the behaviour of the calibration parameters of accessibility, underlying the importance of the municipality's main town. It is necessary to establish a more sophisticated measure that could be able to differentiate closely located cells. Suitability measure must also be improved. This component of the transition potential must be assessed in a way that ensures a good distinction between land use suitabilities for nearby cells. Physical land characteristics (slope and hill shade for example) can also be used to better evaluate land use suitabilities.

5.2.2 Prospective analyses

After a calibration of model parameters for the present case study, a simulation was made for land use change within a 20 years time horizon. The prospective simulation was made for year 2011 and to 2021 starting from the reference situation of 2001. A population growth rate of 10 percent was considered for the 2001/2011 and 2011/2021 periods.

Figure 46 and Figure 47 depicts land use maps for 2011 and 2021 respectively. It is interesting to note that although the model was unable to correctly match changes in numerous cells during calibration phase, when applied in a prospective simulation it was able to produce likely urban change behaviour.

Simulation results led to the densification of the municipality main town, especially in those areas that are close to the industrial area (this can also be seen in Figure 48 and in Figure 49). This behaviour was expected and verified both for 2011 and for 2021. Calibration parameters obtained on the first modelling phase indicated that neighbourhood effect was by far the main component of transition potential (θ_{pot} was high as compared with the other two parameters, χ_{pot} and v_{pot}). They also showed that the distance to the industrial area was the main contributor to the formation of the accessibility measure (see Table 14).

Although these prospective simulation results can be seen as likely land use layouts, they were obtained from a set of calibration parameters that may not correctly simulate urban phenomena.

[20] The association between transition potential to urban potential is a basic concept of this study as it is a rather intuitive concept that can be easily observed in real world problems.

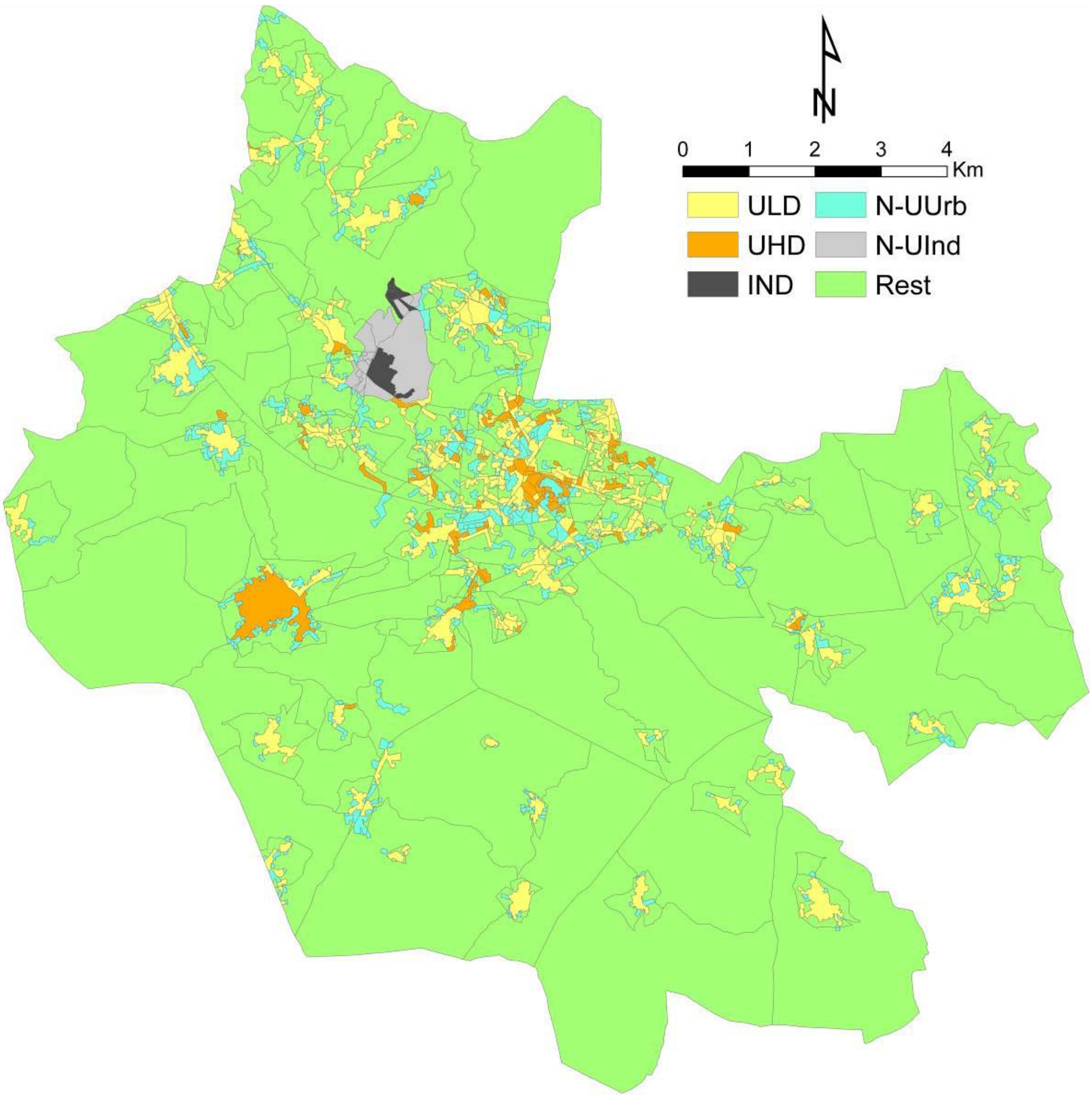
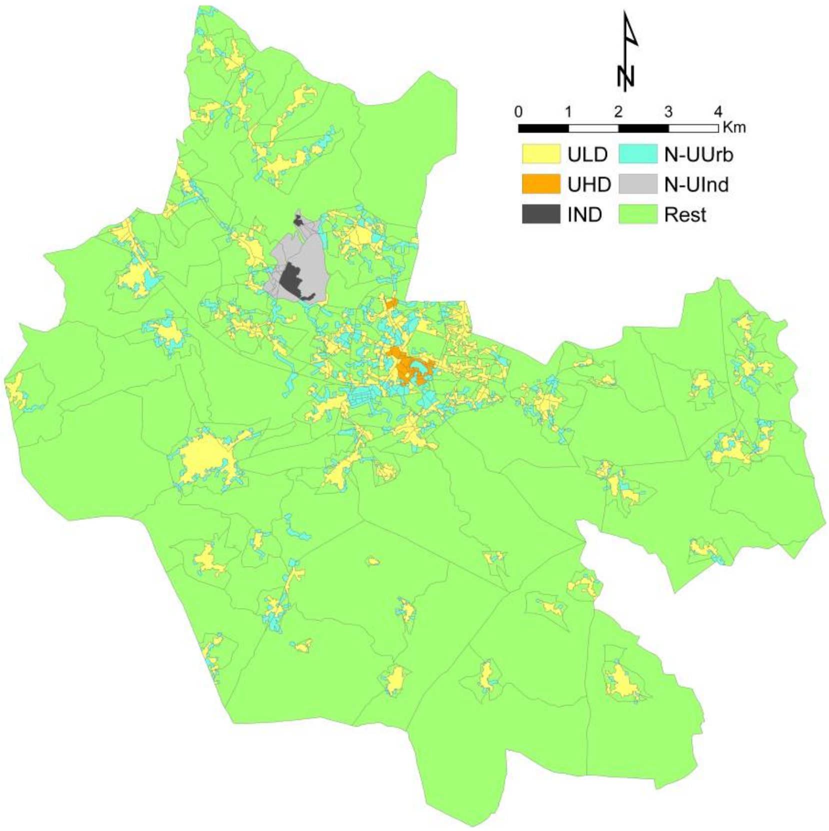
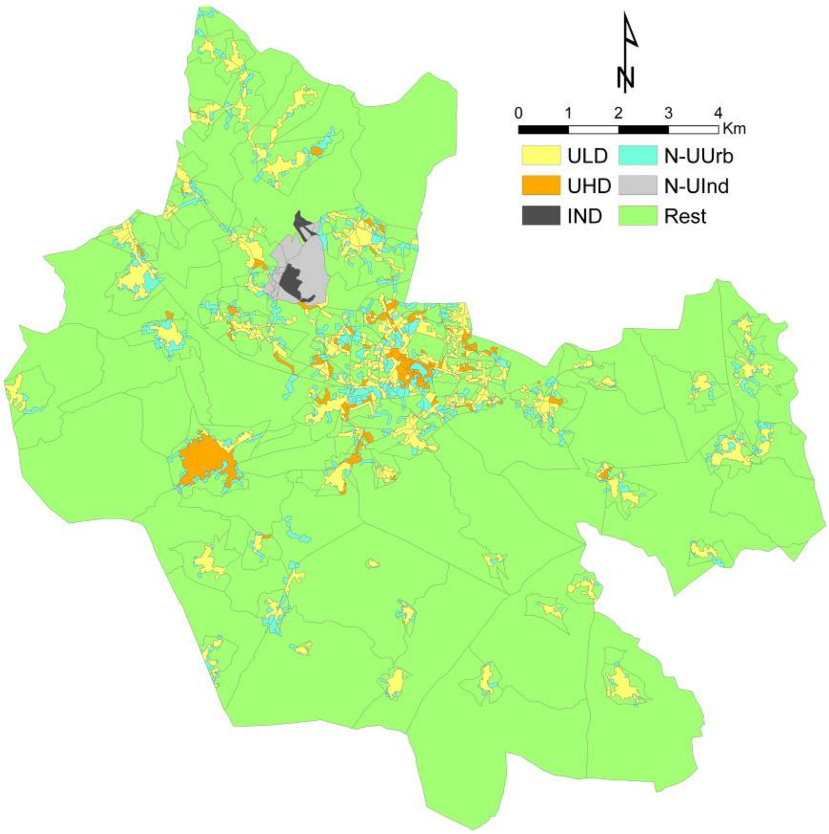


Figure 43 Simulation land use map in 2001

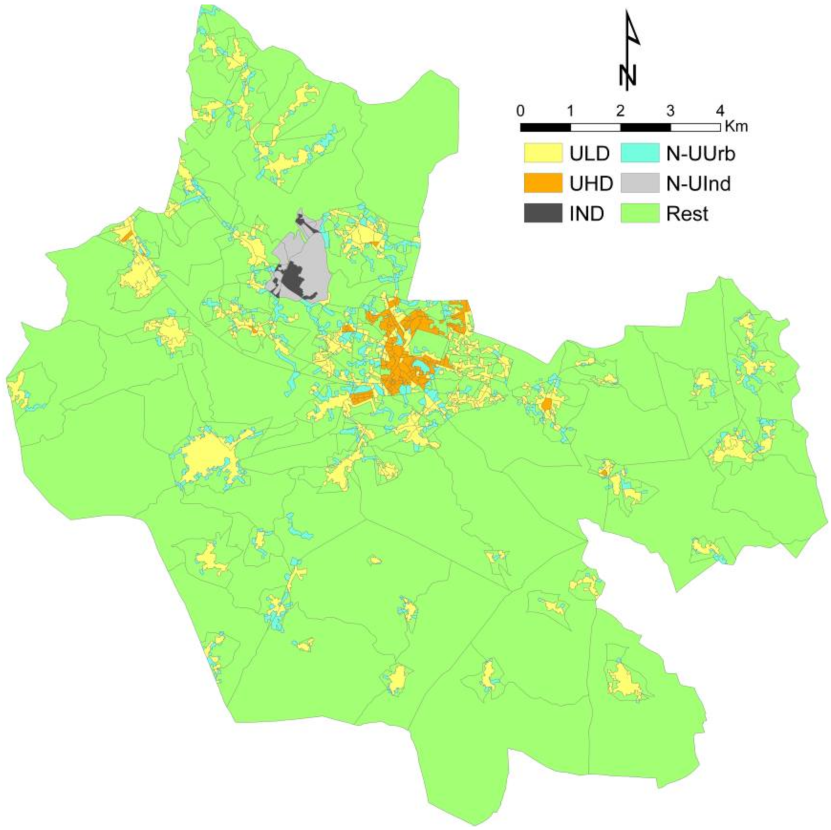


(a)

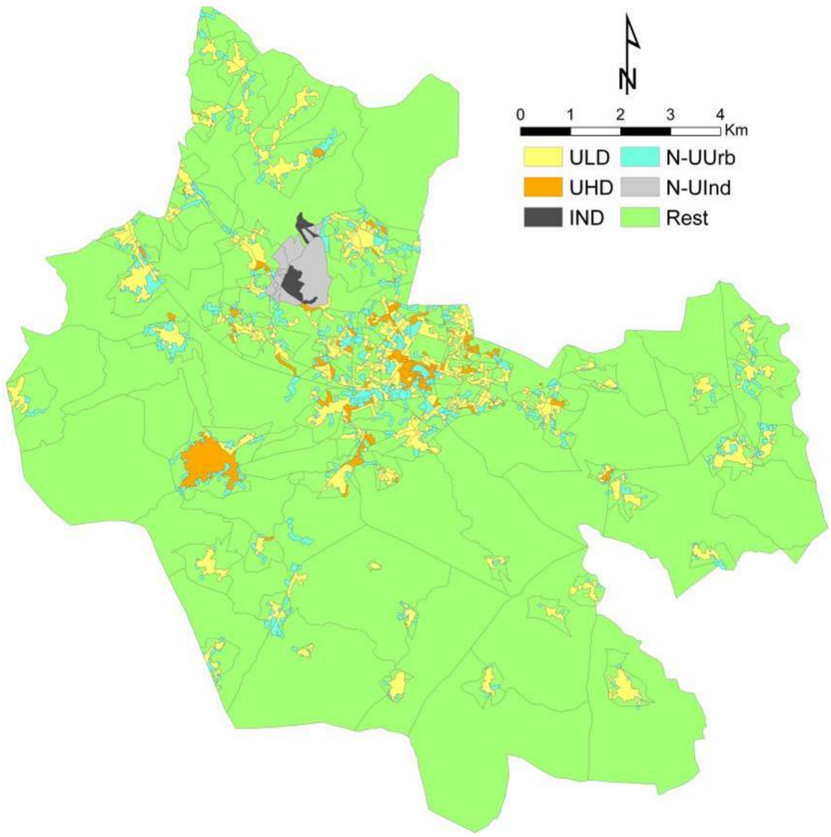


(b)

Figure 44 Evolution from reference land use map in 1991 (a) to simulation land use map in 2001 (b)



(a)



(b)

Figure 45 Comparison between reference land use map in 2001 (a) and simulation land use map in 2001 (b)

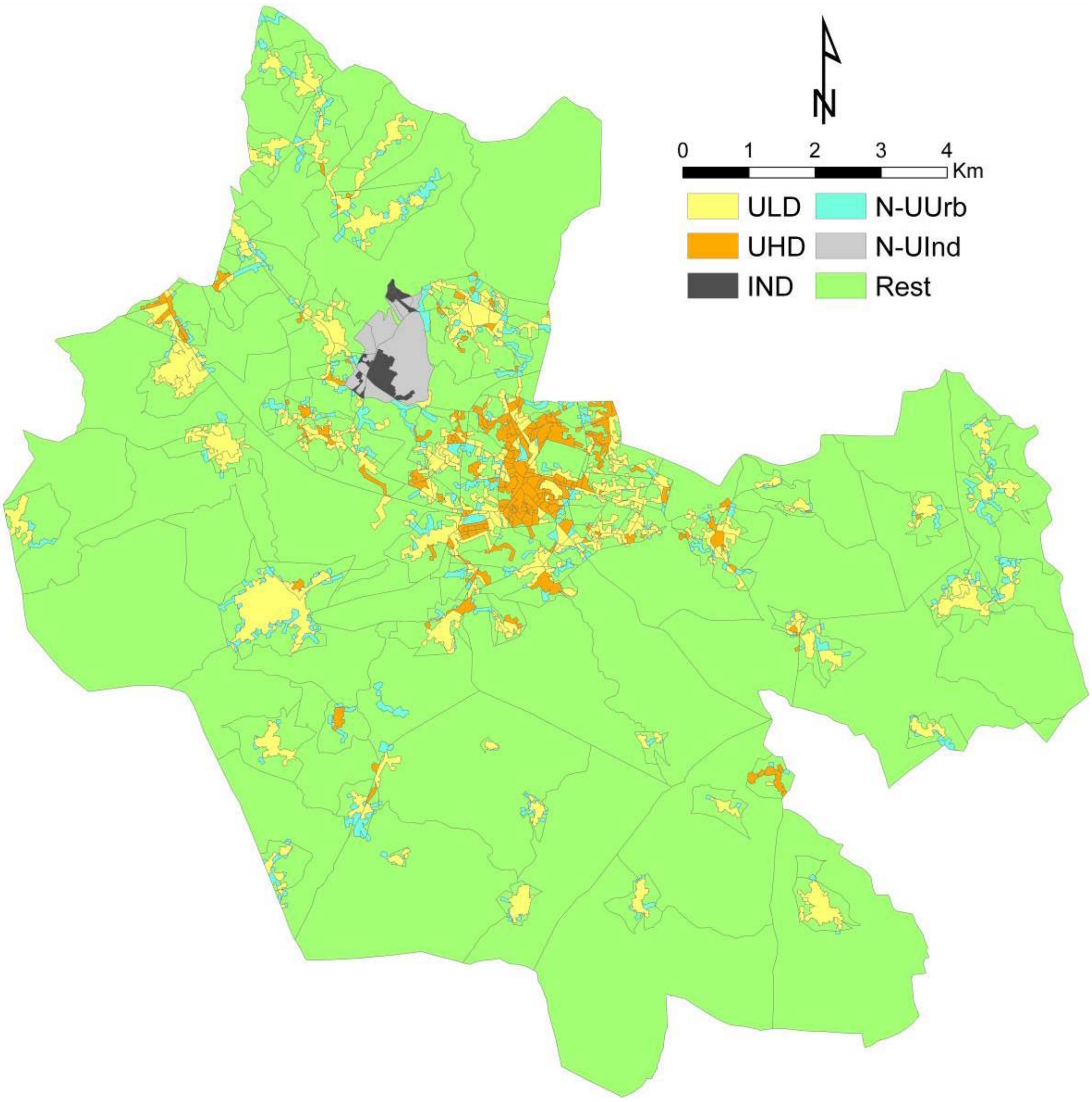


Figure 46 Simulation land use map for 2011

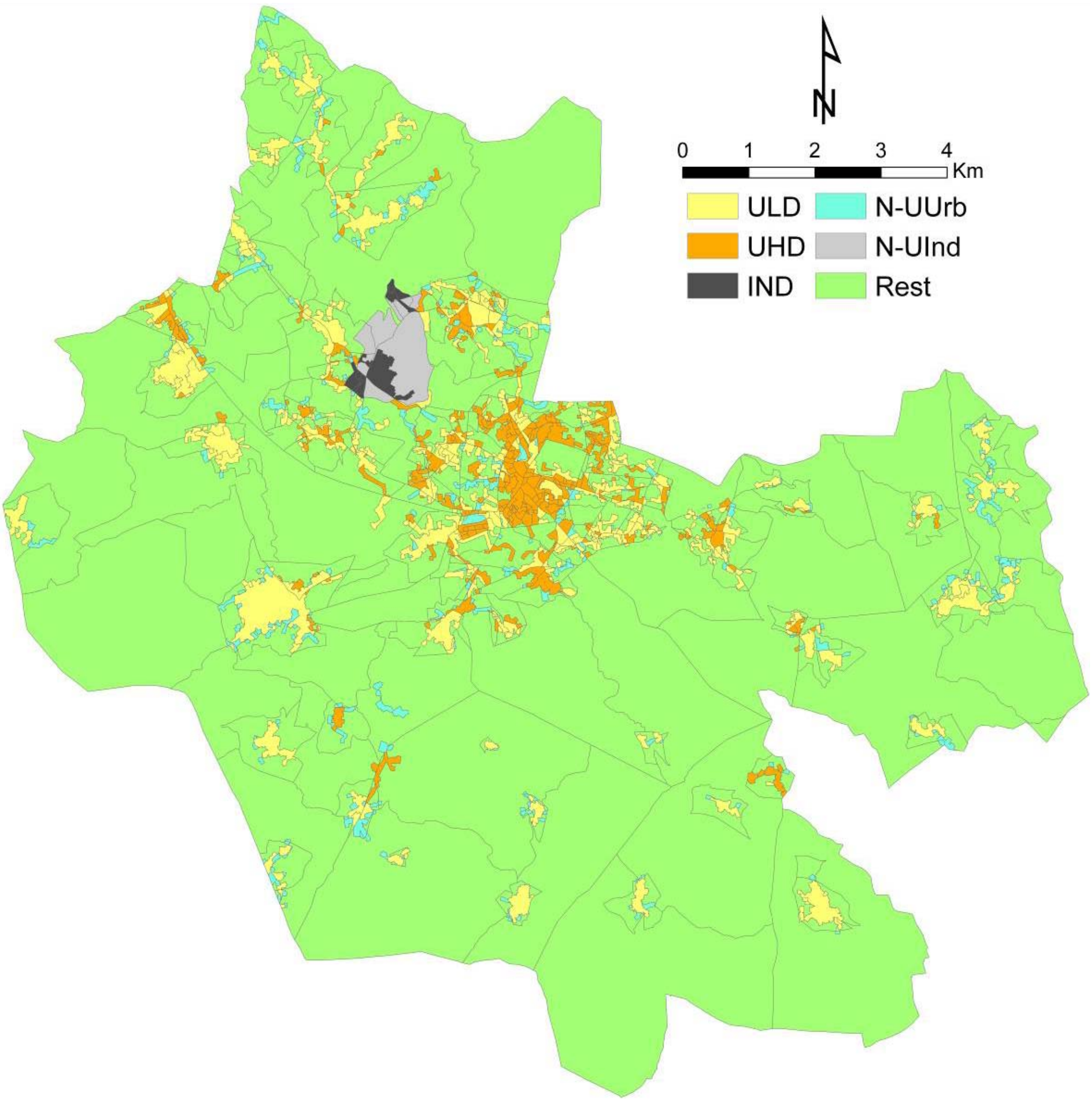
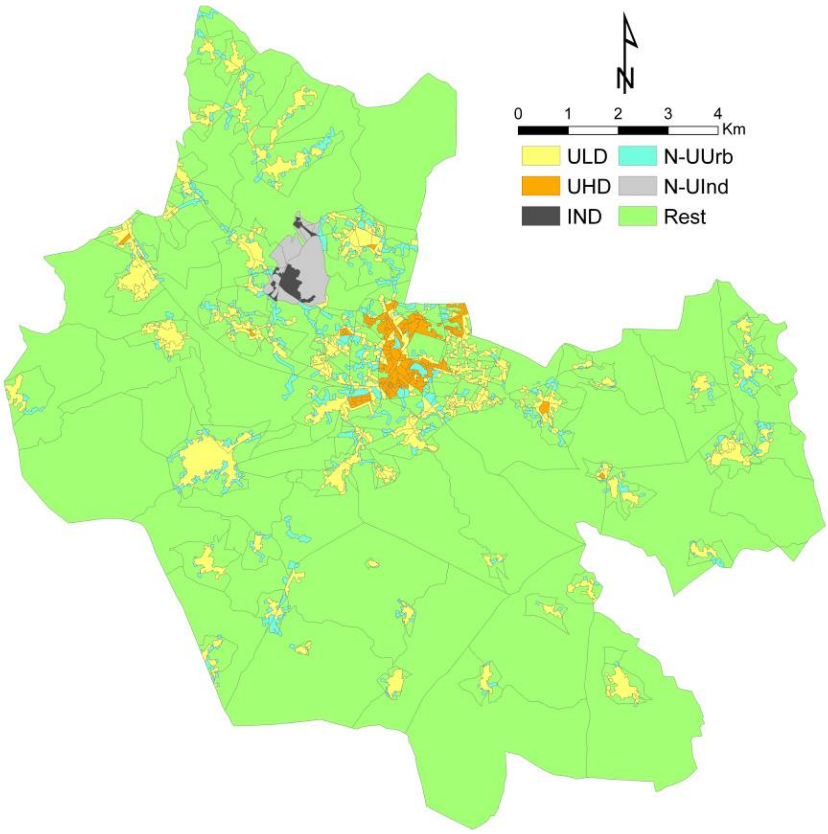
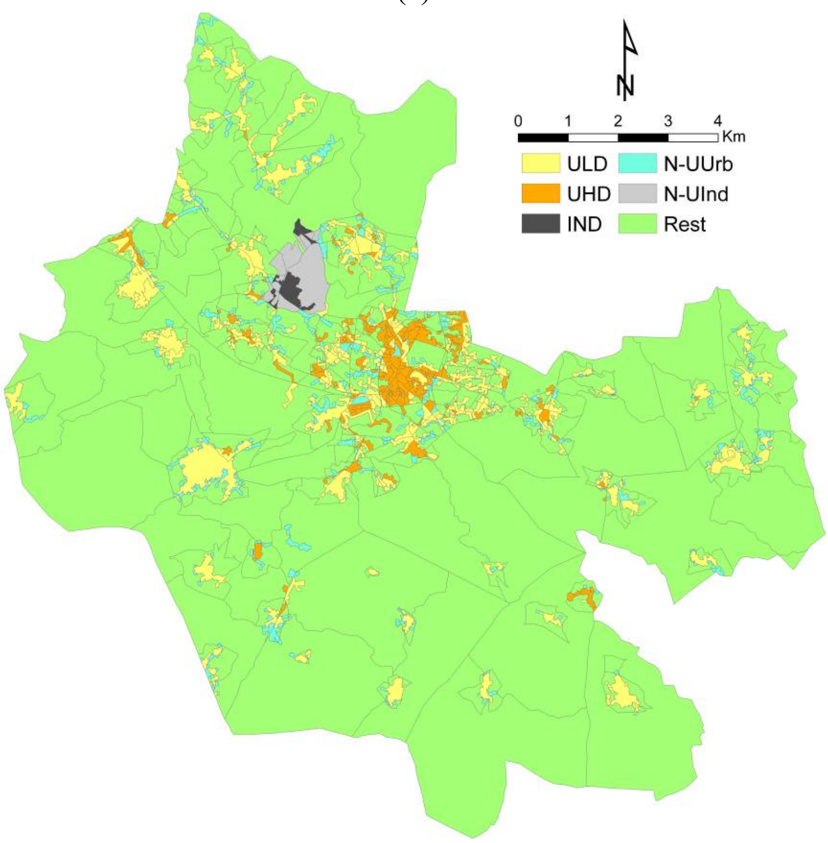


Figure 47 Simulation land use map for 2021

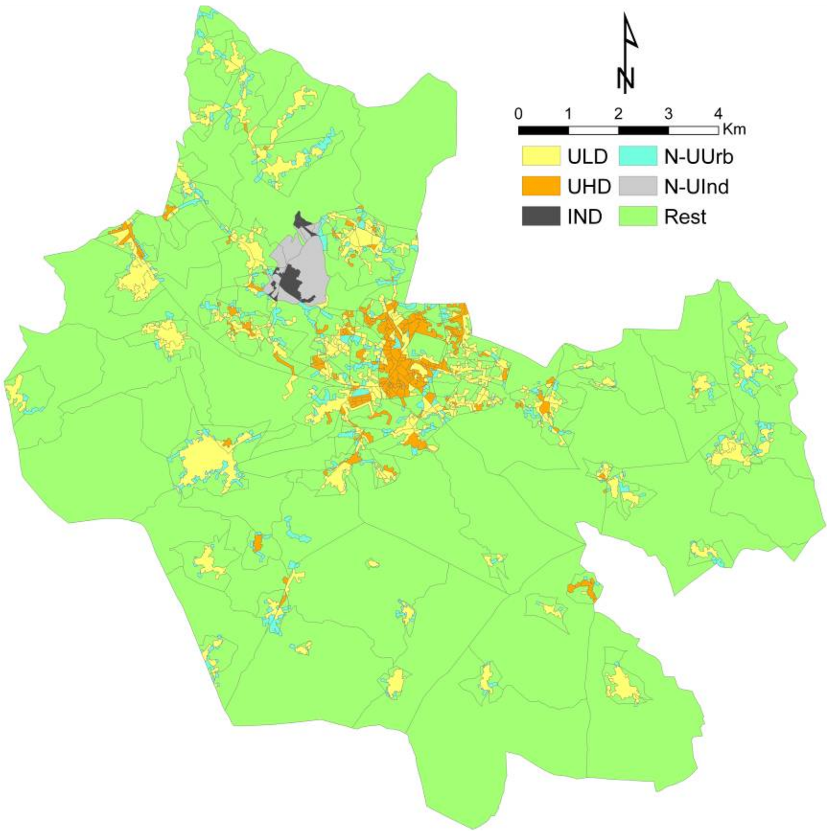


(a)

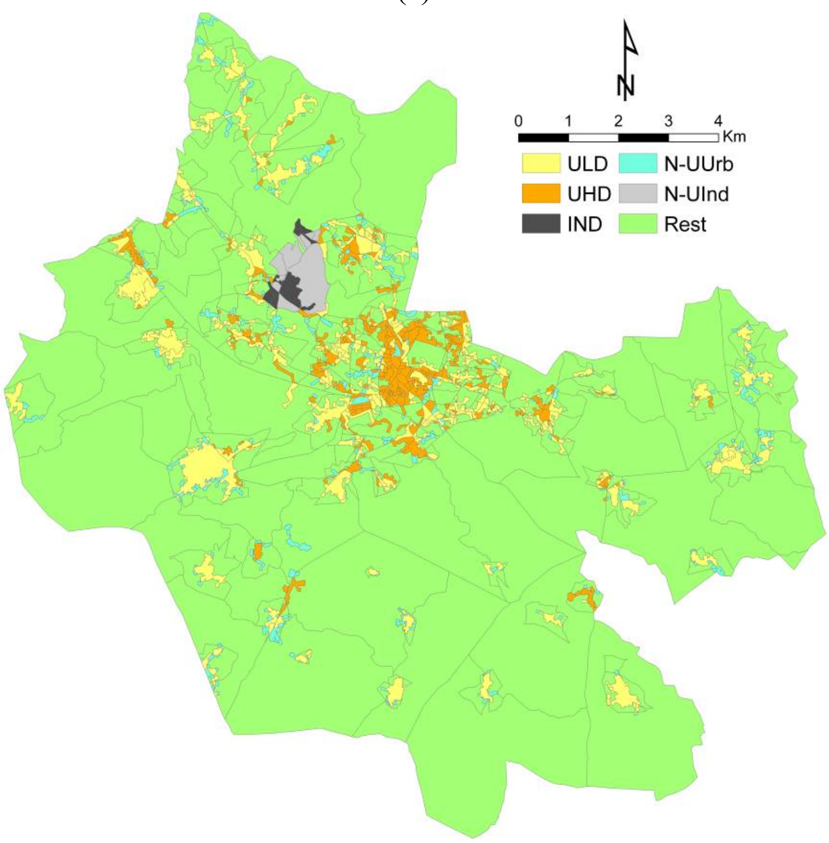


(b)

Figure 48 Evolution from reference land use map in 2001 (a) to simulation land use map in 2011 (b)



(a)



(b)

Figure 49 Evolution from simulation land use map in 2011 (a) to simulation land use map in 2021 (b)

6 Conclusion

"Is society reducible to a computer program?" Ballot and Weisbuch (2000) pose the question when they edited the proceedings of a meeting on simulation and social sciences. The question is also pertinent to urban studies. The answer for this question is quite obvious. Cities and other geographic features are complex by nature and their complexity tends to increase throughout time, because of the evolution of societies and to globalisation (that introduced a higher degree of uncertainty to urban phenomena). But, although planning activity is based on participation and planners tend to look at modelling as a hermetic field that has not been able to reproduce real urban phenomena so far, it is also obvious that it has an important role to play as a means for supporting the planning process.

In Chapter 2 an attempt was made to gather basic information both on modelling theory and on cellular automata (CA) theory and application to urban studies. Urban modelling is presented as a fundamental tool for promoting a scientific approach to the study of urban phenomena (Batty, 1994; Couclelis, 2005). The majority of the authors agree in presenting modelling as an important tool to assist planning, although they also point out its shortcomings. Many of these faults have been overcome in the last few years, with the development of computation science, mathematics and operations research. Planners started to consider modelling as a valid tool for scenario making and visioning (Couclelis, 2005).

Microsimulation is presented as the ultimate modelling approach because it deals with much more disaggregate information than other approaches (Benenson and Torrens, 2004; Waddell and Ulfarsson, 2004). CA and agent-based simulation have gained followers all around the world during the last decade and a series of theoretical and operational models based on these techniques are being developed and implemented.

CA are considered to be a powerful tool for dealing with complex spatial problems and there are several studies on this technique, considering both its theoretical issues (Couclelis, 1985; White and Engelen, 1993; Batty and Longley, 1994) and its application to real world problems (Clarke *et al.*, 1997; Ward, Murray and Phinn, 2000; Li and Yeh, 2001; Barredo *et al.*, 2003). This mathematical technique experienced an important evolution from its classical formulation (Wolfram, 1983, 1984) to a formalisation better suited to urban problems, as a results of successive evolutions and relaxations of its main components (Couclelis, 1997; Benenson and Torrens, 2004).

The application of different measures and calibration techniques is also under intensive research, with studies based on user-dependent procedures (Barredo *et al.*, 2003), on optimization techniques (Li and Yeh, 2001) and on mixed approaches (Clarke *et al.*, 1997).

6.1 Model results

The model was applied to a set of theoretical test problems and to a real world case study in order to understand its behaviour.

Given the results obtained for the set of test problems and for the case study of Condeixa-a-Nova, the use of CA models for simulating small urban areas is considered to be feasible. The use of irregular cells also ensured good results when compared with other CA models that use regular cells.

It is important to stress the innovative nature of this approach based on irregular cells. The CA model developed in this study differs from common applications of urban CA mostly by considering the use of irregular cells instead of traditional regular cells. The use of irregular cells based on census tracts is considered an important development of the approach. It shifts the focus of the model from land use – usually obtained from remote sensing image – to spatial information on both land use and demographics – census tracts contain disaggregate demographic and socio-economic information and can be easily classified in terms of land use. The simultaneous consideration of land use and demographic information is believed to be a major strength of the approach. It enhances spatial representativeness because cell space is closer to urban structure than traditional regular cell spaces.

The other innovative element is the consideration of land use demand as a function of observed and expected population densities. It can be said that urban growth depends on population growth and on the variation of construction density. Rather than assume that land use demand is assessed by a given number of cells[21], it is believed to be far more representative of urban change phenomena the consideration of population (and employment) growth as a tangible indicator for land demand.

The results obtained show the model's potential to simulate urban change for small urban areas. Despite some shortcomings, generally the model was able to achieve good agreement between reference and simulation maps. The high level of agreement obtained for the set of test problems – 50 percent of the problems achieved agreements of 0.800 of higher – opens promising possibilities for the application of this approach to real world small urban areas. Although these test problems are what it may be called "fake reality", they are useful for an exploratory understanding of model behaviour. The values obtained for the agreement are considerably high when compared with other studies that used similar measures.

[21] Traditional CA models that use regular cells generally based on remote sensing pixels can use the number of cells as land use demand because cell area is constant for the entire space.

The model did not achieve a strong correlation between its performance (measured by the agreement between simulation and reality) and problem size. As it can be seen in Figure 30 , when a correlation is made between performance and number of cells the correspondent Pearson value is only of -0.372. When a correlation is made between performance and the proportion of cells with active states (see Figure 31), the model produced a better correlation (Pearson value of -0.470) but still far from granting a solid relationship. Nevertheless, the trends obtained by these correlations indicate that the smaller the size of the problem is, the better the model can simulate it.

The model did not achieve such a great result for the case study, only reaching an agreement of 0.621. However, it may be also considered a promising result because, on the one hand, this lower value is in line with the trend deduced for the set of test problems (the larger the problem is, the worse the agreement is); on the other hand, although the agreement achieved a small value, it must be considered good because it is close to commonly accepted thresholds, for example for remote sensing agreements. This result has its causes in a series of shortcomings that were already identified.

The agreement produced by simulation was also assessed by the number of cells whose change was matched in simulation and in reality. For 80 percent of the test problems, matching proportions of 35 percent or higher were achieved; for 20 percent of the problems the value of 50 percent of matches was exceeded. These results can also be considered satisfactory, because from visual observation (see the problems fact sheets in Appendix: Test Problems Fact Sheets) it is clear that the model chose neighbouring cells of cells that really changed state. This may be a consequence of an insufficient differentiation of transition potentials for neighbouring cells. For the case study this proportion dropped to 20 percent of the cells. This behaviour has two main reasons. The first reason is related with problem definition. The cell lattice that is used by the model is obtained from the intersection of urban areas with census tracts. However, it is necessary to carefully define the final cell space. Urban cells located outside the main town centre had much larger areas than the ones located in central town (for example, see Figure 43). These large urban cells are well located both in terms of accessibility and neighbourhood, which tend to grant these cells higher transition potential values. For this reason the model tends to quickly satisfy land use demand by assigning large amounts of population to large cells with high potentials rather than to consider a series of smaller cells where small amounts of population can be placed. The second reason regards the variation of population density over the reference years. In 2001, population density for land use UHD (urban – high density) is only a little higher than population density for land use ULD (urban – low density) in 1991. Because of this small difference, the model can only allocate small amounts of population in small-size and mid-size cells in state ULD, favouring its appetite for large cells where large amounts of population can be placed. This behaviour was not clear for the set of test problems because in those cases cells presented less variance in their dimensions.

To overcome this problem it is imperative to divide large urban and industrial cells into a series of average-size cells in order to balance the supply of land.

Another important issue that needs to be carefully addressed regards the assessment of the value for the transition potential. This indicator of land use change probability has three components: accessibility, land suitability and neighbourhood effect. Calibration generated a set of calibration parameters for each one of these components. The model was able to choose cells close to the ones that had really changed state. This shows that the value for the transition potential should be very similar for these neighbouring cells. This can be overcome by using better accessibility and suitability measures.

Accessibility was measured by a very simple function of distance to functional centres and industrial areas measured over a very simple road network. The choice for this type of measure results from the fact that problems under consideration are simple in their nature and accessibility can be measured by aggregate values of distance to main centres. However, this measure has very tight limits because it may under estimate the distribution of employment all over the territory, enhancing the importance of industrial areas as the main employment centre. The calibration of the set of test problems showed a trend for a higher importance of the distance to the main functional centre (municipality main town). The distance to industrial area is the less important component of accessibility for these problems. However, for the case study the distance to the industrial area had more importance than the distance to the other centres.

The same shortcoming can be pointed out to suitability. In order to maintain model simplicity, suitability was reduced to a binary variable that also included zoning. This simplification can be override by considering a better measure of suitability. Zoning is highly conditioning to land use because it allows some land uses and forbids others. Land suitability was reduced to the legal definition of zoning: a land parcel can be occupied with a given land use if it is allowed by regulations. Although this assumption is close to planning practice, in order to better differentiate cells it may be necessary to distinguish land suitability from zoning. A cell can be suitable for a given land use while its neighbour is less suitable for the same land use because of different physical conditions (slope, hill shade).

The concept of neighbourhood may also be improved. Urban phenomena are far from having linear behaviours between locations or agents. The concept of cell neighbourhood is also very difficult to establish. Traditional neighbourhoods based on fixed distances between locations somehow excessively simplify reality. Real world neighbourhoods do not have regular shapes and the model concept of neighbourhood should be able to simulate the natural concept of neighbourhood. Real world neighbourhoods are influenced by the existence of physical or built barriers that must be taken into account. This is to say that some attention should be given to the study of neighbourhoods in order to consider natural neighbourhoods for modelling purposes.

6.2 Future developments

The study presented in this dissertation is the first stage of a broader study that aims to develop an integrated land use change model based on CA. This first stage had the goal of developing a CA model capable of simulating urban change in small urban areas. The study also had an implicit goal of exploring CA as a microsimulation technique. The results discussed in the previous section show promising possibilities of using CA for modelling urban change in small urban areas.

The next phase of the study is now starting and it will be developed under a doctoral programme at the Polytechnic University of Catalonia on to the same subject of microsimulation of urban change phenomena. The goal is to use the expertise on CA acquired in this study to develop a more ambitious integrated land use model, aimed to simulate urban change phenomena with a multi-scale approach. The final objective is to develop an innovative technology that could be applied to multi-scale urban studies after a process of validation based on its application to well studied urban areas.

The CA model developed in this study is oriented for the simulation of small urban areas. The model is essentially based on the assessment of a transition potential that is a function of several components, such as accessibility, land suitability and neighbourhood relationships. As it was discussed in the previous section, the assessment of the three components of transition potential must be improved in order to better simulate these phenomena. A multi-modal accessibility model will be developed to correctly assess local measures of accessibility. Land suitability will also be subjected to further research in order to establish suitable measures of physical characteristics. The concept of neighbourhood and the assessment of neighbourhood relationships is another field that needs careful research.

The use of local scale CA model is believed to produce good simulation results for urban change phenomena. The integration of CA with other models specifically designed for simulating accessibility, physical suitabilities and land use demand is believed to lead to the development of powerful modelling tools ultimately aimed to assist planning.

7 Bibliography

Allen, E. (2001). "INDEX: software for community indicators." Planning Support Systems: Integrating Geographical InformationSystems, Models and Visualization Tools. R. K. Brail and R. E. Klosterman. Redlands, CA, ESRI Press: 229-262.

Antunes, A. (1999). "Location analysis helps manage solid waste in Central Portugal." Interfaces 29(4): 32-43.

Antunes, A. and D. Peeters (2000). "A dynamic optimization model for school network planning." Socio-Economic Planning Sciences 34(2): 101-120.

Antunes, A., A. Seco and N. Pinto (2003). "An Accessibility Maximization Approach to Road Network Planning." Computer-Aided Civil & Infrastructure Engineering 18(3): 224-240.

Ballot, G. and G. Weisbuch, Eds. (2000). "Applications of Simulation to Social Sciences." Stanmore, Middlesex, Hermes Science Publishing.

Barredo, J., M. Kasanko, N. McCormick and C. Lavalle (2003). "Modelling dynamic spatial processes: simulation of urban future scenarios through cellular automata." Landscape and Urban Planning 64: 145-160.

Barros, J. (2005). "Urban dynamics in latin american cities: an agent-based simulation approach." Ninth Computers in Urban Planning and Urban Management, London, United Kingdom.

Batty, M. (1994). "A chronicle of scientific planning - The Anglo-American modelling experience." Journal of the American Planning Association 60(1): 7-16.

Batty, M. (2004). "Dissecting the streams of planning history: technology versus policy trough models." Environment and Planning B: Planning and Design 31: 326-330.

Batty, M., H. Couclelis and M. Eichen (1997). "Editorial: Urban systems as cellular automata." Environment and Planning B: Planning and Design 24: 159-164.

Batty, M. and P. Longley (1994). "Fractal cities : a geometry of form and function." London ; San Diego, Academic Press.

Batty, M. and P. Torrens (2001). "Modelling complexity: the limits to prediction." CASA Working Papers Series.

Batty, M. and Y. Xie (1997). "Possible urban automata." Environment and Planning B: Planning and Design 24: 175-192.

Batty, M., Y. Xie and Z. Sun (1999). "Modelling urban dynamics through GIS-based cellular automata." Computers, Environment and Urban Systems 23: 205-233.

Benenson, I., S. Aronovich and S. Noam (2005). "Let's talk objects: generic methodology for urban high-resolution simulation." Computers, Environment and Urban Systems 29: 425-453.

Benenson, I., I. Omer and E. Hatna (2002). "Entity-based modelling of urban residential dynamics: the case of Yaffo, Tel Aviv." Environment and Planning B: Planning and Design 29: 491-512.

Benenson, I. and P. M. Torrens (2004). "Geosimulation - Automata-based modeling of urban phenomena." Chchester, John Wiley & Sons Ltd.

Benguigui, L., D. Czamanski, M. Marinov and Y. Portugali (2000). "When and where is a city fractal?" Environment and Planning B: Planning and Design 27: 507-519.

Candau, J. (2000). "Calibrating a cellular automaton model of urban growth in a timely manner." 4th International Conference on Integrating GIS and Environmental Modeling (GIS/EM4), Banff, Alberta, Canada,.

Casti, J. (1997). "Would-be worlds: how simulation is changing the frontiers of science." New York, John Wiley and Sons.

Clarke, K. (2002). "Land USe Change Modeling Using SLEUTH." Advanced Training Workshop on Land Use and Land Cover Change Study, Taiwan, National Central University/National Taiwan Unniversity/ START.

Clarke, K., S. Hoppen and L. Gaydos (1996). "Methods and Techniques for Rigourous Calibration of a Cellular Automaton Model of Urban Growth." Third International Conference/Workshop on Integrating Geographic Information Systems and Environmental Modeling, Santa Fe, NM.

Clarke, K., S. Hoppen and L. Gaydos (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.

Costa, A. and R. N. Markellos (1997). "Evaluating public transport efficiency with neural network models." Transportation Research Part C: Emerging Technologies 5(5): 301-312.

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

Couclelis, H. (1997). "From cellular automata to urban models: new principles for model development and implementation." Environment and Planning B: Planning and Design 24: 165-174.

Couclelis, H. (2005). ""Where has the future gone?" Rethinking the role of integrated lad-use models in spatial planning." Environment and Planning A 37: 1353-1371.

Couto, P. (2003). "Assessing the accuracy of spatial simulation models." Ecological Modelling 167: 181–198.

De Keersmaecker, M.-L., P. Frankhauser and I. Thomas (2003). "Using fractal dimensions for characterizing intra-urban diversity: the example of Brussels." Geographical Analysis 35(4): 310-328.

de la Barra, T. (2001). "Integrated land use and transport modelling: the TRANUS experience." Planning Support Systems: Integrating Geographical InformationSystems, Models and Visualization Tools. R. K. Brail and R. E. Klosterman. Redlands, CA, ESRI Press: 129-156.

Eloranta, K. (1997). "Critical growth phenomena in cellular automata." Physica D: Nonlinear Phenomena Lattice Dynamics 103(1-4): 478-484.

Engelen, G., R. White, I. Uljee and P. Drazan (2002). "The MURBANDY and MOLAND models for Dublin." Ispra, European Comission Joint Reseach Centre: 172.

Fragkias, M. and K. Seto (2005). "Modelling urban growth in data-sparse environments: a new approach." XII CUPUM, London, UK.

Frankhauser, P. (1991). "Aspects fractals des structures urbaines." L'Espace Géographique 1: 45-69.

Frankhauser, P. and R. Sadler (1991). "Fractal analysis of agglomerations." Second International Colloquium of the Sonderforshungsbereich 230: Natural Structures Sonderforshungsbereich 230, Stuttgard.

Friedman, B., P. Kahn and A. Borning (2002). "Value Sensitive Design: Theory and Methods." Seattle, University of Washington: 8.

Gerling, R. W. (1990). "Classification of triangular and honeycomb cellular automata." Physica A: Statistical and Theoretical Physics 162(2): 196-209.

Herold, M., H. Couclelis and K. C. Clarke (2005). "The role of spatial metrics in the analysis and modelling of urban land use change." Computers, Environment and Urban Systems 29: 369-399.

Hunt, J. D. and D. C. Simmonds (1993). "Theory and application of an integrated land-use and transport modelling framework." Environment and Planning 20B: 221–244.

INE (2001). "Censos 2001." On-line at www.ine.pt Accessed in n/a.

Julião, R. P. (2001). "Tecnologias de Informação Geográfica e Ciência Regional - Contributos metodológicos para a definição de modelos de apoio à decisão em desenvolvimento regional." PhD Thesis, Faculty of Human and Social Sciences, New University of Lisbon, Lisbon

Kennedy, J. (1997). "The particle swarm: Social adaptation of knowledge." 303-308.

Klosterman, R. (1994). "Large-Scale urban models - Restrospect and prospect." Journal of the American Planning Association 60(1): 3-6.

Klosterman, R. and C. Petit (2005). "Editorial: An update on planning support systems." Environment and Planning B: Planning and Design 32: 477-484.

Klosterman, R. E. (2001). "The What If? planning support system." Planning Support Systems: Integrating Geographical InformationSystems, Models and Visualization Tools. R. K. Brail and R. E. Klosterman. Redlands, CA, ESRI Press: 263-284.

Klosterman, R. E., L. Siebert, M. A. Hoque, J. W. Kim and A. Parveen (2003). "Using an operational planning support system to evaluate farmland preservation policies." Planning Support Systems in Practice. S. Geertman and J. Stillwell: 391-407.

Kwartler, M. and R. N. Bernar (2001). "CommunityViz: an integrated planning support system." Planning Support Systems: Integrating Geographical InformationSystems, Models and Visualization Tools. R. K. Brail and R. E. Klosterman. Redlands, CA, ESRI Press: 285-308.

Landis, J. (2001). "CUF, CUF II and CURBA: a family of spatially explicit urban growth and land-use policy simulation models." Planning Support Systems: Integrating Geographical InformationSystems, Models and Visualization Tools. R. K. Brail and R. E. Klosterman. Redlands, CA, ESRI Press: 157-200.

Latuso, K. (2003). "The SPARTACUS system for defining and analysing sustainable land use and trasnport policies." Planning Support Systems in Practice. S. Geertman and J. Stillwell. Heidelberg, Springer: 453-463.

Lee, D. (1973). "Requiem for large-scale models." Journal of the American Planning Association 39(3): 163-178.

Lee, D. (1994). "Retrospective on large-scale urban models." Journal of the American Planning Association 60(1): 35-40.

Leontief, W. W. (1966). "Input-output economics." New York,, Oxford University Press.

Levy, S. (1992). "Artificial life : the quest for a new creation." New York, Pantheon Books.

Li, X. and A. Yeh (2001). "Calibration of cellular automata by using neural networks for the simulation of complex urban systems." Environment and Planning A 33: 1445-1462.

Lindenmayer, A. (1968). "Mathematical models for cellular interaction in development, Parts I and II." Journal of Theoretical Biology 18: 280-315.

Longley, P. A. and V. Mesev (2000). "On the measurement and generalisation of urban form." Environment and Planning A 32: 473-488.

Mandelbrot, B. (1983). "The Fractal Geometry of Nature." New York, W. H. Freeman.

Marques, T. S. (2004). "Portugal na Transição do Século - Retratos e Dinâmicas Territoriais." Porto, Edições Afrontamento.

McLoughlin, J. B. (1969). "Urban and Regional Planning." London, Faber and Faber.

Ménard, A. and D. J. Marceau (2005). "Exploration of spatial scale sensitivity in geographical cellular automata." Environment and Planning B: Planning and Design 32: 693-714.

Miller, E., J. Hunt, J. Abraham and P. Salvini (2004). "Microsimulating urban systems." Computers, Environment and Urban Systems 28: 9-44.

O'Sullivan, D. (2001a). "Exploring Spatial Process Dynamics Using Irregular Cellular Automaton Models." Geographical Analysis 33(1): 1-18.

O'Sullivan, D. (2001b). "Graph-based cellular automaton models of urban spatial processes." PhD Thesis, Bartlett School of Architecture and Planning, University College, London

Parsopoulos, K. E. and M. N. Vrahatis (2002). "Recent approaches to global optimization problems through Particle Swarm Optimization." Natural Computing 1: 235–306.

Putman, S. H. and C. Shih-Liang (2001). "The METROPILUS planning support system: urban models and GIS." Planning Support Systems: Integrating Geographical InformationSystems, Models and Visualization Tools. R. K. Brail and R. E. Klosterman. Redlands, CA, ESRI Press: 99-128.

Semboloni, F. (1997). "An urban and regional model based on cellular automata." Environment and Planning B-Planning & Design 24(4): 589-612.

Semboloni, F. (2000). "The growth of an urban cluster into a dynamic self-modifying spatial pattern." Environment and Planning B-Planning & Design 27(4): 549-564.

Semboloni, F. (2005). "Multi-agents simulation of urban dynamic." XII CUPUM, London.

Silva, E. (2002). "Cenários da Expansão Urbana na Área Metropolitana de Lisboa." Revista de Estudos Regionais 2º Semestre 2002: 23-41.

Silva, E. and K. C. Clarke (2002). "Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal." Computers, Environment and Urban Systems 26: 525-552.

Snyder, K. (2001). "Tools for community design and decision-making." Planning Support Systems in Practice. S. Geertman and J. Stillwell. Hedelberg, Springer: 99-120.

Straatman, B., R. White and G. Engelen (2004). "Towards an automatic calibration procedure for constrained cellular automata." Computers, Environment and Urban Systems 28(1-2): 149-170.

Swarm (2002). "Swarm - Multi-Agent Simulation Environment." On-line at http://www.swarm.org/wiki/Main_Page Accessed in 2006/09/07.

Takeyama, M. and H. Couclelis (1997). "Map dynamics: integrating cellular automata and GIS through Geo-Algebra." International Journal of Geographical Information Science, 11(1): 73-91.

Teixeira, J., A. P. Antunes and J. P. de Sousa (2004). "Recyclable waste collection planning--a case study." European Journal of Operational Research 158(3): 543-554.

Tobler, W. (1970). "A Computer Movie Simulating Urban Growth in the Detroit Region." Economic Geography 46(2): 234-240.

Tobler, W. (1979). "Cellular geography." Philosophy in Geography. S. Gale and G. Olsson. Boston, D. Reidel: 379-386.

Torrens, P. (2000). "How cellular models of urban systems work (1. Theory)." On-line at http://www.casa.ucl.ac.uk/working_papers.htm Accessed in

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

Trelea, I. C. (2003). "The particle swarm optimization algorithm: convergence analysis and parameter selection." Information Processing Letters 85: 317–325.

van den Bergh, F. and A. P. Engelbrecht (2005). "A study of particle swarm optimization particle trajectories." Information Sciences in press(in press).

Vandergue, D., J.-P. Treuil and D. Drogoul (2000). "Modelling urban phenomena with cellular automata." Applications of Simulation to Social Science, Paris, Hermes Science Publishing.

Waddell, P. (2001). "Between politics and planning:UrbanSim as a decision-support system for metropolitan planning." Planning Support Systems: Integrating Geographical InformationSystems, Models and Visualization Tools. R. K. Brail and R. E. Klosterman. Redlands, CA, ESRI Press: 201-228.

Waddell, P., C. Bhat, E. Ruiter, S. Bekhor, M. Outwater and E. L. Schroer (2001). "Review of the Literature and Operational Models." Puget Sound Region, University of Washington.

Waddell, P. and G. Ulfarsson (2004). "Introduction to urban simulation: design and development of operational models." Handbook of Transport Geography and Spatial Systems, Volume 5 (Handbooks in Transport). D. Hensher, K. Button, K. Haynes and P. Stopher, Elsevier Science.

Wagner, D. F. (1997). "Cellular automata and geographic information systems." Environment and Planning B: Planning and Design 24: 219-234.

Ward, D., A. Murray and S. Phinn (2000). "A stochastically constrained cellular model of urban growth." Computers, Environment and Urban Systems 24: 539-558.

Ward, D., A. Murray and S. Phinn (2003). "Integrating spatial optimization and cellular automata for evaluating urban change." The Annals of Regional Science 37: 131-148.

Wegener, M. (1994). "Operational urban models - State of the Art." Journal of the American Planning Association 60(1): 17-29.

White, R. and G. Engelen (1993). "Cellular Automata and fractal urban form: a cellular modelling approach to the evolutionb of urban land-use patterns." Environment and Planning A 25: 1175-1199.

White, R. and G. Engelen (1997). "Cellular automata as the basis of integrated dynamic regional modelling." Environment and Planning B: Planning and Design 24: 235-246.

White, R. and G. Engelen (2000). "High-resolution integrated modelling of the spatial dynamics of urban and regional systems." Computers, Environment and Urban Systems 24: 383-400.

Wolfram, S. (1983). "Statistical mechanics of cellular automata." Reviews of Modern Physics 55(3): 601-644.

Wolfram, S. (1984). "Computation theory of cellular automata." Communications in Mathematical Physics 96(1): 15-57.

Wolfram, S. (2005). "Historical notes: history of cellular automata." On-line at http://www.wolframscience.com/reference/notes/876b Accessed in 2005-12-12.

Wu, F. (2002). "Calibration of stochastic cellular automata: the application to rural-urban land conversions." Internationa l Journal of Geographical Information Science 16(8): 795–818.

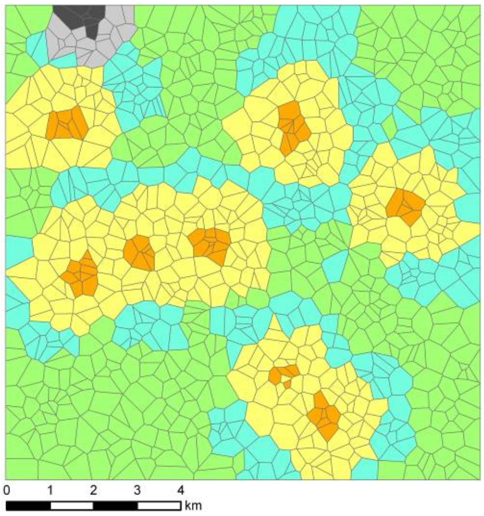
Xie, Y. (1996). "A generalized model for cellular urban dynamics." Geographical Analysis 284: 350-373.

Zeigler, B. P. (1976). "Theory of Modeling and Simulation." New York, Wiley.

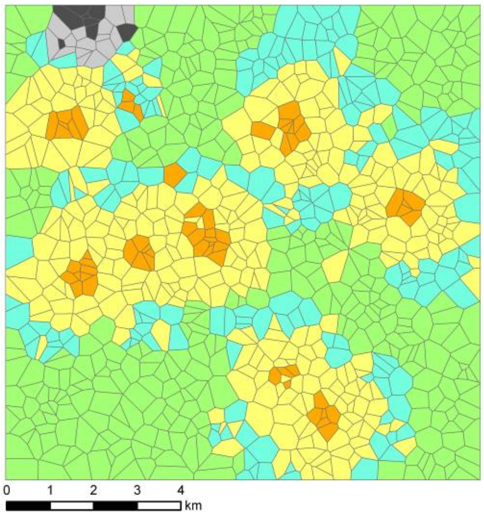
8 Appendix: Test Problems Fact Sheets

Problem#01

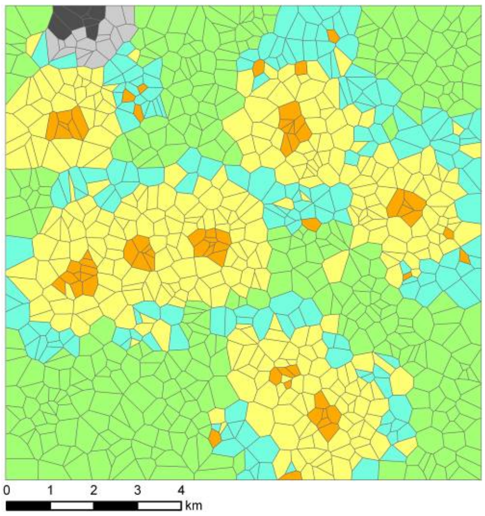
Initial Reference



Final Reference



Simulation



Problem Characteristics

Number of cells	Maximum Dimension (m)	Avgerage Cell Area (ha)	Area (ha)
933	10681	12.228	11408

	ULD	UHD	IND	N-UUrb	N-UInd	Rest
Population Density by State 1991 (inh/ha)	12.97	21.89	10.00	1.00		
Population Density by State 2001 (inh/ha)	7.19	13.36	10.00	1.00		
Total Area by State 1991 (ha)	3528	396	48	2790	170	4478
Total Area by State 2001 (ha)	4010	483	64	2220	153	4478
Area variation 1991/2001	14%	22%	34%	-20%	-10%	0%

Simulation Results

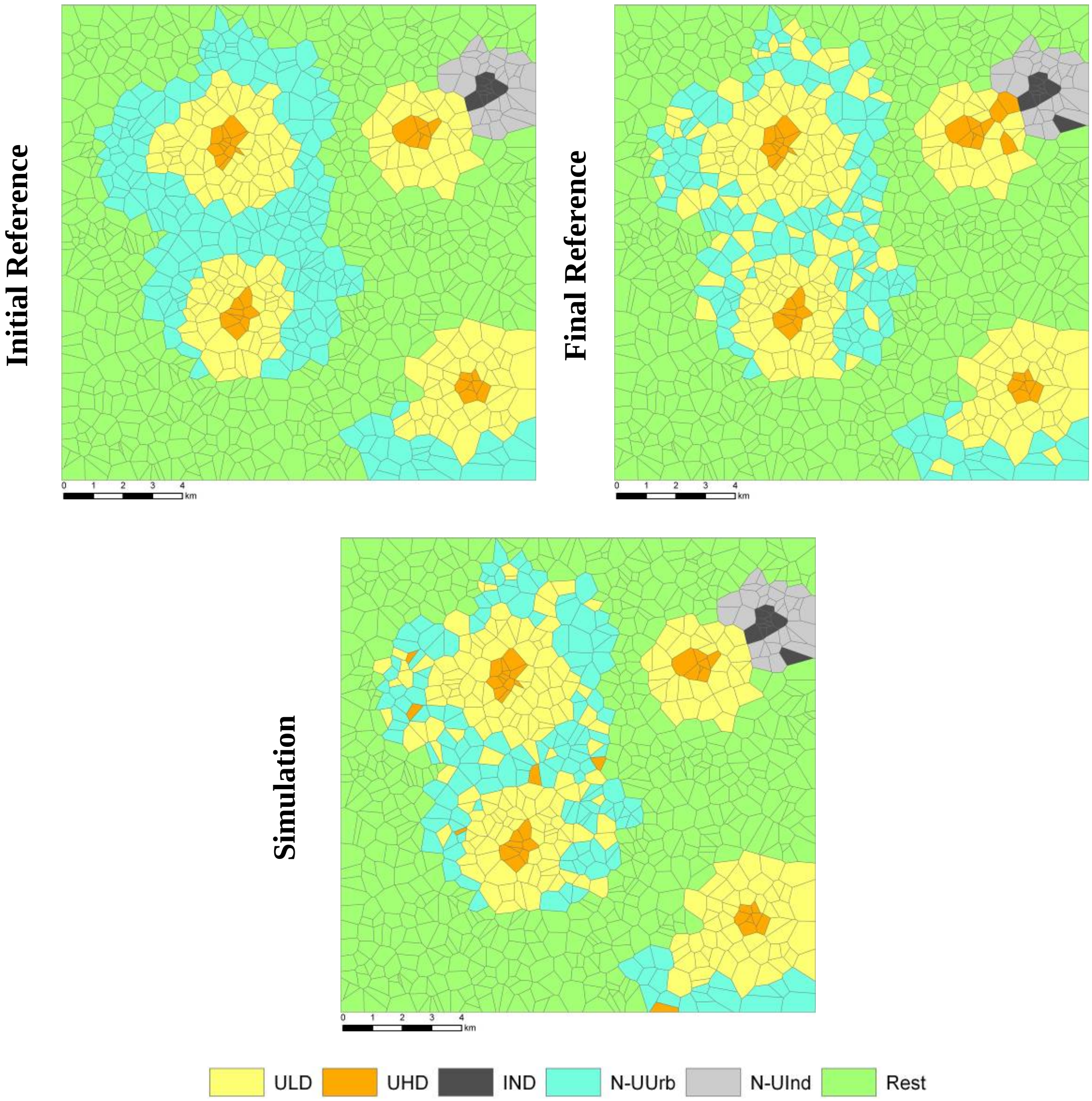
ModKvalue	*ModkValue* Increase	*kValue*	Overall Accuracy	Number of active cells	Proportion of active cells
0.827	24%	0.907	0.897	594	64%

Number of changed cells	Number of matching cells	Similarity (%)
56	25	45%

	ULD	UHD	IND	N-UUrb	N-UInd
Θ_S (1991/2001)	0%	-2%	11%	1%	-5%

Neigbourhood Distance (km)	α_{acc}	β_{acc}	γ_{acc}	χ_{pot}	ν_{pot}	θ_{pot}
3.0	0.618	0.429	0.154	0.739	0.536	0.377

Problem#02



Problem Characteristics

Number of cells	Maximum Dimension (m)	Avgerage Cell Area (ha)	Area (ha)
1033	15703	23.871	24658

Population Density by State 1991 (inh/ha)	ULD	UHD	IND	N-UUrb		
	11.69	23.80	10.00	1.00		
Population Density by State 2001 (inh/ha)	ULD	UHD	IND	N-UUrb		
	7.23	11.95	10.00	1.00		
Total Area by State 1991 (ha)	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	4902	474	107	5339	633	13203
Total Area by State 2001 (ha)	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	6194	568	141	3954	598	13203
Area variation 1991/2001	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	26%	20%	32%	-26%	-5%	0%

Simulation Results

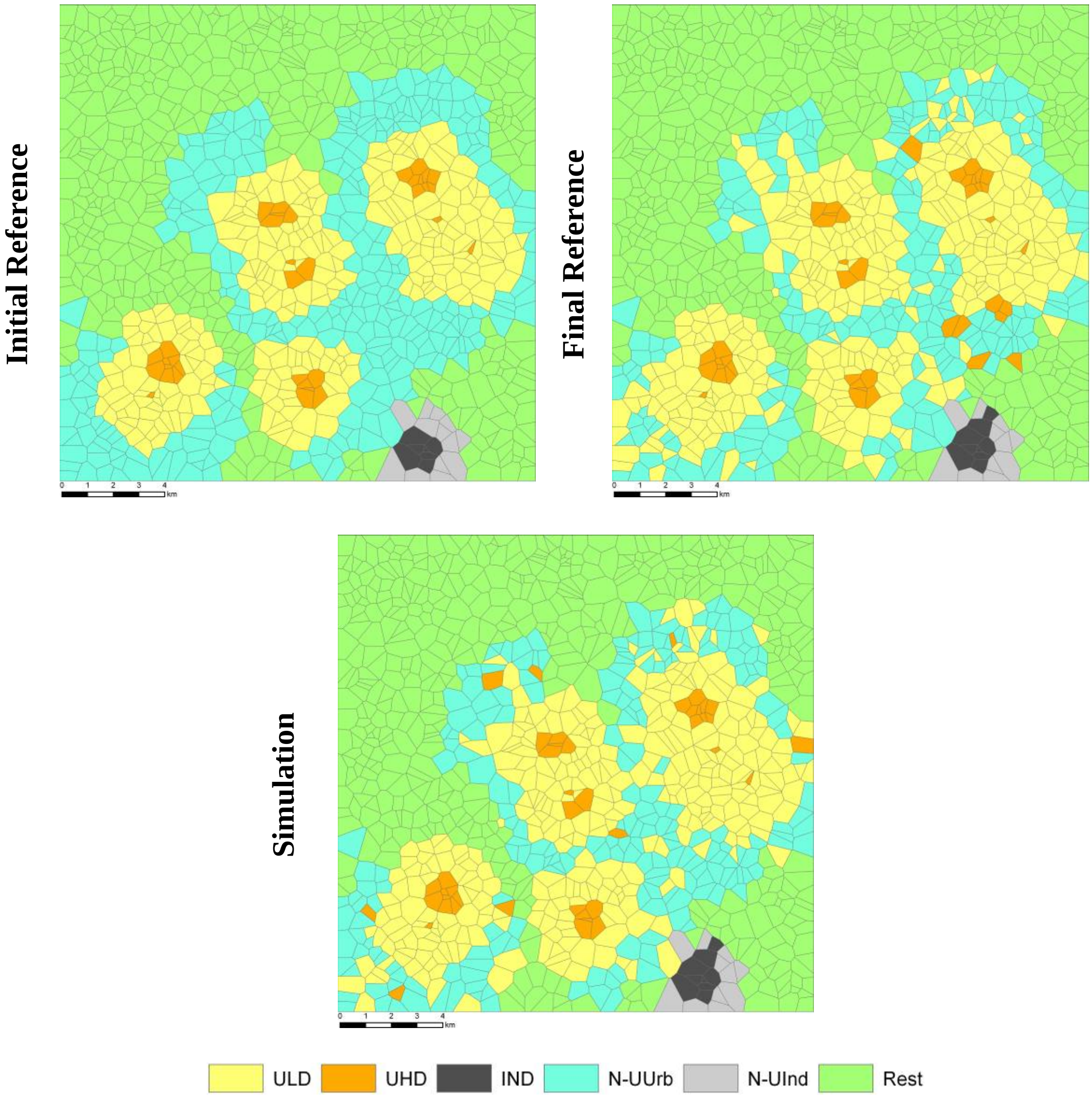
ModKvalue	*ModkValue* Increase	*kValue*	Overall Accuracy	Number of active cells	Proportion of active cells
0.770	24%	0.896	0.856	465	45%

Number of changed cells	Number of matching cells	Similarity (%)
57	24	42%

Θ_S (1991/2001)	ULD	UHD	IND	N-UUrb	N-UInd
	0%	-4%	0%	0%	0%

Neigbourhood Distance (km)	α_{acc}	β_{acc}	γ_{acc}	χ_{pot}	v_{pot}	θ_{pot}
3.7	0.580	0.545	0.277	0.396	0.596	0.150

Problem#03



Problem Characteristics

Number of cells	Maximum Dimension (m)	Avgerage Cell Area (ha)	Area (ha)
1082	18214	30.661	33175

	ULD	UHD	IND	N-UUrb
Population Density by State 1991 (inh/ha)	13.54	29.75	10.00	1.00
Population Density by State 2001 (inh/ha)	6.15	11.24	10.00	1.00

	ULD	UHD	IND	N-UUrb	N-UInd	Rest
Total Area by State 1991 (ha)	7972	693	228	9474	500	14308
Total Area by State 2001 (ha)	10097	953	299	7090	429	14308
Area variation 1991/2001	27%	37%	31%	-25%	-14%	0%

Simulation Results

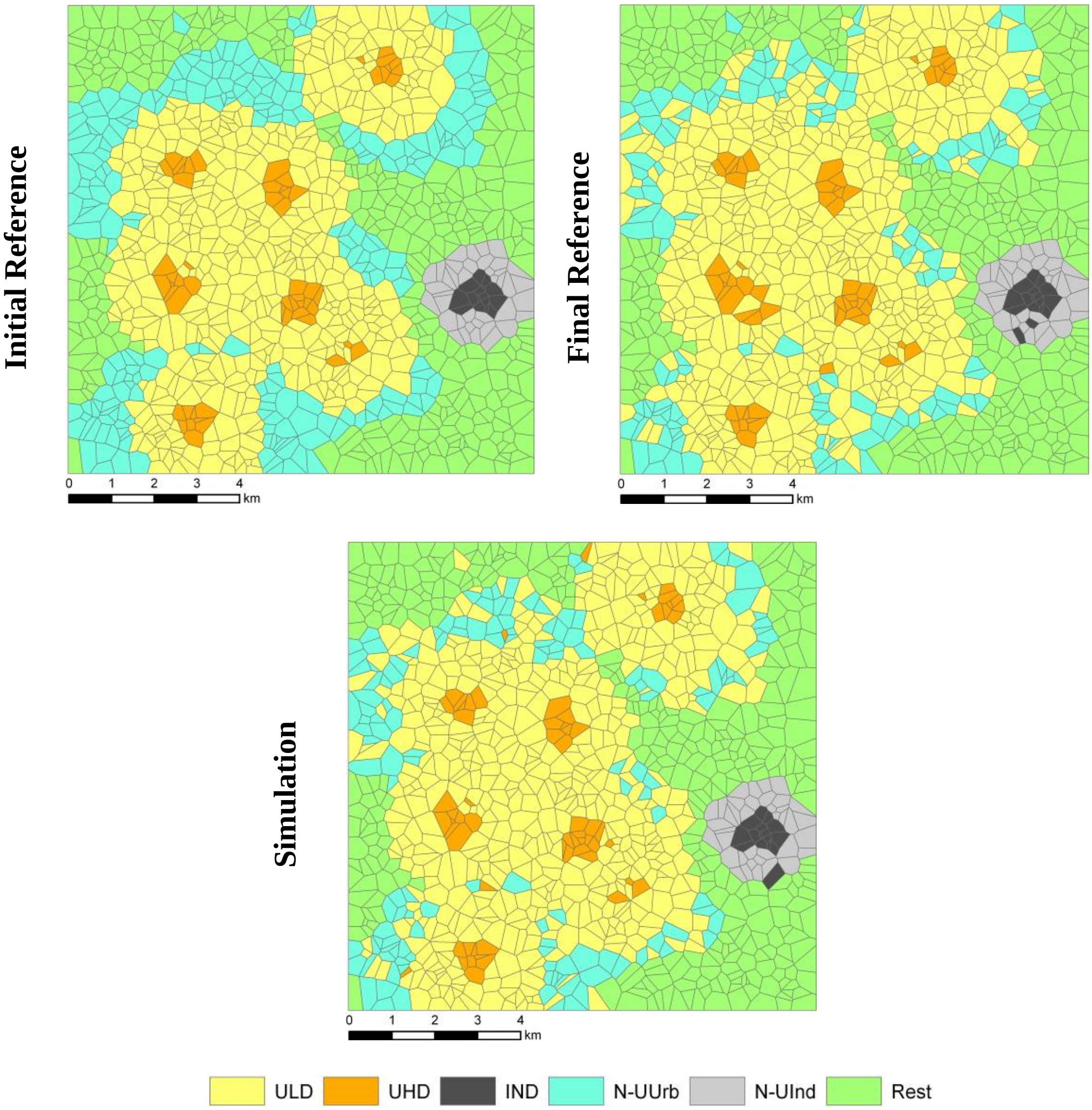
ModKvalue	*ModkValue* Increase	*kValue*	Overall Accuracy	Number of active cells	Proportion of active cells
0.726	22%	0.868	0.842	609	56%

Number of changed cells	Number of matching cells	Similarity (%)
73	24	33%

	ULD	UHD	IND	N-UUrb	N-UInd
Θ_S (1991/2001)	1%	-6%	0%	-1%	0%

Neigbourhood Distance (km)	α_{acc}	β_{acc}	γ_{acc}	χ_{pot}	v_{pot}	θ_{pot}
3.0	0.710	0.207	0.480	0.929	0.490	0.439

Problem#04



Problem Characteristics

Number of cells	Maximum Dimension (m)	Avgerage Cell Area (ha)	Area (ha)
1132	10725	10.161	11503

Population Density by State 1991 (inh/ha)	ULD	UHD	IND	N-UUrb
	10.40	25.71	10.00	1.00

Population Density by State 2001 (inh/ha)	ULD	UHD	IND	N-UUrb
	7.27	14.93	10.00	1.00

Total Area by State 1991 (ha)	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	4701	443	92	2468	378	3421

Total Area by State 2001 (ha)	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	5640	491	103	1480	367	3421

Area variation 1991/2001	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	20%	11%	12%	-40%	-3%	0%

Simulation Results

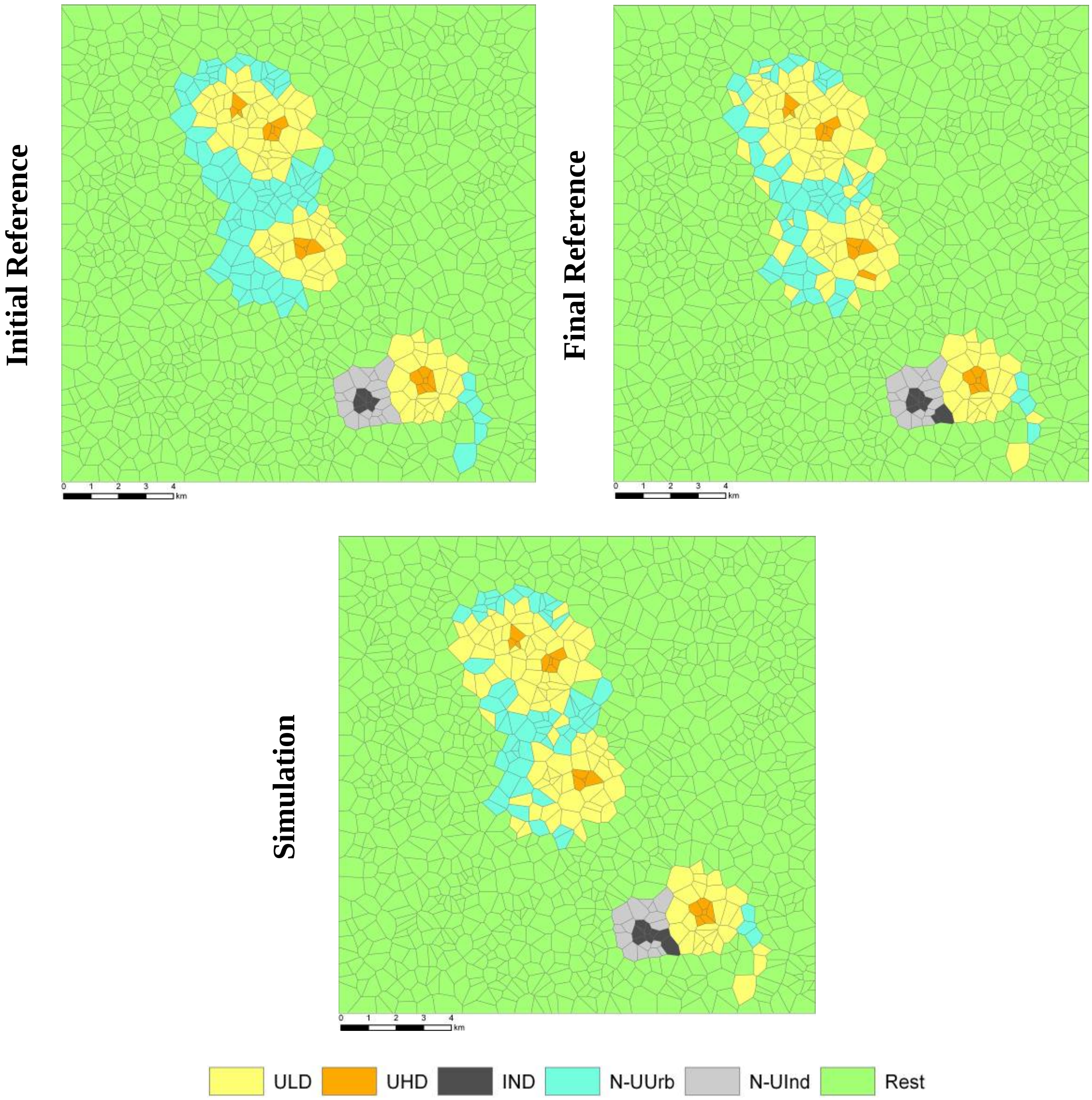
ModKvalue	*ModkValue* Increase	*kValue*	Overall Accuracy	Number of active cells	Proportion of active cells
0.770	20%	0.878	0.884	795	70%

Number of changed cells	Number of matching cells	Similarity (%)
93	50	54%

Θ_S (1991/2001)	ULD	UHD	IND	N-UUrb	N-UInd
	1%	-6%	2%	-2%	-1%

Neigbourhood Distance (km)	α_{acc}	β_{acc}	γ_{acc}	χ_{pot}	v_{pot}	θ_{pot}
3.7	0.350	0.517	0.500	0.161	0.047	0.894

Problem#05



Problem Characteristics

Number of cells	Maximum Dimension (m)	Avgerage Cell Area (ha)	Area (ha)
1157	16981	24.923	28835

Population Density by State 1991 (inh/ha)	ULD	UHD	IND	N-UUrb
	13.83	28.68	10.00	1.00

Population Density by State 2001 (inh/ha)	ULD	UHD	IND	N-UUrb
	5.63	14.58	10.00	1.00

Total Area by State 1991 (ha)	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	2395	204	59	2115	382	23680

Total Area by State 2001 (ha)	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	3069	219	98	1426	342	23680

Area variation 1991/2001	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	28%	7%	67%	-33%	-10%	0%

Simulation Results

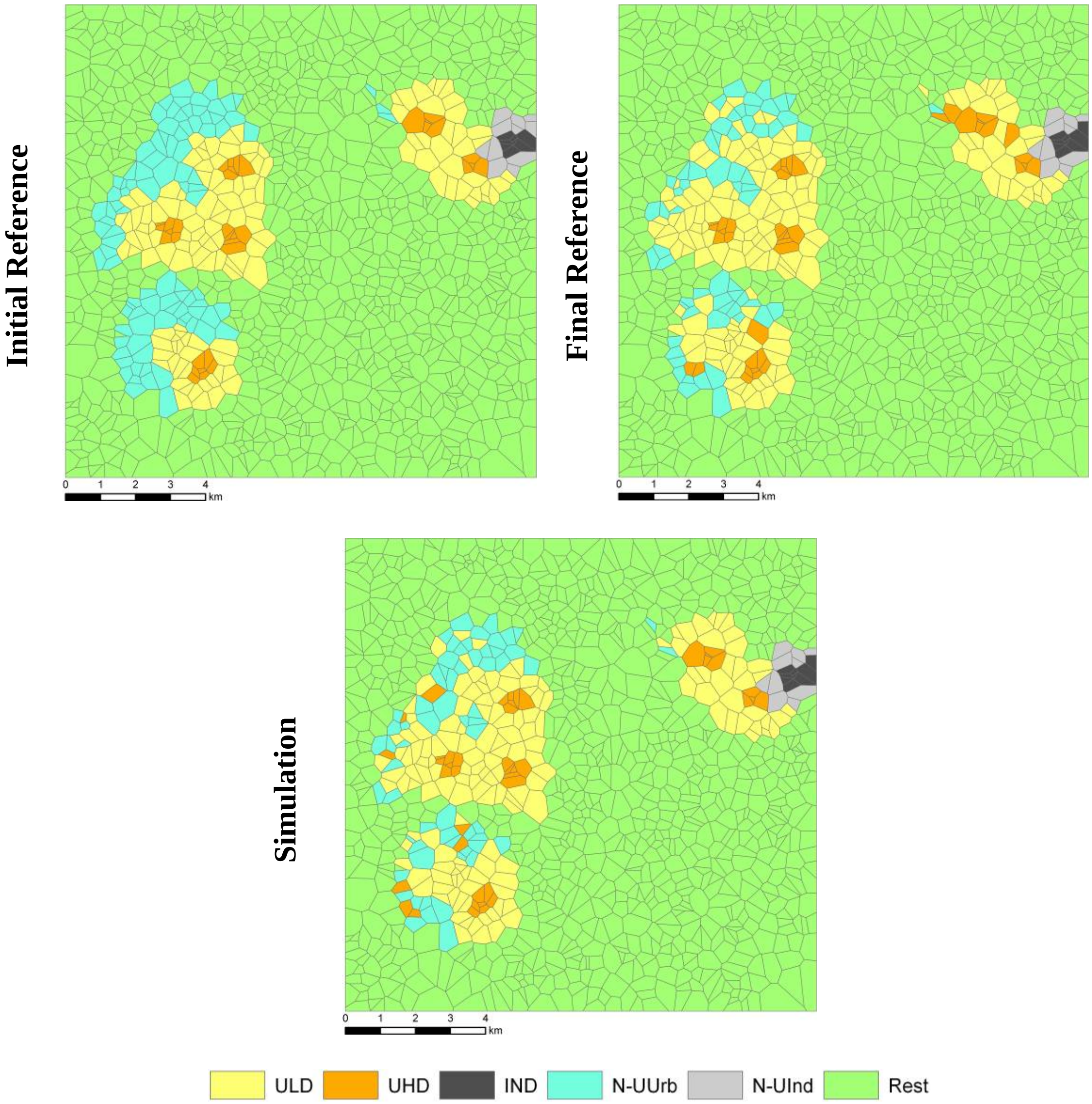
ModKvalue	*ModkValue* Increase	*kValue*	Overall Accuracy	Number of active cells	Proportion of active cells
0.801	51%	0.927	0.874	222	19%

Number of changed cells	Number of matching cells	Similarity (%)
27	14	52%

Θ_S (1991/2001)	ULD	UHD	IND	N-UUrb	N-UInd
	1%	-4%	3%	-1%	-1%

Neigbourhood Distance (km)	α_{acc}	β_{acc}	γ_{acc}	χ_{pot}	v_{pot}	θ_{pot}
6.1	1.104	0.949	0.455	0.589	0.469	0.760

Problem#06



Problem Characteristics

Number of cells	Maximum Dimension (m)	Avgerage Cell Area (ha)	Area (ha)
1182	13236	14.822	17519

Population Density by State 1991 (inh/ha)	ULD	UHD	IND	N-UUrb
	12.26	21.66	10.00	1.00

Population Density by State 2001 (inh/ha)	ULD	UHD	IND	N-UUrb
	6.19	14.22	10.00	1.00

Total Area by State 1991 (ha)	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	2118	296	55	1418	162	13470

Total Area by State 2001 (ha)	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	2537	392	61	903	156	13470

Area variation 1991/2001	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	20%	32%	10%	-36%	-3%	0%

Simulation Results

ModKvalue	*ModkValue* Increase	*kValue*	Overall Accuracy	Number of active cells	Proportion of active cells
0.784	35%	0.921	0.867	278	24%

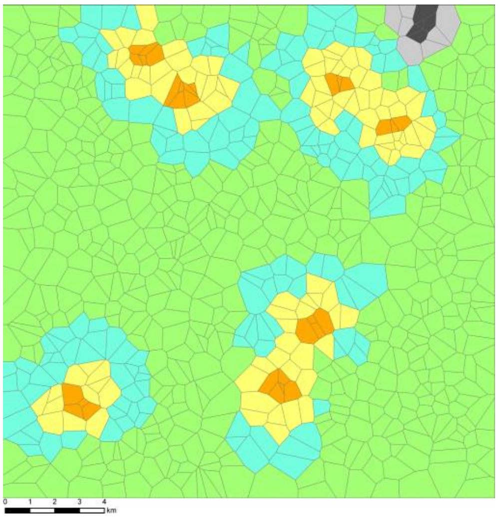
Number of changed cells	Number of matching cells	Similarity (%)
34	15	44%

Θ_S (1991/2001)	ULD	UHD	IND	N-UUrb	N-UInd
	2%	-9%	1%	-2%	0%

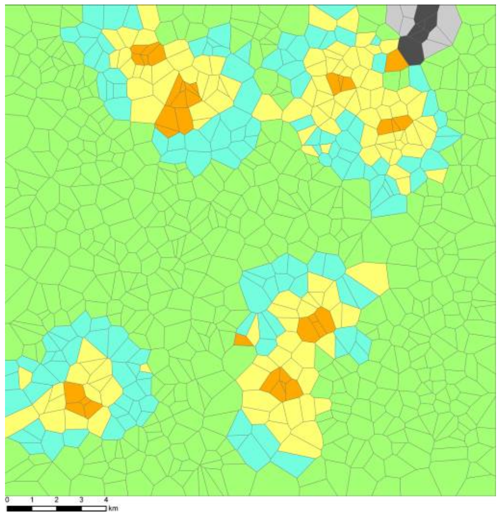
Neigbourhood Distance (km)	α_{acc}	β_{acc}	γ_{acc}	χ_{pot}	ν_{pot}	θ_{pot}
2.7	0.861	1.049	0.395	0.604	0.540	0.454

Problem#07

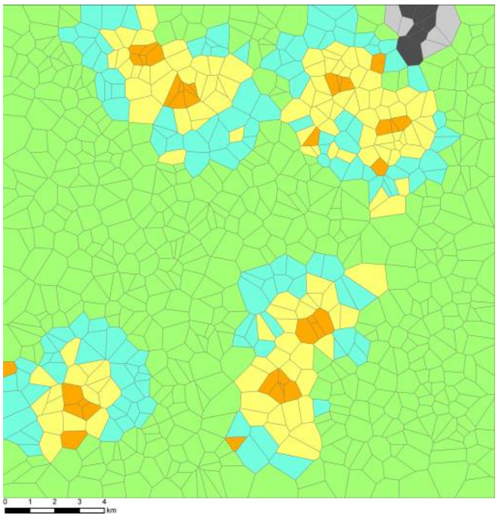
Initial Reference



Final Reference



Simulation



ULD UHD IND N-UUrb N-UInd Rest

Problem Characteristics

Number of cells	Maximum Dimension (m)	Avgerage Cell Area (ha)	Area (ha)
807	19492	47.080	37994

Population Density by State 1991 (inh/ha)	ULD	UHD	IND	N-UUrb
	10.69	24.64	10.00	1.00

Population Density by State 2001 (inh/ha)	ULD	UHD	IND	N-UUrb
	6.75	13.87	10.00	1.00

Total Area by State 1991 (ha)	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	4401	792	109	7814	380	24497

Total Area by State 2001 (ha)	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	5704	1024	202	6279	287	24497

Area variation 1991/2001	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	30%	29%	85%	-20%	-24%	0%

Simulation Results

ModKvalue	*ModkValue* Increase	*kValue*	Overall Accuracy	Number of active cells	Proportion of active cells
0.792	29%	0.912	0.869	306	38%

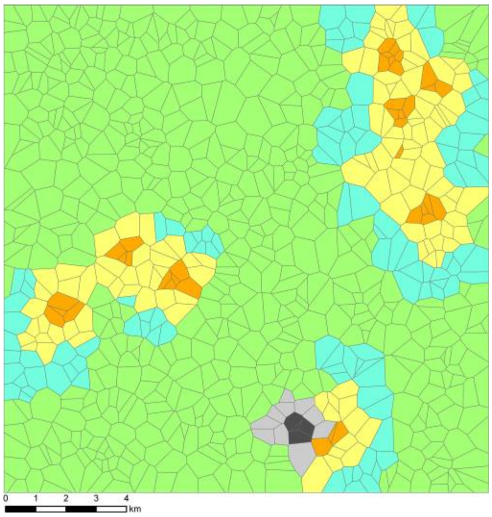
Number of changed cells	Number of matching cells	Similarity (%)
30	10	33%

Θ_S (1991/2001)	ULD	UHD	IND	N-UUrb	N-UInd
	0%	0%	28%	0%	-20%

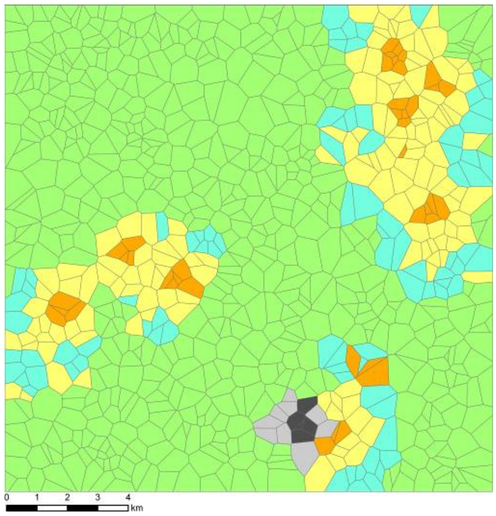
Neigbourhood Distance (km)	α_{acc}	β_{acc}	γ_{acc}	χ_{pot}	v_{pot}	θ_{pot}
8.0	0.607	0.508	0.470	1.071	0.681	1.051

Problem#08

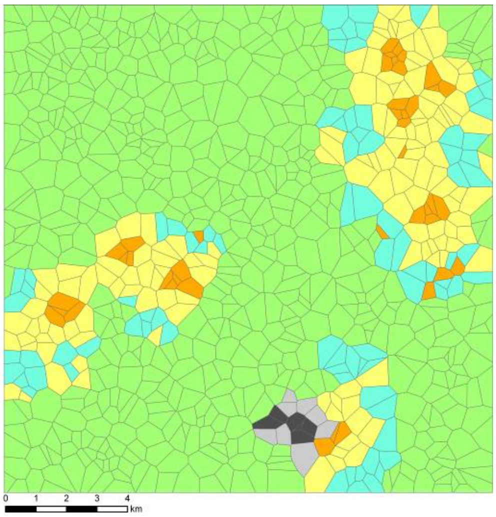
Initial Reference



Final Reference



Simulation



Problem Characteristics

Number of cells	Maximum Dimension (m)	Avgerage Cell Area (ha)	Area (ha)
832	15747	29.804	24797

Population Density by State 1991 (inh/ha)	ULD	UHD	IND	N-UUrb
	14.12	27.61	10.00	1.00

Population Density by State 2001 (inh/ha)	ULD	UHD	IND	N-UUrb
	7.32	13.51	10.00	1.00

	ULD	UHD	IND	N-UUrb	N-UInd	Rest
Total Area by State 1991 (ha)	3545	606	74	3299	309	16963
Total Area by State 2001 (ha)	4212	722	102	2516	281	16963
Area variation 1991/2001	19%	19%	37%	-24%	-9%	0%

Simulation Results

ModKvalue	*ModkValue* Increase	*kValue*	Overall Accuracy	Number of active cells	Proportion of active cells
0.841	25%	0.934	0.899	276	33%

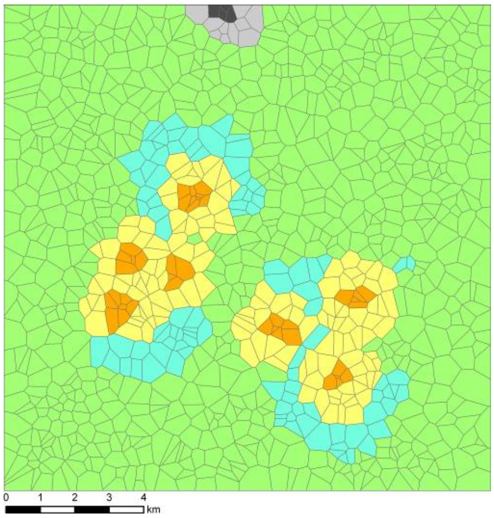
Number of changed cells	Number of matching cells	Similarity (%)
27	11	41%

	ULD	UHD	IND	N-UUrb	N-UInd
Θ_S (1991/2001)	1%	-2%	23%	-2%	-8%

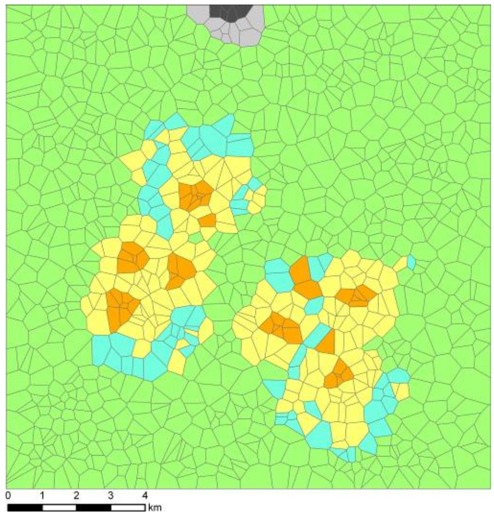
Neigbourhood Distance (km)	α_{acc}	β_{acc}	γ_{acc}	χ_{pot}	ν_{pot}	θ_{pot}
3.7	0.457	0.706	0.482	0.282	0.737	0.520

Problem#09

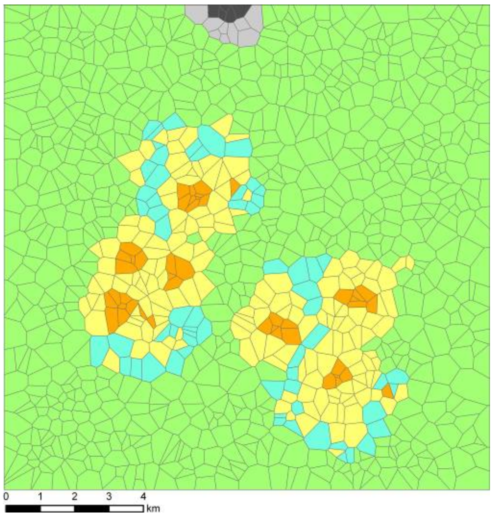
Initial Reference

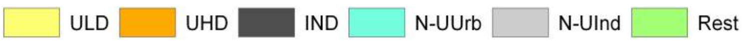


Final Reference



Simulation





Problem Characteristics

Number of cells	Maximum Dimension (m)	Avgerage Cell Area (ha)	Area (ha)
1045	13875	18.423	19252

	ULD	UHD	IND	N-UUrb
Population Density by State 1991 (inh/ha)	10.83	29.10	10.00	1.00

	ULD	UHD	IND	N-UUrb
Population Density by State 2001 (inh/ha)	6.50	11.14	10.00	1.00

	ULD	UHD	IND	N-UUrb	N-UInd	Rest
Total Area by State 1991 (ha)	2304	433	25	1829	160	14499

	ULD	UHD	IND	N-UUrb	N-UInd	Rest
Total Area by State 2001 (ha)	2759	544	35	1265	151	14499

	ULD	UHD	IND	N-UUrb	N-UInd	Rest
Area variation 1991/2001	20%	25%	37%	-31%	-6%	0%

Simulation Results

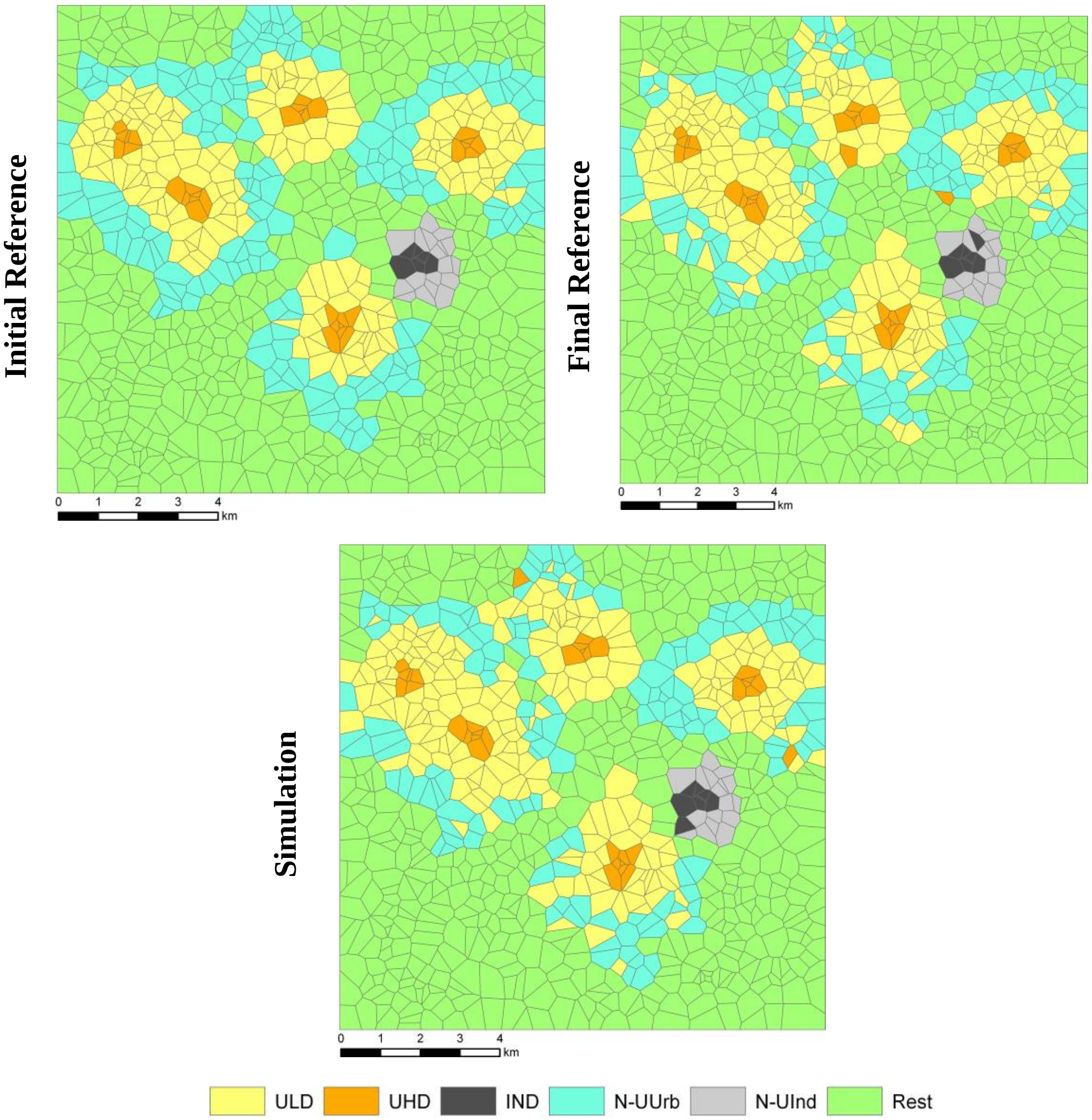
ModKvalue	*ModkValue* Increase	*kValue*	Overall Accuracy	Number of active cells	Proportion of active cells
0.802	28%	0.926	0.881	285	27%

Number of changed cells	Number of matching cells	Similarity (%)
40	20	50%

	ULD	UHD	IND	N-UUrb	N-UInd
Θ_S (1991/2001)	4%	-10%	31%	-4%	-7%

Neigbourhood Distance (km)	α_{acc}	β_{acc}	γ_{acc}	χ_{pot}	v_{pot}	θ_{pot}
2.3	0.668	0.841	0.193	0.418	0.390	0.654

Problem#10



Problem Characteristics

Number of cells	Maximum Dimension (m)	Avgerage Cell Area (ha)	Area (ha)
857	12003	16.811	14407

Population Density by State 1991 (inh/ha)	ULD	UHD	IND	N-UUrb		
	12.55	20.59	10.00	1.00		
Population Density by State 2001 (inh/ha)	ULD	UHD	IND	N-UUrb		
	5.68	13.16	10.00	1.00		
Total Area by State 1991 (ha)	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	2730	305	68	3200	245	7859
Total Area by State 2001 (ha)	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	3407	337	79	2492	234	7859
Area variation 1991/2001	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	25%	10%	16%	-22%	-5%	0%

Simulation Results

ModKvalue	*ModkValue* Increase	*kValue*	Overall Accuracy	Number of active cells	Proportion of active cells
0.803	21%	0.909	0.880	417	49%

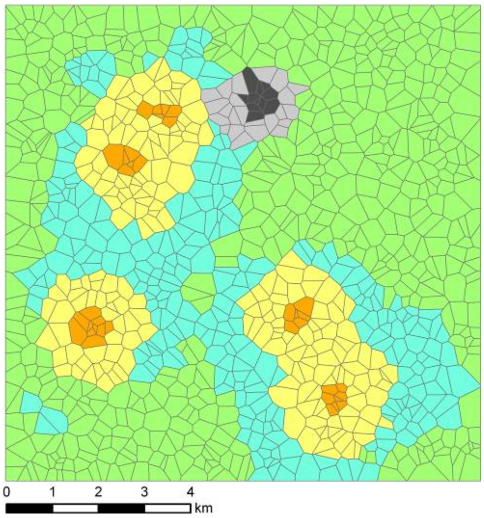
Number of changed cells	Number of matching cells	Similarity (%)
40	17	43%

Θ_S (1991/2001)	ULD	UHD	IND	N-UUrb	N-UInd
	0%	-2%	0%	0%	0%

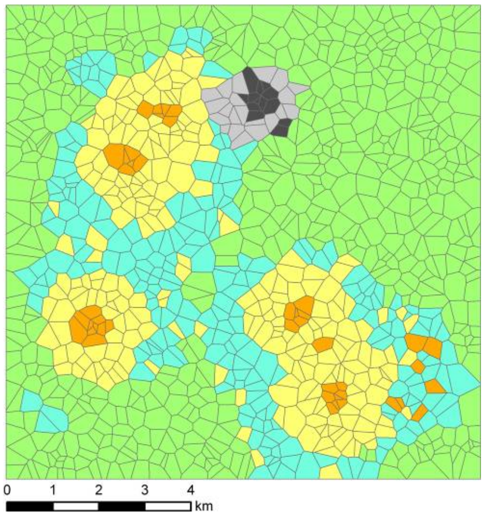
Neigbourhood Distance (km)	α_{acc}	β_{acc}	γ_{acc}	χ_{pot}	ν_{pot}	θ_{pot}
3.2	0.844	0.173	0.624	0.366	0.429	0.388

Problem#11

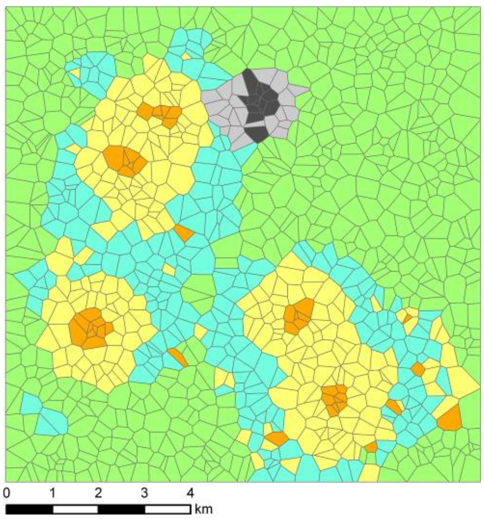
Initial Reference



Final Reference



Simulation



Problem Characteristics

Number of cells	Maximum Dimension (m)	Avgerage Cell Area (ha)	Area (ha)
1069	10130	9.599	10262

Population Density by State 1991 (inh/ha)	ULD	UHD	IND	N-UUrb
	14.26	22.08	10.00	1.00

Population Density by State 2001 (inh/ha)	ULD	UHD	IND	N-UUrb
	7.06	15.18	10.00	1.00

Total Area by State 1991 (ha)	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	1918	203	59	2519	208	5354

Total Area by State 2001 (ha)	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	2238	276	71	2126	196	5354

Area variation 1991/2001	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	17%	36%	20%	-16%	-6%	0%

Simulation Results

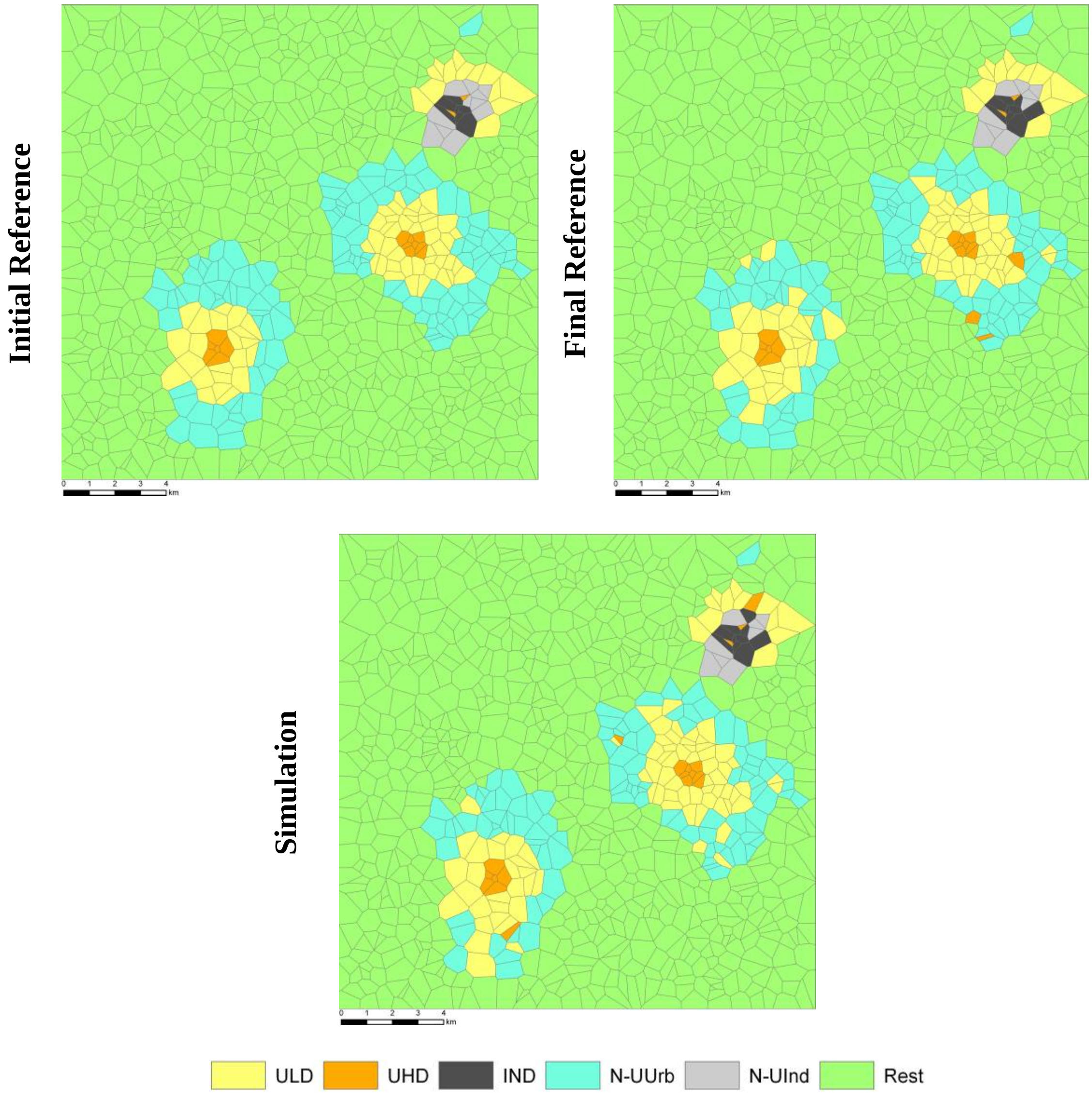
ModKvalue	*ModkValue* Increase	*kValue*	Overall Accuracy	Number of active cells	Proportion of active cells
0.811	15%	0.910	0.882	533	50%

Number of changed cells	Number of matching cells	Similarity (%)
43	12	28%

Θ_S (1991/2001)	ULD	UHD	IND	N-UUrb	N-UInd
	1%	-4%	3%	0%	-1%

Neigbourhood Distance (km)	α_{acc}	β_{acc}	γ_{acc}	χ_{pot}	v_{pot}	θ_{pot}
3.7	0.608	1.069	1.433	1.006	0.924	0.695

Problem#12



Problem Characteristics

Number of cells	Maximum Dimension (m)	Avgerage Cell Area (ha)	Area (ha)
882	18258	37.795	33335

	ULD	UHD	IND	N-UUrb
Population Density by State 1991 (inh/ha)	10.98	23.57	10.00	1.00
Population Density by State 2001 (inh/ha)	6.24	12.81	10.00	1.00

	ULD	UHD	IND	N-UUrb	N-UInd	Rest
Total Area by State 1991 (ha)	2636	228	143	4426	371	25532
Total Area by State 2001 (ha)	3096	290	194	3903	320	25532
Area variation 1991/2001	17%	27%	35%	-12%	-14%	0%

Simulation Results

ModKvalue	*ModkValue* Increase	*kValue*	Overall Accuracy	Number of active cells	Proportion of active cells
0.865	20%	0.948	0.914	233	26%

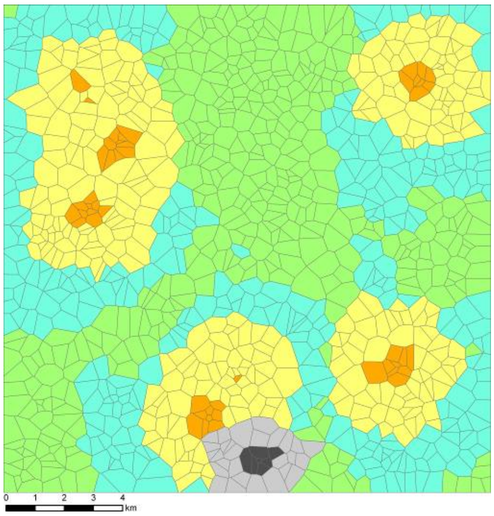
Number of changed cells	Number of matching cells	Similarity (%)
21	8	38%

	ULD	UHD	IND	N-UUrb	N-UInd
Θ_S (1991/2001)	2%	-1%	9%	-1%	-6%

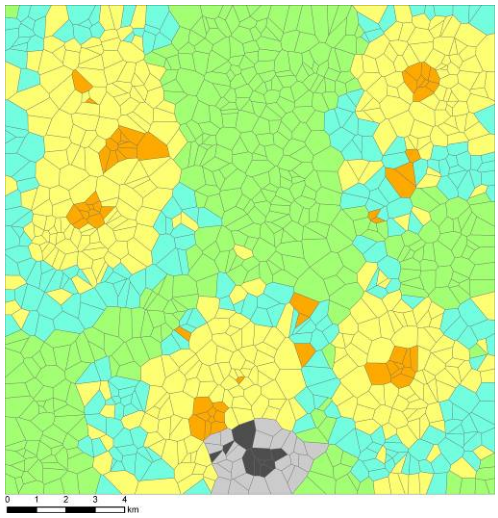
Neigbourhood Distance (km)	α_{acc}	β_{acc}	γ_{acc}	χ_{pot}	ν_{pot}	θ_{pot}
5.1	0.452	0.447	0.900	0.476	0.402	0.394

Problem#13

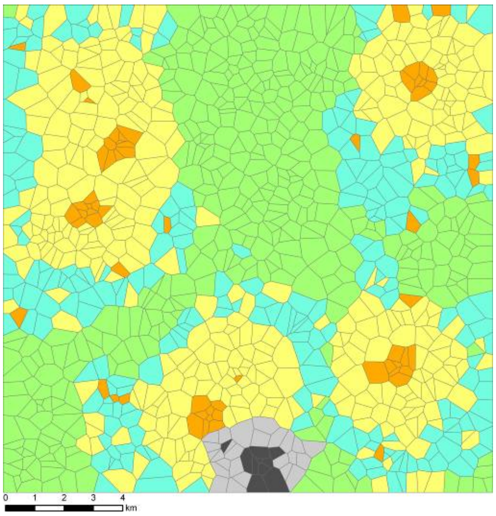
Initial Reference

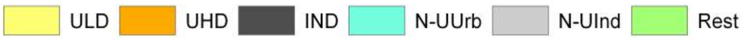


Final Reference



Simulation





Problem Characteristics

Number of cells	Maximum Dimension (m)	Avgerage Cell Area (ha)	Area (ha)
1094	16386	24.543	26850

Population Density by State 1991 (inh/ha)	ULD	UHD	IND	N-UUrb		
	12.69	25.05	10.00	1.00		
Population Density by State 2001 (inh/ha)	ULD	UHD	IND	N-UUrb		
	7.62	14.83	10.00	1.00		
Total Area by State 1991 (ha)	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	8105	714	105	8025	625	9276
Total Area by State 2001 (ha)	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	9857	1014	178	5974	552	9276
Area variation 1991/2001	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	22%	42%	70%	-26%	-12%	0%

Simulation Results

ModKvalue	*ModkValue* Increase	*kValue*	Overall Accuracy	Number of active cells	Proportion of active cells
0.745	20%	0.859	0.848	716	65%

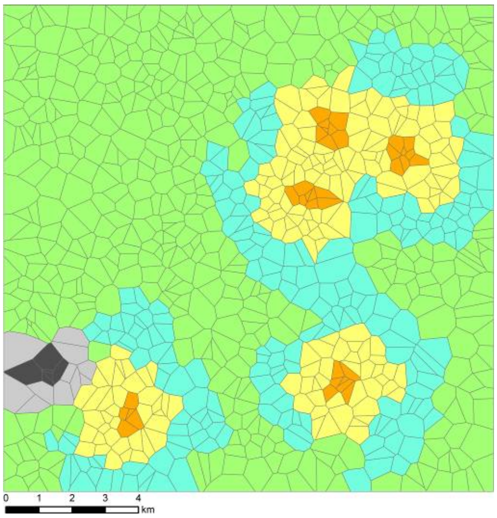
Number of changed cells	Number of matching cells	Similarity (%)
86	27	31%

Θ_S (1991/2001)	ULD	UHD	IND	N-UUrb	N-UInd
	2%	-8%	9%	-1%	-3%

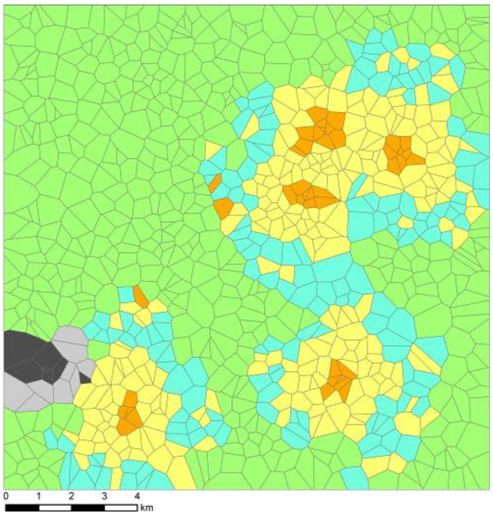
Neigbourhood Distance (km)	α_{acc}	β_{acc}	γ_{acc}	χ_{pot}	v_{pot}	θ_{pot}
5.6	0.621	0.566	0.840	1.318	0.930	0.583

Problem#14

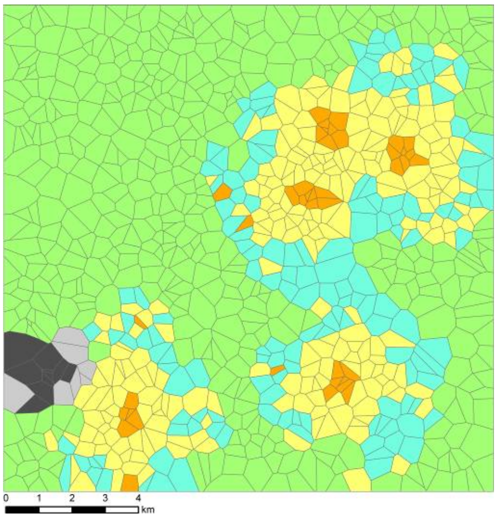
Initial Reference



Final Reference



Simulation



Problem Characteristics

Number of cells	Maximum Dimension (m)	Avgerage Cell Area (ha)	Area (ha)
907	14514	23.226	21066

	ULD	UHD	IND	N-UUrb	N-UInd	Rest
Population Density by State 1991 (inh/ha)	14.41	26.54	10.00	1.00		
Population Density by State 2001 (inh/ha)	6.80	12.45	10.00	1.00		
Total Area by State 1991 (ha)	3381	442	131	5077	405	11630
Total Area by State 2001 (ha)	4611	545	224	3743	313	11630
Area variation 1991/2001	36%	23%	70%	-26%	-23%	0%

Simulation Results

ModKvalue	*ModkValue* Increase	*kValue*	Overall Accuracy	Number of active cells	Proportion of active cells
0.735	32%	0.879	0.842	450	50%

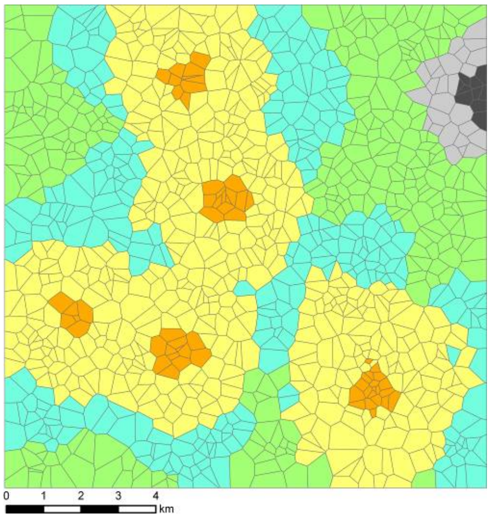
Number of changed cells	Number of matching cells	Similarity (%)
72	34	47%

	ULD	UHD	IND	N-UUrb	N-UInd
Θ_S (1991/2001)	0%	-5%	0%	0%	0%

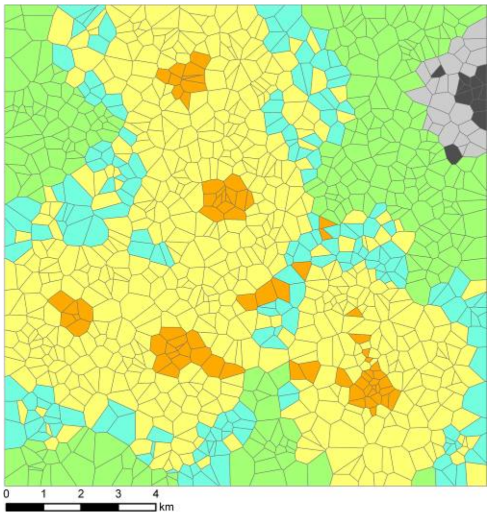
Neigbourhood Distance (km)	α_{acc}	β_{acc}	γ_{acc}	χ_{pot}	v_{pot}	θ_{pot}
3.0	0.481	0.487	0.245	0.972	0.569	0.584

Problem#15

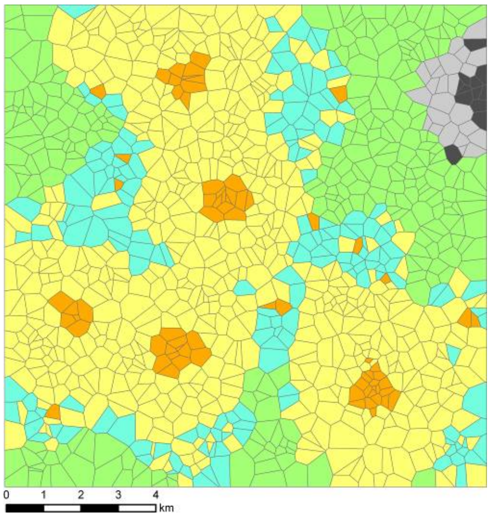
Initial Reference



Final Reference



Simulation



ULD UHD IND N-UUrb N-UInd Rest

Problem Characteristics

Number of cells	Maximum Dimension (m)	Avgerage Cell Area (ha)	Area (ha)
1119	12641	14.280	15979

Population Density by State 1991 (inh/ha)	ULD	UHD	IND	N-UUrb
	11.12	28.03	10.00	1.00
Population Density by State 2001 (inh/ha)	ULD	UHD	IND	N-UUrb
	5.98	14.47	10.00	1.00

Total Area by State 1991 (ha)	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	6943	593	71	4043	401	3929
Total Area by State 2001 (ha)	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	8243	809	96	2527	376	3929
Area variation 1991/2001	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	19%	36%	36%	-37%	-6%	0%

Simulation Results

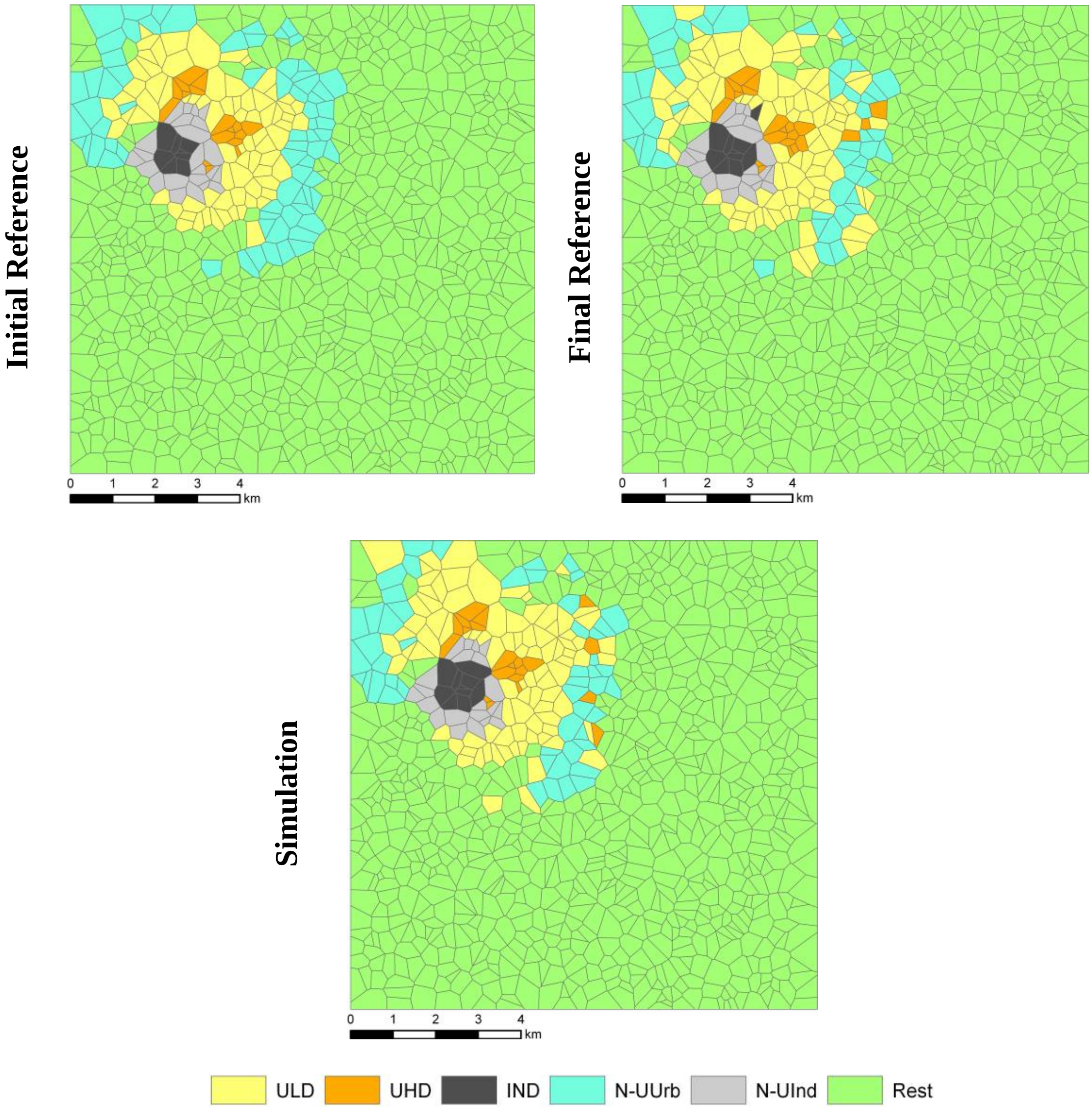
ModKvalue	*ModkValue* Increase	*kValue*	Overall Accuracy	Number of active cells	Proportion of active cells
0.718	27%	0.834	0.857	853	76%

Number of changed cells	Number of matching cells	Similarity (%)
110	46	42%

Θ_S (1991/2001)	ULD	UHD	IND	N-UUrb	N-UInd
	2%	-13%	14%	-4%	-4%

Neigbourhood Distance (km)	α_{acc}	β_{acc}	γ_{acc}	χ_{pot}	ν_{pot}	θ_{pot}
4.4	0.735	0.917	0.300	0.479	0.309	0.581

Problem#16



Problem Characteristics

Number of cells	Maximum Dimension (m)	Avgerage Cell Area (ha)	Area (ha)
932	10769	12.443	11597

	ULD	UHD	IND	N-UUrb
Population Density by State 1991 (inh/ha)	12.84	29.52	10.00	1.00
Population Density by State 2001 (inh/ha)	7.36	12.10	10.00	1.00

	ULD	UHD	IND	N-UUrb	N-UInd	Rest
Total Area by State 1991 (ha)	989	121	87	1071	254	9076
Total Area by State 2001 (ha)	1250	162	108	769	233	9076
Area variation 1991/2001	26%	34%	24%	-28%	-8%	0%

Simulation Results

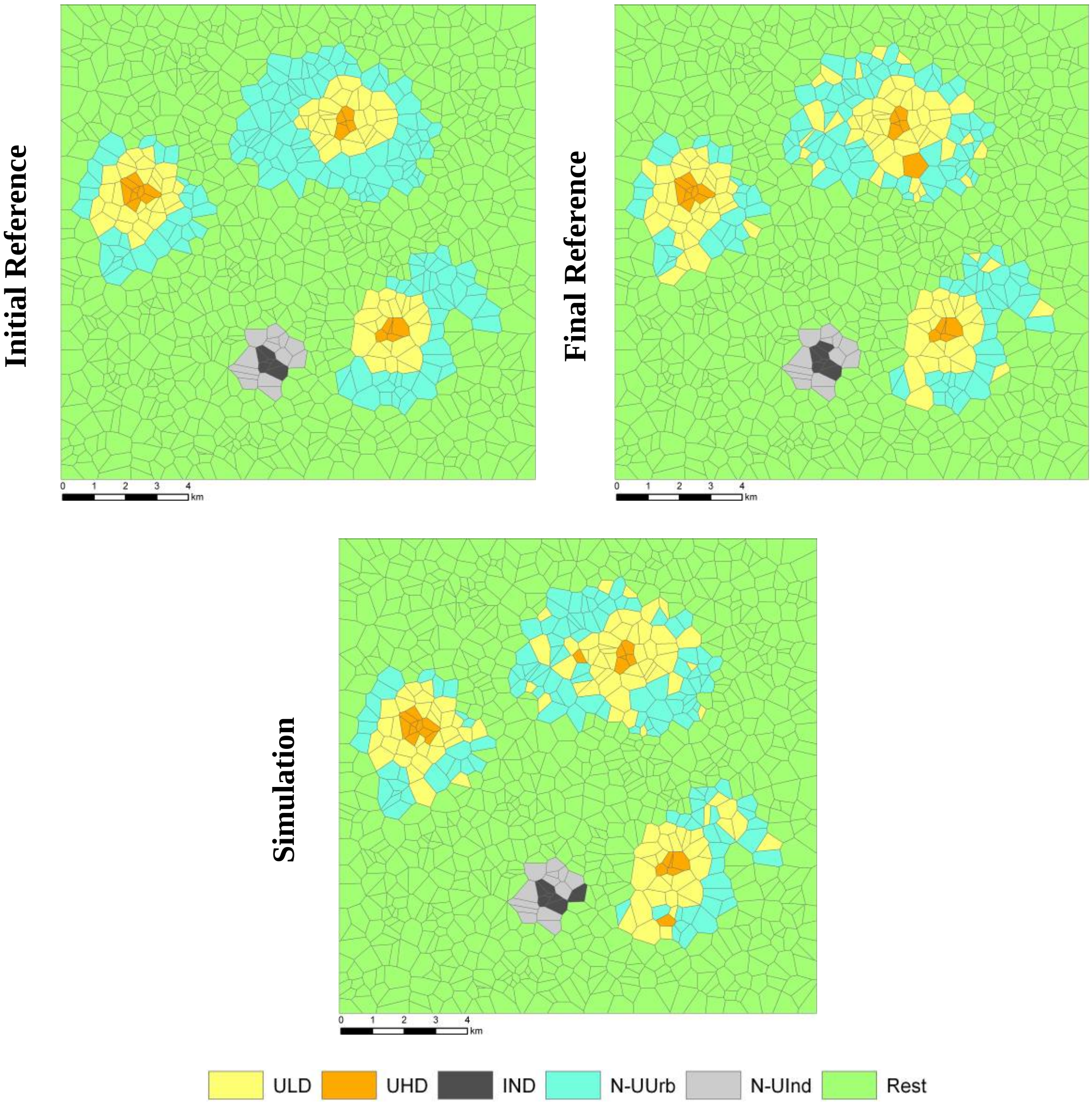
ModKvalue	*ModkValue* Increase	*kValue*	Overall Accuracy	Number of active cells	Proportion of active cells
0.811	27%	0.925	0.870	193	21%

Number of changed cells	Number of matching cells	Similarity (%)
21	10	48%

	ULD	UHD	IND	N-UUrb	N-UInd
Θ_S (1991/2001)	2%	-8%	13%	-1%	-6%

Neigbourhood Distance (km)	α_{acc}	β_{acc}	γ_{acc}	χ_{pot}	ν_{pot}	θ_{pot}
2.1	0.332	0.360	0.716	0.618	0.584	0.291

Problem#17



Problem Characteristics

Number of cells	Maximum Dimension (m)	Avgerage Cell Area (ha)	Area (ha)
1038	14833	21.196	22002

Population Density by State 1991 (inh/ha)	ULD	UHD	IND	N-UUrb
	13.69	25.26	10.00	1.00

Population Density by State 2001 (inh/ha)	ULD	UHD	IND	N-UUrb
	5.85	15.31	10.00	1.00

Total Area by State 1991 (ha)	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	1554	196	66	3419	288	16479

Total Area by State 2001 (ha)	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	2273	243	78	2652	276	16479

Area variation 1991/2001	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	46%	24%	18%	-22%	-4%	0%

Simulation Results

ModKvalue	*ModkValue* Increase	*kValue*	Overall Accuracy	Number of active cells	Proportion of active cells
0.746	35%	0.903	0.841	277	27%

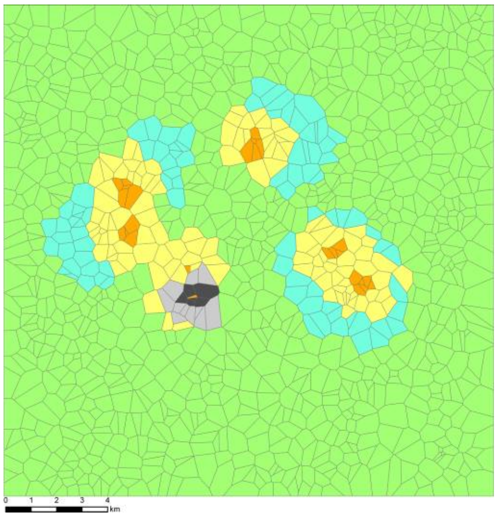
Number of changed cells	Number of matching cells	Similarity (%)
39	15	38%

Θ_S (1991/2001)	ULD	UHD	IND	N-UUrb	N-UInd
	3%	-7%	16%	-2%	-5%

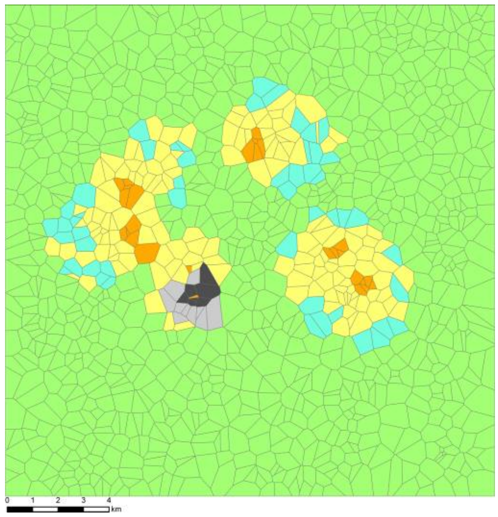
Neigbourhood Distance (km)	α_{acc}	β_{acc}	γ_{acc}	χ_{pot}	ν_{pot}	θ_{pot}
4.0	0.674	0.696	0.598	0.166	0.124	0.483

Problem#18

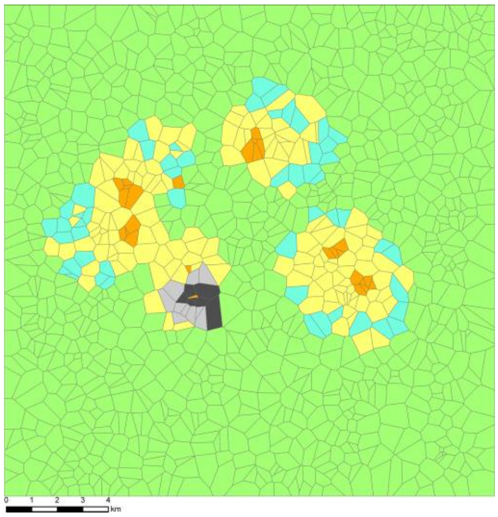
Initial Reference



Final Reference



Simulation



Problem Characteristics

Number of cells	Maximum Dimension (m)	Avgerage Cell Area (ha)	Area (ha)
1144	18897	31.215	35710

	ULD	UHD	IND	N-UUrb	N-UInd	Rest
Population Density by State 1991 (inh/ha)	14.55	21.01	10.00	1.00		
Population Density by State 2001 (inh/ha)	6.54	14.12	10.00	1.00		
Total Area by State 1991 (ha)	3154	345	107	2906	309	28889
Total Area by State 2001 (ha)	4257	410	145	1738	271	28889
Area variation 1991/2001	35%	19%	35%	-40%	-12%	0%

Simulation Results

ModKvalue	*ModkValue* Increase	*kValue*	Overall Accuracy	Number of active cells	Proportion of active cells
0.791	37%	0.929	0.879	240	21%

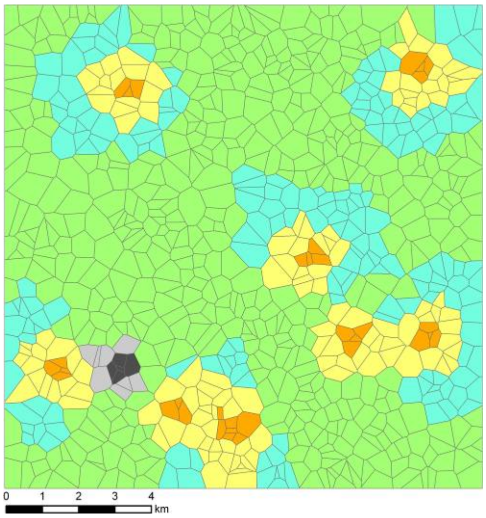
Number of changed cells	Number of matching cells	Similarity (%)
34	21	62%

	ULD	UHD	IND	N-UUrb	N-UInd
Θ_S (1991/2001)	1%	-13%	11%	0%	-6%

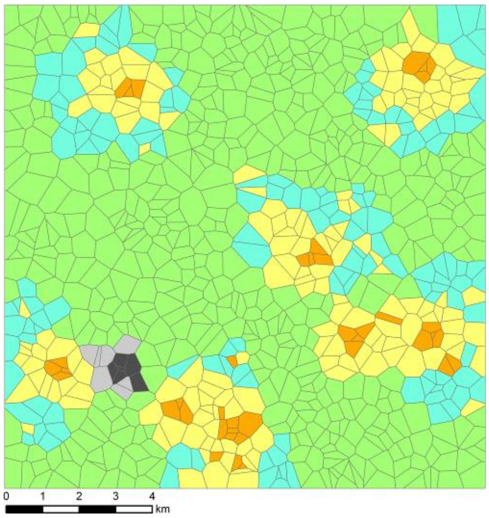
Neigbourhood Distance (km)	α_{acc}	β_{acc}	γ_{acc}	χ_{pot}	v_{pot}	θ_{pot}
4.9	0.739	0.502	0.293	0.608	0.272	0.544

Problem#19

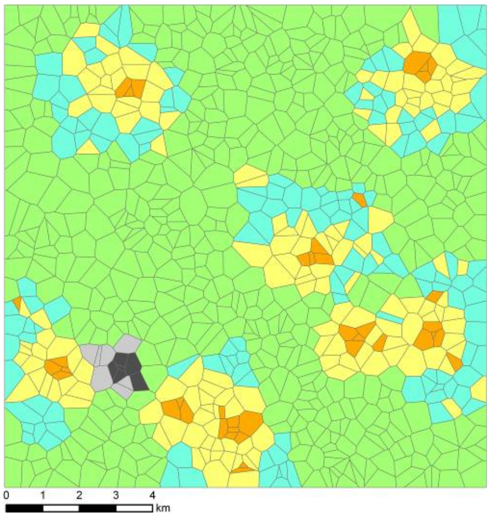
Initial Reference



Final Reference



Simulation



Problem Characteristics

Number of cells	Maximum Dimension (m)	Avgerage Cell Area (ha)	Area (ha)
850	12961	19.763	16799

Population Density by State 1991 (inh/ha)	ULD	UHD	IND	N-UUrb
	10.41	26.75	10.00	1.00

Population Density by State 2001 (inh/ha)	ULD	UHD	IND	N-UUrb
	7.23	12.93	10.00	1.00

Total Area by State 1991 (ha)	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	2512	376	52	3580	151	10129

Total Area by State 2001 (ha)	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	3178	438	70	2852	132	10129

Area variation 1991/2001	ULD	UHD	IND	N-UUrb	N-UInd	Rest
	27%	17%	35%	-20%	-12%	0%

Simulation Results

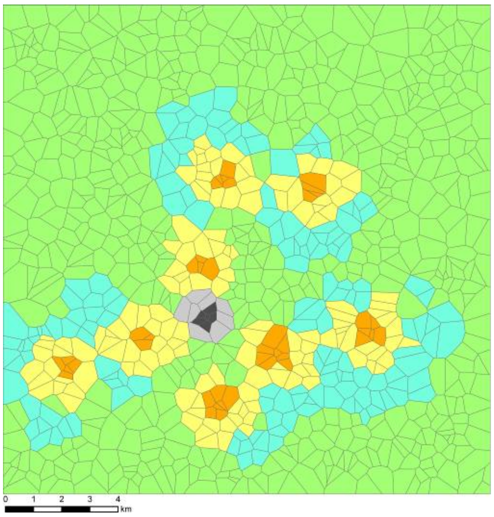
ModKvalue	*ModkValue* Increase	*kValue*	Overall Accuracy	Number of active cells	Proportion of active cells
0.757	26%	0.893	0.848	356	42%

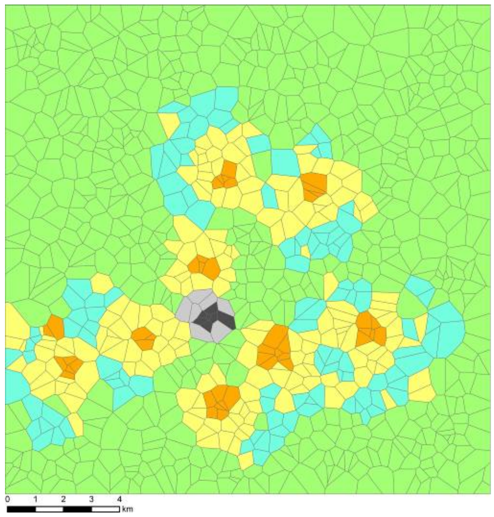
Number of changed cells	Number of matching cells	Similarity (%)
44	19	43%

Θ_S (1991/2001)	ULD	UHD	IND	N-UUrb	N-UInd
	0%	-2%	14%	0%	-7%

Neigbourhood Distance (km)	α_{acc}	β_{acc}	γ_{acc}	χ_{pot}	v_{pot}	θ_{pot}
2.4	0.760	0.493	0.475	0.383	0.252	0.285

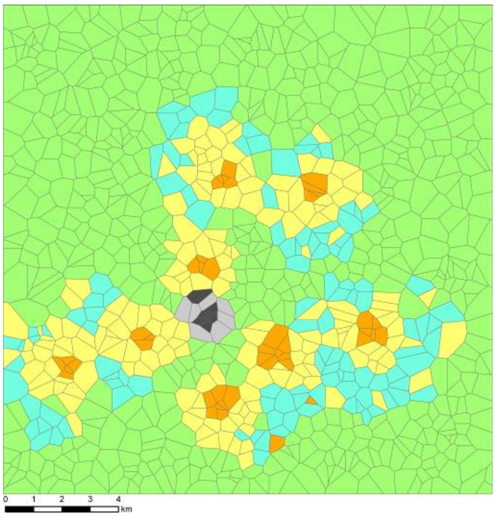
Problem#20





Final Reference

Simulation



Problem Characteristics

Number of cells	Maximum Dimension (m)	Avgerage Cell Area (ha)	Area (ha)
950	17025	30.511	28985

	ULD	UHD	IND	N-UUrb
Population Density by State 1991 (inh/ha)	11.27	22.50	10.00	1.00

	ULD	UHD	IND	N-UUrb
Population Density by State 2001 (inh/ha)	5.72	11.74	10.00	1.00

	ULD	UHD	IND	N-UUrb	N-UInd	Rest
Total Area by State 1991 (ha)	3818	665	50	4719	236	19498

	ULD	UHD	IND	N-UUrb	N-UInd	Rest
Total Area by State 2001 (ha)	4972	712	76	3518	210	19498

	ULD	UHD	IND	N-UUrb	N-UInd	Rest
Area variation 1991/2001	30%	7%	52%	-25%	-11%	0%

Simulation Results

ModKvalue	*ModkValue* Increase	*kValue*	Overall Accuracy	Number of active cells	Proportion of active cells
0.781	25%	0.909	0.863	344	36%

Number of changed cells	Number of matching cells	Similarity (%)
38	15	39%

	ULD	UHD	IND	N-UUrb	N-UInd
Θ_S (1991/2001)	1%	-3%	49%	-1%	-18%

Neigbourhood Distance (km)	α_{acc}	β_{acc}	γ_{acc}	χ_{pot}	ν_{pot}	θ_{pot}
4.3	0.818	0.142	0.665	0.432	0.099	0.761